

OVERVIEW OF OCEAN WAVE STATISTICS

Author: Mercè Casas Prat

Internal tutor: Joan Pau Sierra Pedrico

External tutor: Leo H. Holthuijsen (TU Delft)

Abstract

In order to predict high waves, understanding ocean wave height statistics has become an important matter for civil engineering, especially for safety and prevention reasons.

The present study, which focuses on deep water, has gained insight into some of the most important existing theories of short term statistics. It has been based on the analysis of 40,000 time wave records (approx. 10 million waves) from 4 buoys off the Catalan coast (Mediterranean Sea) from XIOM and complemented with about 9,000 waves registered by two laser altimeters from the WADIC project (North Sea). The original raw Mediterranean data (approx. 40 million waves) have been filtered through a quality control. Such a quality control and the subsequent statistical and spectral analysis have been carried out with a MATLAB code, developed by the author for the present study. The analysis has mainly consisted of the comparison of the observed and predicted values for the most important descriptive parameters of surface elevation: the significant and maximum wave height (also for the crest and trough).

The large amount of data has served to increase the reliability of the results since the sampling errors are diminished. The length of the samples (in terms of number of data points) is limited by a stationary requirement. However, by normalizing the records (dividing by the standard deviation of the surface elevation of each record), sets of large samples have been constructed, permitting the analysis of the maximum wave height (also crest and trough) for large numbers of waves. The comparison between the buoy and laser data has served to make evident, as in many other studies, that the buoy tends to avoid high crests and, because the buoy follows the orbital motion, it makes the surface profile more symmetric than it is in reality. However, the measurements of wave heights do not seem to be affected.

The classical linear theory, apart from the assumption of the surface elevation being a large sum of independent harmonic waves (and, therefore, being Gaussian distributed), assumes a narrowband spectrum. This theory results in the well-known Rayleigh distribution for the prediction of the wave heights (Longuet-Higgins, 1952). Nevertheless, measured waves are slightly smaller than the predictions made by this theory. Other, more advanced, theories have been presented in the present study. They have been grouped into two, depending on whether they consider the surface elevation Gaussian distributed or not. In the first case, they try to solve the overprediction of the linear theory which is mainly related to the wide character of the spectrum. Secondly, the possible interactions between waves are considered, which lead to the presence of nonlinearities. They appear to be related to the kurtosis and skewness of the surface elevation: the kurtosis to an enhancement of wave heights and the skewness to the presence of an asymmetric profile (sharper crests and more rounded troughs).

After the analysis of both sets of observations, it has been found that the significant wave height is overpredicted by 7% by the linear theory. Such a discrepancy is higher for the maximum wave height; in fact, the error made using the Rayleigh distribution increases toward the low-probability tail of the distribution. The presence of nonlinearities in the wave height appears to be weak. The overprediction of the wave height has been found to be mainly caused by the assumption in the Rayleigh theory of the wave height being double the crest height. The crest-to-trough theory (Tayfun, 1981b), which assumes a certain time lag between crests and troughs, has been found to be the most appropriate theory but its results are not convincing in the sense that errors are 2-7 % for the most characteristic wave heights such as the mean wave height, rms wave height and significant wave height. However, if, for simplicity, results are needed within a 10% accuracy for estimating these wave heights, the Rayleigh distribution is quite acceptable (except for the maximum wave height). However, such accuracy may not be acceptable for engineering purposes. More research is needed to better quantify the effect of the spectral width.

Contrary to the wave height results, the observations of crests and troughs have been found to differ noticeably between the buoy and laser records. In the first case, crest and trough levels were reasonably well predicted by the linear theory, the overprediction being, in general, less than 2%. In the laser records, they clearly manifested a nonlinear character typical of the physical pattern of pointed higher crests and shallower flattened troughs. Such nonlinearities were especially strong in the maximum wave crest / trough, in which the nonlinear theory of Tayfun (1994), and using an effective number of waves (Cartwright, 1958), underpredicted observed crests whereas overpredicted measured troughs. A possible reason is the sensitivity of results to the skewness, which is not a very robust parameter.

CONTENTS

| | |
|--|-----------|
| ACKNOWLEDGEMENTS | vi |
| 1. INTRODUCTION | 1 |
| 1.1. Research motivation..... | 1 |
| 1.2. Objectives..... | 1 |
| 2. ANALYSED DATA | 3 |
| 2.1. Mediterranean data | 3 |
| 2.2. Quality control of Mediterranean data..... | 4 |
| 2.2.1. Introduction..... | 4 |
| 2.2.2. Rough anomalies..... | 5 |
| 2.2.3. Buoy limitations | 9 |
| 2.2.4. Visual check | 11 |
| 2.2.5. Shallow water | 11 |
| 2.2.6. Comments | 12 |
| 2.3. North Sea data | 12 |
| 3. STATISTICAL ANALYSIS | 15 |
| 3.1. Introduction..... | 15 |
| 3.2. Definitions of height, crest, trough and period | 15 |
| 3.3. Parameters..... | 16 |
| 4. SPECTRAL ANALYSIS | 18 |
| 4.1. Introduction..... | 18 |
| 4.2. Fourier theory | 18 |
| 4.2.1. Continuous function..... | 18 |
| 4.2.2. Discrete function..... | 19 |
| 4.3. The wave spectrum | 19 |
| 4.3.1. Theoretical definition | 19 |
| 4.3.2. Practical limitations | 19 |
| 4.3.3. Windowing | 23 |
| 4.4. Wave spectrum calculation..... | 26 |
| 4.5. Spectral parameters | 28 |
| 4.5.1. Parameters related with the surface elevation..... | 28 |
| 4.5.2. Shape parameters | 29 |
| 4.5.3. BFI..... | 30 |
| 5. LINEAR THEORY | 31 |
| 5.1. Probability distributions..... | 31 |
| 5.1.1. Surface elevation..... | 31 |
| 5.1.2. Wave crest..... | 31 |
| 5.1.3. Wave height..... | 34 |
| 5.1.4. Maximum wave crest/height | 35 |
| 5.2. Parameters | 38 |
| 5.2.1. The significant wave height | 38 |
| 5.3. Limitations of Rayleigh theory..... | 40 |

| | |
|--|-----------|
| 6. NONLINEAR THEORY..... | 42 |
| 6.1. Introduction..... | 42 |
| 6.2. Inclusion of kurtosis | 42 |
| 6.2.1. Surface elevation..... | 42 |
| 6.2.2. Wave crest..... | 43 |
| 6.2.3. Wave height..... | 44 |
| 6.2.4. Estimation of kurtosis from the spectrum: the BFI | 44 |
| 6.3. Inclusion of skewness..... | 46 |
| 6.3.1. Surface elevation..... | 46 |
| 6.3.2. Wave crest/trough | 46 |
| 6.3.3. Wave height..... | 50 |
| 6.3.4. Estimation of skewness from the spectrum | 50 |
| 7. CREST-TO-TROUGH WAVE HEIGHT DISTRIBUTION | 51 |
| 7.1. Introduction..... | 51 |
| 7.2. Wave height | 51 |
| 8. IMPROVEMENTS OF RAYLEIGH DISTRIBUTION..... | 56 |
| 8.1. Introduction..... | 56 |
| 8.2. Modified Rayleigh distribution I (Longuet-Higgins) | 56 |
| 8.3. Modified Rayleigh distribution II (Naess) | 57 |
| 8.4. Comparison | 58 |
| 9. CONSIDERATIONS FOR EXPECTED MAXIMA | 59 |
| 10. SUMMARY OF PROBABILITY DISTRIBUTIONS | 62 |
| 11. THE MAXIMUM WAVE HEIGHT PARADOX..... | 64 |
| 11.1. Introduction..... | 64 |
| 11.2. Possible reasoning | 64 |
| 12. ANALYSIS OF THE MEDITERRANEAN DATA..... | 66 |
| 12.1. Introduction..... | 66 |
| 12.2. Wave height..... | 66 |
| 12.2.1. Significant wave height..... | 66 |
| 12.2.2. Maximum wave height..... | 73 |
| 12.2.3. Summary of results..... | 75 |
| 12.3. Wave crest/trough | 76 |
| 12.3.1. Significant wave crest/trough..... | 76 |
| 12.3.2. Maximum wave crest/trough..... | 77 |
| 12.4. Nonlinearities..... | 79 |
| 12.4.1. Introduction..... | 79 |
| 12.4.2. Kurtosis and Skewness | 79 |
| 12.4.3. BFI..... | 85 |
| 12.5. Conclusion..... | 87 |
| 13. ANALYSIS OF THE NORTH SEA DATA | 88 |
| 13.1. Introduction..... | 88 |
| 13.2. Results | 88 |
| 14. CONCLUSIONS..... | 99 |

| | |
|---|------------|
| APPENDIX A: Quality control | 101 |
| APPENDIX B: Freak waves | 103 |
| APPENDIX C: Matlab code..... | 106 |
| REFERENCES..... | 124 |
| OTHER CONSULTED BIBLIOGRAPHY | 127 |

ACKNOWLEDGEMENTS

First of all, I would like to thank dr.ir. Leo H. Holtuijsen, my thesis project supervisor at *Technische Universiteit Delft* (TU Delft), for giving me the chance to carry out this research which will be presented at the 2008 ICCE, and, above all, for his daily theoretical, practical and personal support. Thanks are also given to dr.ir Pieter H.A.J.M. Van Gelder for his valuable assistance.

I also want to express my gratitude to all the staff at the Department of Hydraulic Engineering at TU Delft, especially the Section of Environmental Fluid Mechanics, for their pleasant welcome and help throughout my work.

I acknowledge Dr. Joan Pau Sierra Pedrico from *Universitat Politècnica de Catalunya* (UPC), for accepting to be my supervisor in Barcelona.

Furthermore, I am very grateful to the people who have provided me with data, without which this study would not have been possible; firstly to all the staff at *Laboratori d'Enginyeria Marítima* of my home university (UPC), especially Jesús Gómez Aguar, who supplied the data for the four buoys off the Catalan coast, which are managed by XIOM (*Xarxa d'Instruments Oceanogràfics i Meteorològics de la Generalitat de Catalunya*). Secondly, I thank Ph. D. Stephen Barstow, senior ocean wave climatologist, who gave me permission to use some of the WADIC project storm data, in Norway.

Last but not least, I need to express my best deepest appreciation to the invaluable support of my partner, Iván, and my parents for their constant support and encouragement.

1. INTRODUCTION

1.1. Research motivation

In order to predict high waves, understanding ocean wave height statistics has become an important matter for civil engineering, especially for safety and prevention reasons. Since Longuet-Higgins (1952) introduced the Rayleigh distribution for describing the probability of wave height occurrence under certain conditions, based on results from Rice (1945), it has become the basis for describing wave height statistics. Many studies (Al-Humoud et al., 2001; Cartwright & Longuet-Higgins, 1956; Forristall, 1978; Longuet-Higgins, 1957, 1963, 1975, 1980, 1984; Mori & Yasuda, 2001; Tayfun, 1981a, 1981b, 1983, 1984, 1990, 1994, 2004, 2006, 2007; etc.) have been carried out in the last half century in order to find a better adjusted probability distribution, either theoretical or empirical, or simply to demonstrate agreement by using observations in the field. However, no clear conclusion has been derived as yet. The present study intends to give an overview of the most important existing theories and compare them with the observations from the Mediterranean and North Sea.

The motivation for the present research was the apparent “Maximum wave height paradox”. On the one hand, one found that the significant wave height was overpredicted by linear theory (Holthuijsen, 2007), having a discrepancy of approximately 7.5%. On the other hand, Cartwright (1958) found that the maximum wave crest was reasonably well predicted by a modified linear theory, including the spectral bandwidth of the spectrum. For his observations, such an inclusion caused a difference with the original linear theory lower than 2%, being far from 7.5%. This discrepancy led to investigate deeply the wave height statistics.

1.2. Objectives

The present study, which focuses on deep water, aims to gain an insight into some of the most important existing theories of short term statistics. It is based on the analysis of 40,000 time wave records (approx. 10 million waves) from 4 buoys off the Catalan coast (Mediterranean Sea) and 69 wave records (approx. 9,000 waves) from 2 laser altimeters in the North Sea (the WADIC project).

The Mediterranean data is the result of a meticulous quality control process performed on raw data directly obtained from the buoys during the period of 1991-2006. This large amount of data will hopefully serve to obtain more reliable results and diminish sampling errors. Although the length for each record is limited by the requirement of stationary, by normalizing the records with the standard deviation of the surface elevation, one can construct sets of large samples.

Although the amount of the Mediterranean data is clearly enough, a complementary analysis has been made with 69 wave records from the North Sea obtained with two laser altimeters, which were used in the WADIC project (Allender et. al, 1989). The reasons are mainly twofold: the inclusion of higher wave heights and a comparison between statistics of buoy and laser registrations. The buoy might tend to avoid the wave crests and therefore underestimate them whereas a laser altimeter does not suffer from this effect.

The starting point is the well-known Rayleigh distribution for wave heights for a narrow-band spectrum, assuming the surface elevation being Gaussian distributed. It will be shown that,

although the Rayleigh distribution is derived for such limited conditions, its capacity of prediction is reasonably good. However, the Rayleigh distribution often overpredicts the observed wave height (Holthuijsen, 2007; Forristall, 1978 1984; Longuet-Higgins, 1980; Naess, 1985; Nolte et al., 1979; Rodríguez et al., 2002; Vinje 1989, etc.) but sometimes the effect is the opposite, observed higher wave heights are larger than those predicted (Mori & Janssen 2005). Therefore, other theories are studied, trying to describe better wave height statistics. Firstly, some versions of linear theory are made (Longuet-Higgins, 1980; Tayfun, 1981b, 1990; Naess, 1985) in order to try to solve the overprediction problem. That is partly related with assumption of narrow-band spectrum. Secondly, nonlinearities are invoked (Al-Humoud et al., 2001; Mori & Yasuda, 2001; Tayfun, 1983, 1984, 1994, 2006, 2007) with the aim of understanding the reason for encountering large wave heights in some records. In that case, the surface elevation is not considered to be Gaussian distributed since terms of higher order are included.

In general, as the theories include more general situations and are less limited by assumptions, the formulation becomes noticeably more complicated, in some cases leading to complex expressions for the probability density function and therefore requiring numerical integration. One important question is whether such more complicated probability distributions should be preferred, or, otherwise, the simpler Rayleigh distribution. One consideration is that it is preferable to have a formulation depending only on spectral parameters (which can be predicted from wind prediction) than on analysis from time records themselves.

In conclusion, the principal objectives are:

- Statistical and spectral analysis of the Mediterranean data, being the raw data previously filtered.
- Comparison of observations to linear theory predictions, and, if necessary, to more advanced theories.
- Complementary analysis, using the North Sea data (from 2 laser altimeters).

2. ANALYSED DATA

2.1. Mediterranean data

Several observation techniques are nowadays available which are based on different wave measurement devices. One of the most common is a buoy because of its simplicity and lower associated costs. The buoy follows the three-dimensional motion of wave particles at the sea surface and their vertical acceleration is measured. By integrating the acceleration twice, the vertical motion of the buoy is obtained and therefore the surface elevation.

The analysed data in the present study consists of raw data from several time series from four buoys off the Catalan coast; Spain (Mediterranean Sea): Roses, Tordera, Llobregat and Tortosa (see Figure 2.1). Their names come from, respectively: the Roses gulf, the Tordera delta, the Llobregat delta and the Cape of Tortosa. The buoy of Tordera is usually called Blanes, which is the name of the coast city nearest the buoy's location.

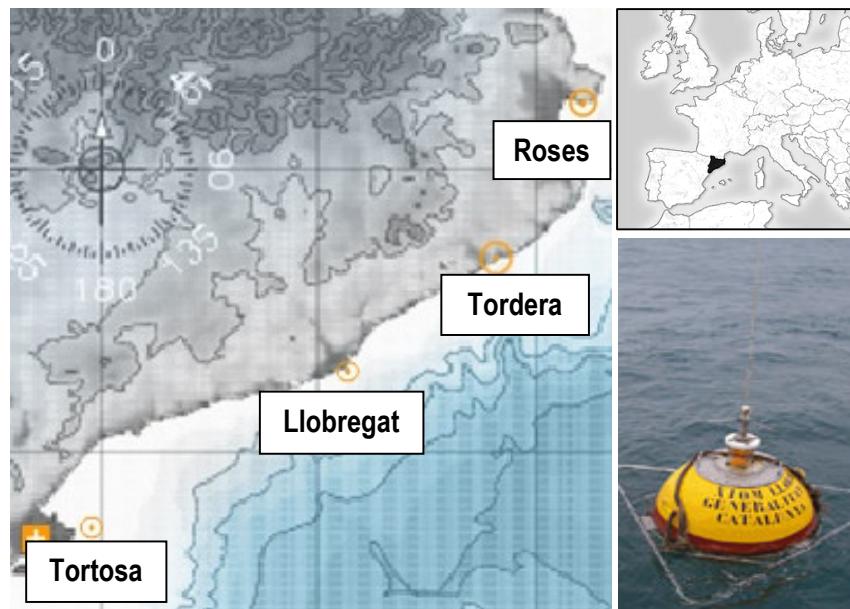


Figure 2.1 Situation of the four buoys

The waves were recorded during the period 1991-2006. The recording frequency (number of records per day) is not constant and many gaps occur (see Table 2.1). For example, in 2003 the data from Tortosa are missing. The reasons are possibly associated with either buoy breakdowns or computing problems. The memory requirement of the raw data (of surface elevations) is high. In situations where such memory is unavailable, only the relevant parameters of the surface elevation such as, for example, the significant and maximum wave heights are sent to the control station, and then the rest of the recorded information is deleted.

All the buoys belong to the XIOM network (*Xarxa d'Instruments Oceanogràfics i Meteorològics de la Generalitat de Catalunya*) and are Waverider buoys of DATAWELL. The most southern buoy (Tortosa) is a directional buoy whereas the remaining ones are scalar buoys. Typically, two types of directional buoys exist: those measuring the slope of

the sea surface (heave and pitch-roll motion) or those measuring the horizontal motion (surge and sway). In this case, a second type is used, more specifically, the Directional Waverider MK II.

Due to their duration of 20 minutes, the records are considered to be stationary. The sample time interval differs between the directional buoy and the rest (see Table 2.1). The registered files are saved in their respective year folder and their format is month-day-time-extension. Each buoy has a different extension (see Table 2.1) as well as the number of columns into which the data is organized. For example 03290900.RAW corresponds to the Tortosa buoy and was recorded on 29/03 at 09:00h.

Table 2.1: Description of buoys and their measurements

| Buoy | Roses | Blanes | Llobregat | Tortosa |
|-----------------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| Coordinates | 42 10.79 N 03 11.99 E | 41 38.81 N 02 48.93 E | 41 16.69 N 02 08.48 E | 40 43.29 N 00 58.89 E |
| Depth (m) | 46 | 74 | 45 | 60 |
| Diameter (m) | 0.7 | 0.7 | 0.7 | 0.9 |
| Sample interval (s) | 1/2.56 | 1/2.56 | 1/2.56 | 1/1.28 |
| Resolution | cm | cm | cm | Cm |
| Type of buoy | Scalar | Scalar | Scalar | Directional |
| Record duration (min) | 20 | 20 | 20 | 20 ¹ |
| Available recorded period (years) | 2001-2006 | 2002-2006 | 2001-2004 | 1991-1997, 2001-2006 |
| File extension | .2RW | .3RW | .1RW | RAW |

The raw data, as the name indicates, comes from the direct integration of the measured vertical acceleration of the buoy and has not been filtered through any quality control to detect anomalies, for instance, gaps or spikes. Therefore, a code has been developed to reject any records which do not pass certain quality requirements. The option of repairing the records containing errors has not been considered because, apart from its difficulty, it is not necessary since the remaining number of records after quality control is still sufficient. These quality requirements are explained in Section 2.2 .

After applying the quality control to the data the statistical and spectral analysis are carried out, which has been programmed in MATLAB code (see Appendix C). This is done separately for each year and buoy.

2.2. Quality control of Mediterranean data

2.2.1. Introduction

The main objective of the quality control is the automatic rejection of any record which contains anomalies derived from errors in the buoy's registration. Different types of errors have been found, such as rough errors, which are mostly spikes and gaps. But one has to consider other errors, which are not rough but are associated with intrinsic limitations of the device's mechanism. One of the most important is the trend in the buoy signal, which is

¹ There are some exceptions (see Section 2.2.2)

removed by subtracting a fitted straight line. Moreover, a possible effect of aliasing due to the sampling interval is considered. In addition, errors due to externalities are present. When for whatever reason the buoy is hit (a bump), a low frequency oscillation is artificially added, distorting the results.

Although the quality control may seem an easy task, it is not. In fact, despite this not being the main objective of the present study, it became tricky and laborious.

2.2.2. Rough anomalies

First of all, definitions of the most common anomalies are necessary:

Spike: very large elevation (either positive or negative) caused by an error in the buoy registration (see Figure 2.2 and Table 2.2).

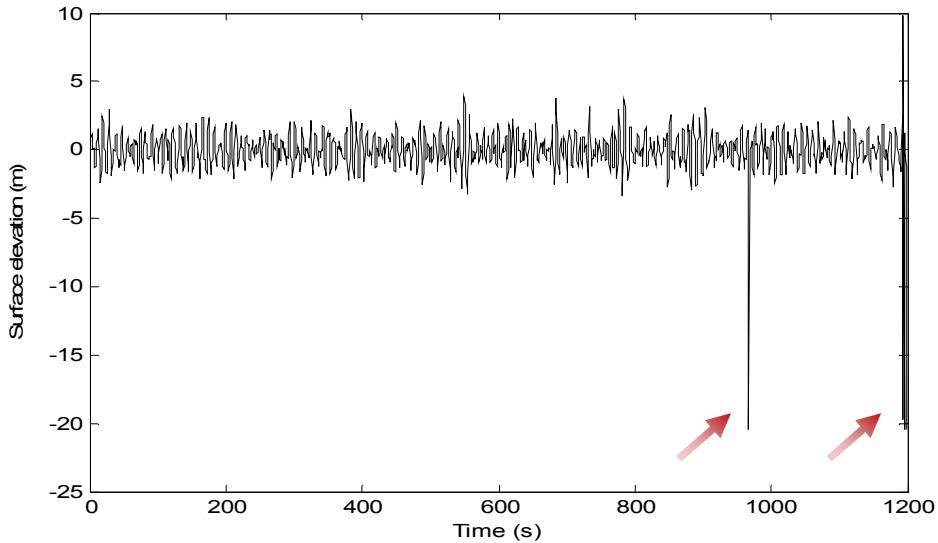


Figure 2.2 Record with spikes (Tortosa 29/03/2004 09:00h)

Gap: missing information in the record, being present in different forms: the values are simply absent (and therefore the number of data points is less) or they are replaced by zeros or other configurations (see Figure 2.3 and Table 2.2).

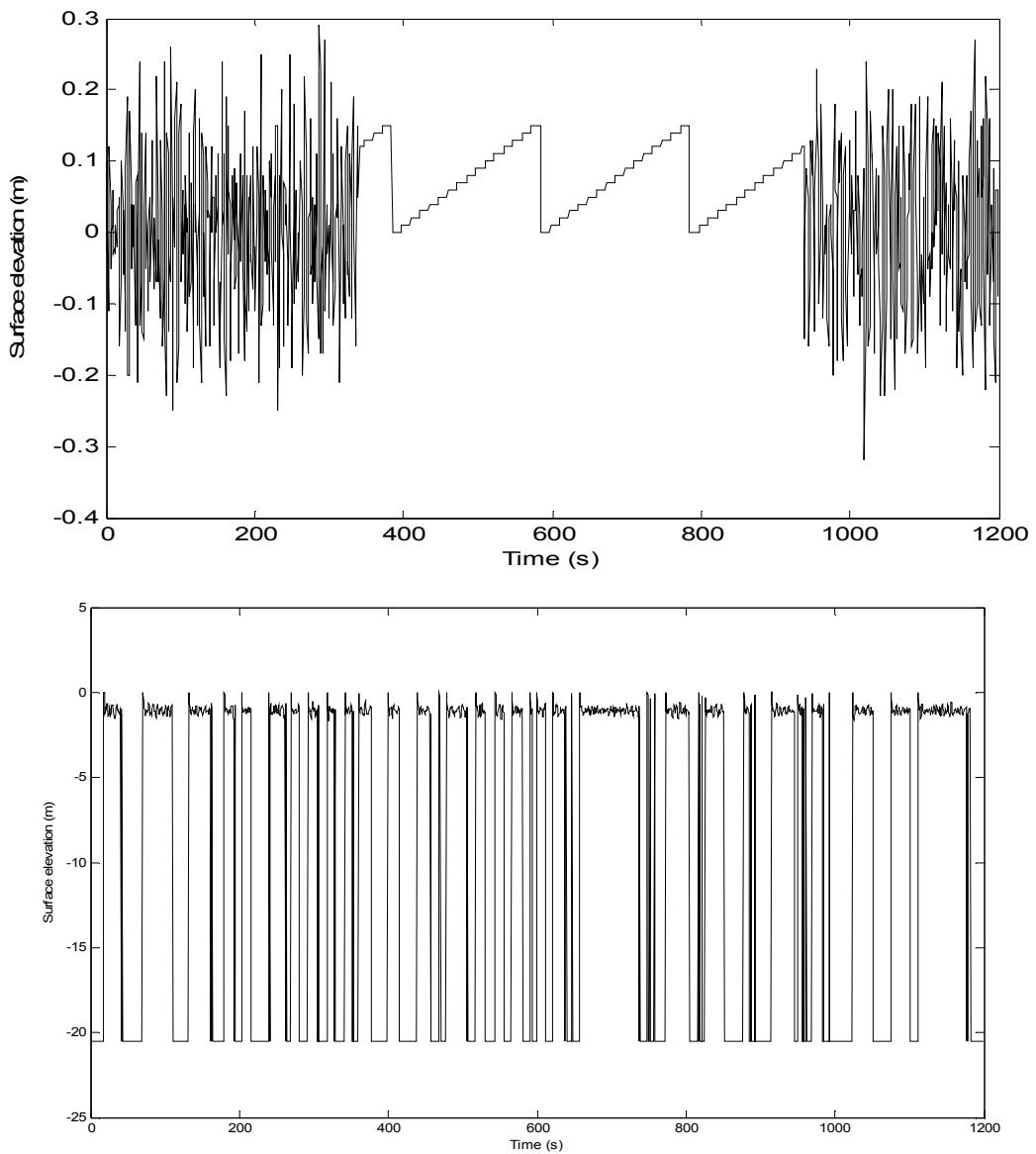


Figure 2.3 Upper panel: Record with a large gap in the middle (Tortosa 14/07/2004 02:00h); lower panel: Record with several gaps of constant value (Roses 30/04/2004 15:20h)

Taking into account the above mentioned definitions, the mathematical criterion in order to detect them is summarized in the Table 2.2 and subsequently explained.

Table 2.2 Conditions for the detection of spikes and gaps.

| | Criteria (details below) |
|--------|--|
| Spikes | S1) The maximum acceleration of the surface elevation higher than half the gravitational acceleration. |
| | S2) The maximum surface elevation higher than 2.83 times the significant wave crest. |
| Gaps | G1) Length of the record (number of data points) differs from the theoretical length. |
| | G2) Twice three consecutive second derivatives smaller than a certain tolerance. |

S1 criterion

For a Stokes corner flow, the downward acceleration of a water particle in the crest is $-0.5g$ (Tucker, 2001). Therefore, in a deep water wave, whose limiting form at the crest is this Stokes flow, when experiencing accelerations higher than half g , wave breaking occurs, i.e. whitecaps (Tayfun, 1981a). In fact, according to Stokes (1847,1880), the necessary criteria for individual wave breaking to begin are when:

- The particle velocity of fluid at the crest equals the phase velocity
- The crest of the wave attains a sharp point with an angle of 120°
- The ratio of wave height to wave length is approximately $1/7$
- The particle acceleration at the crest of the wave equals $0.5g$

For irregular waves Snyder and Kennedy (1983a,b) stated that the stability of the gravity wave flow is controlled by the vertical acceleration. Longuet-Higgins (1985) added that the real acceleration of the fluid particles (measured by a buoy) should be considered instead of the apparent acceleration (measured by a pole at a fixed point).

Although other theories (Ochi and Tsai (1983), Srokosz (1986)) with high order expansion for Stokes' waves results in a lower threshold, the more conservative one (half g) is considered in this study.

The acceleration has been approximated as shown in Eq. (2.1).

$$Accel_{i+1} \cong \frac{\eta_{i+2} + \eta_i - 2\eta_{i+1}}{\Delta t^2} > \frac{1}{2}g \quad \forall i = 1 \dots n-2 \quad (2.1)$$

in which g is the gravitation, n the length of the record, η the surface elevation and Δt the time interval. Therefore, the records with only one value of acceleration higher than $0.5g$ (in absolute value), they are considered unphysical.

S2 criterion

The intention of this requirement is to reject spikes by comparing the maximum wave crest to the significant wave crest (according to linear theory, the wave crest is half the significant wave height), which represents the mean of the highest one third wave crests from the record.

The threshold of 2.83 is not arbitrary. According to studies of freak wave occurrence, a freak wave is defined as a wave with:

$$H \geq 2.83H_s \quad (2.2)$$

which has been found by considering the maximum wave height with an exceedance probability of 0.01 in a storm of 6 h (approx. 2000 waves) and BFI = 0.8 (see Section 6.2.4). In other words, a freak wave occurs once in every 100 hundred such storms. Then, if the threshold is exceeded by three consecutive points (not just one), it would mean a real freak wave instead of a spike and therefore it is not removed. In Appendix B, examples of freak waves are given.

In addition, by using linear theory one can appreciate that the associated probabilities of a threshold of 3 (rounded up from a value of 2.83) are clearly low:

- Probability of the maximum wave height in a record of 100,000 waves exceed $3H_s$ is approx. 0.9985
- Size of the record (number of waves) with an expected maximum wave height of $3H_s$: 10^8 waves approx.

To have an impression of what 100,000 waves mean, note that a wave record of 20 min has approximately 300 waves and between 1990 and 2003 in Tortosa the mean duration of a storm was 0,9006 days (Rotés, 2004), which represents approximately 20,000 waves. Rotés (2004) defined a storm as an event with the significant wave height per record higher than 1.50 m and with a minimum duration of 6 h.

G1 criterion

Once the duration D and the time interval are known, the theoretical length of the surface elevation's vector (in terms of number of data points of the record) can be defined as (see Table 2.3):

$$\text{Length} = D/\Delta t \quad (2.3)$$

Table 2.3 Duration, sample interval and consequent length of buoys' records

| Buoy | Roses | Blanes | Llobregat | Tortosa |
|-----------------------|--------|--------|-----------|---------|
| Duration (min) | 20 | 20 | 20 | 20 |
| Sampling interval (s) | 1/2.56 | 1/2.56 | 1/2.56 | 1/1.28 |
| Length | 3072 | 3072 | 3072 | 1536 |

However, in the period 2001-2006, the Tortosa buoy record length is 1535: one element is missing, possibly due to a computing error. In general, this does not present any particular anomaly compared to the others and, therefore, instead of rejecting such an amount of data, an exception has been made in these cases, considering the same Δt but a slightly shorter duration (in order to fulfil Eq. (2.3)).

G2 criterion

This last criterion is empirical. The idea is to remove the data with irregularities like the examples shown in Figure 2.3. Such irregularities appear to be constant or combinations of straight lines. Therefore, the easiest way to detect them is to consider the second derivative (which is approximated in the same manner as in the first criterion). If it

becomes null in many consecutive points in time, then the recorded data is not considered to be a wave profile and some type of anomaly is present. The value of 3 consecutive second derivatives (and not 2, for example) is somewhat arbitrary. One does not want to be so strict and subsequently reject a very steep wave which has a short linear stretch. The same comment could be made for considering twice consecutive second derivatives.

2.2.3. Buoy limitations

Buoy signal trend

As has been mentioned, the linear trend of the buoy is systematically removed from all records by subtracting the fitted straight line obtained by the least square method.

Aliasing

One of the most important parts of the data analysis is the spectral analysis. As it will be commented upon in Chapter 4, the fact of having discrete (and not continuous) time series produces the aliasing phenomenon. The spectrum is mirrored around the Nyquist frequency and therefore the obtained spectral energy at nearby frequencies is distorted. Hence, if the mean frequency is close to the Nyquist frequency, the obtained spectrum with the Fourier analysis is not reliable. It has been considered that the Nyquist frequency/mean frequency ratio is not acceptable if it is lower than 2.2:

$$\frac{f_N}{f_m} < 2.2 \quad (2.4)$$

where the mean frequency has been calculated as the inverse of the mean period as in Eq. (4.21).

There is no theoretical explanation about choosing exactly 2.2 but it is empirically reasonable. In fact, higher values are sometimes recommended. Holthuijsen (2007) suggested that, in general, the Nyquist frequency should be higher than 4 times the mean frequency. In the present study, the threshold has been readjusted and a less conservative value has been used. This criterion is more or less equivalent to considering $f_N/f_p > 3$ in which f_p is the peak frequency.

This condition basically causes the rejection of some records from the Tortosa buoy, which has a higher sampling interval time compared to the other buoys off the Catalan coast and also to the common value in sea measurements of 0.5 s.

Bump

The buoy is at open sea and may be hit by some object (e.g. a ship). This produces a long period oscillation to the buoy signal which is added to the surface elevation. Such a phenomenon can be detected by looking at the spectral energy at the zero frequency, which, theoretically, should be zero. However, owing to the averaging process in the spectrum calculation (see Chapter 4), a little energy can be found. A threshold of 0.004 m²/Hz has been found acceptable, very similar to the one used by Rotés (2004).

$$E(f \approx 0) > 0.004 \text{ m}^2 / \text{Hz} \quad (2.5)$$

Note that in Eq. (2.5), the frequency at which the spectrum is evaluated is not exactly zero since the first data point of the calculated spectrum corresponds to half the frequency

interval. As an example, some records from the Roses buoy in March 2005 are rejected due to this criterion. They are consecutive in time, suggesting that the buoy spends one or two days until becoming stable again. In Figure 2.4 another example of such distortion in buoy registration. One can discern a long period of oscillation superposed over shorter real ocean waves. This artificial oscillation produces the enhancement of energy at $f \approx 0$ Hz.

Maybe, other reasons cause such strange shapes for the spectrum as in Figure 2.4. In any case, high energies for very low frequencies are physically unacceptable for ocean waves.

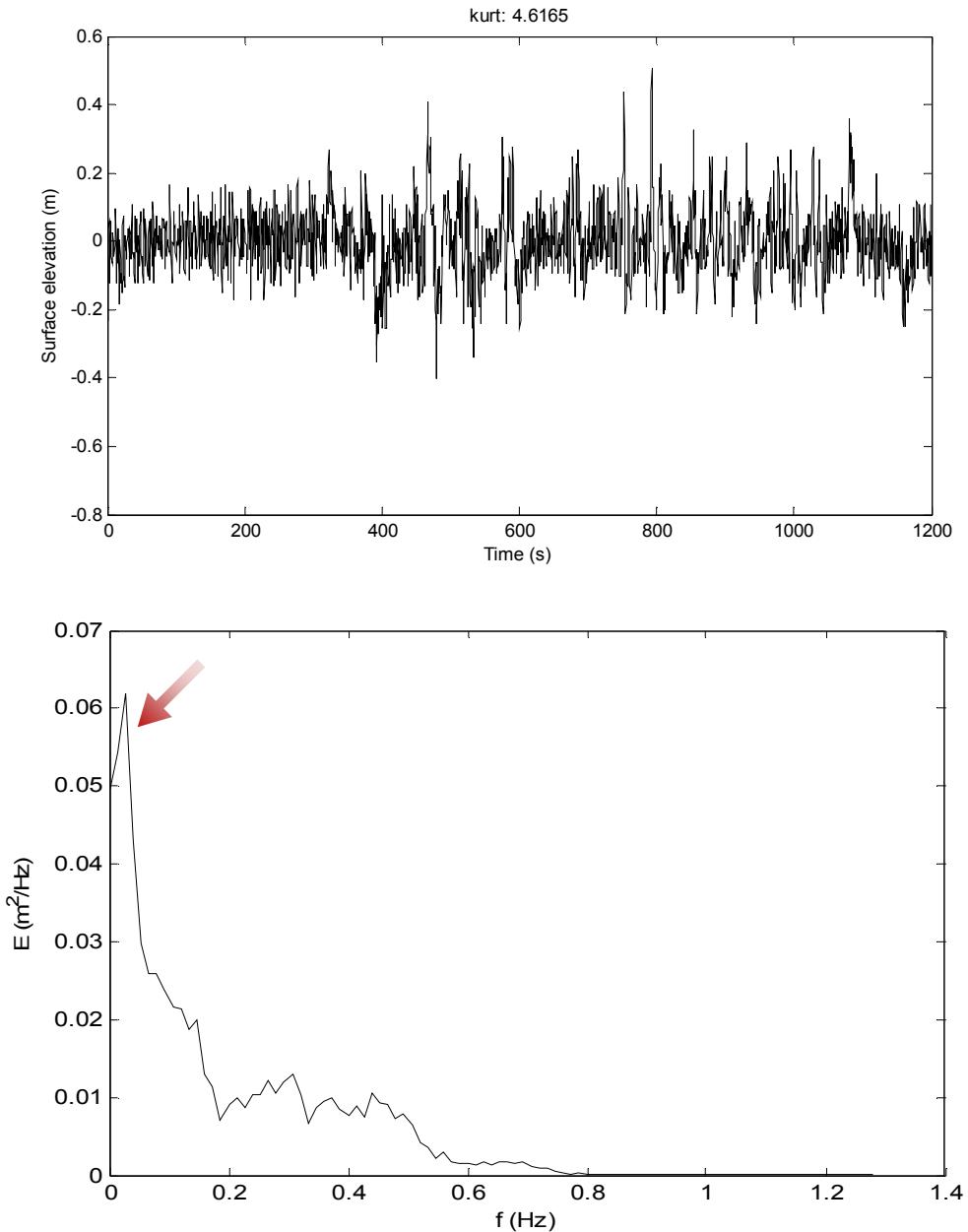


Figure 2.4 Surface elevation and associated spectrum of a distorted record, probably due to a bump against the buoy

Low waves

The buoy does not appear to be capable of properly measuring low wave heights of the same order of magnitude as its diameter. Then, the rejection of data with significant wave height lower than half meter has been carried out. In fact, one of the reasons is that the quantification interval of the buoy (1 cm) becomes proportionally more important in the case of low waves.

In the present data, this restriction produces the loss of a considerable amount of data since such low heights are very common in the Mediterranean Sea. This quality requirement is probably questionable. However, remark that from the engineering point of view, the behaviour of higher wave heights is of greater importance.

2.2.4. Visual check

After the implementation of the above carefully designed criteria, about 100 randomly selected records have been visually checked, looking at the surface elevation itself, the spectrum and the observed exceedance probability. In general, the remaining data seems to be "clean". Special attention has been paid to outliers in all subsequent analysis. In addition, records with high values of skewness and kurtosis of surface elevation have been checked. These two parameters have been found appropriate to detect strange records. As later explained in Chapter 5, the skewness and kurtosis of a Gaussian variable are, respectively, 0 and 3. Records with an absolute value of skewness higher than 0.3 or a value of kurtosis higher than 4, have been visually checked since the deviation from the theoretical values for the Gaussian distribution is suspicious. The following equation summarises such conditions, which are quite arbitrary but they have been found empirically acceptable.

$$\text{kurt} > 4, \quad |\text{ske}| > 0.3 \quad (2.6)$$

About 35 records fulfil with Eq. (2.6) but only 16 have been rejected. Most of them are rejected due to their non-stationary behaviour.

2.2.5. Shallow water

As mentioned in the Introduction, in this study the assumption of deep water is made. In shallow water, there exist other mechanisms which interfere but these are not considered. In principle, the buoys of this study are situated at deep enough sites. However, depending on the wave length some cases can be considered as intermediate or shallow water. The condition of rejection is:

$$h < L_0/2 \quad (2.7)$$

where h is the depth and L_0 the wave length:

$$L_0 = \frac{gT_0}{2\pi} \quad (2.8)$$

and T_0 the mean zero crossing period (see Chapter 4, Eq. (4.20)).

Most of the data from the Catalan coast has been collected from deep water except for a few cases (49 records in total) in the buoys of Roses, Blanes and Llobregat, whose depths are lower compared to the Tortosa one.

2.2.6. Comments

Although possible irregularities were initially considered, subsequent data analysis surprisingly revealed other types of peculiar records. Therefore the design process of the criteria explained above has been iterative and adapted to the results found.

Table 2.4 summarizes the number of rejected data for each buoy. For more detailed information see Appendix A. The total amount of data has then been reduced to 30% (an especially substantial reduction in the buoys from Roses and Tordera) which means a very high percentage of irregular records. Perhaps, the conditions could be readjusted so that some of the rejected data could be used. Moreover, some of the removed records could also be repaired. For example, in the case of a bump, one could remove the energy of low frequencies and then compute (with the inverse FFT) the correct time record of surface elevation without the artificial oscillation. However, the amount of remaining data (42,000 records approx.) is still sufficient for the present study.

Table 2.4 Accepted data of each buoy

| Number of records | Roses | Blanes | Llobregat | Tortosa | TOTAL |
|-------------------|--------|--------|-----------|---------|---------------|
| Initial | 36,256 | 36,823 | 17,596 | 44,708 | 135,383 |
| Final | 7,141 | 12,701 | 4,215 | 18,320 | 42,377 |
| Percentage | 20 % | 34 % | 24 % | 41 % | 32 % |

2.3. North Sea data

The Norwegian data analysed in the present study forms part of the WADIC project (Allender et. al, 1989) whose objective was the evaluation of commercially available directional wave measurement systems under severe open ocean wave conditions. The main data base in that project were comprised of data from October 1985 to January 1986, which included in excess of 100 million individual data values from 20 independent wave measurements systems. In the present study, only 92 time series registered in November-December 1985 by two (EMI) laser altimeters (from a pentagon array) are used. They are mounted on the Phillips Edda platform in the North Sea (see Figure 2.5). In Table 2.5, more features about the data are detailed, such as water depth and sample interval time.

Table 2.5 Description of the laser altimeters and their measurements

| Parameters | Description |
|------------------------------------|---------------------------------|
| Coordinates | 56 28 N 03 28 E |
| Depth (m) | 70 |
| Sample interval (s) | 1 |
| Resolution | mm |
| Record duration (min) ² | 17,07 |
| Recorded dates: | 5-6/11/1985 21-22-23/12/1985 |
| File extension | Text file |

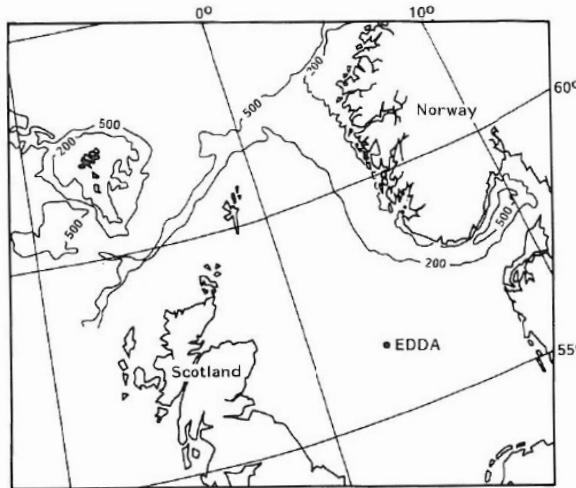


Figure 2.5 The location of the WADIC experiment, near the EDDA platform (Allender et. al., 1989)

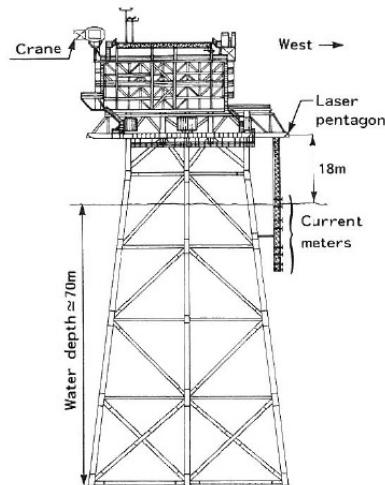


Figure 2.6 Instrument locations on the Eddystone platform (Allender et al., 1989)

² The duration is not a whole number because 1024 data points were used, which, with a sample interval time of 1Hz, is equivalent to a duration of 17 min 4 s.

The EMI laser is a pulsed laser range finder operating in the near-infrared region. Narrow pulses of light are produced by a laser diode, and the radiation from the target is used to take a time interval measurement. The time of travel is converted to an analogue voltage proportional to the distance between it and the reflector. Laser altimeters and other various sensors are mounted on the instrument tower. Their signals are cabled to a central data logging station.

The data represents two independent storms, one of them with the significant wave height greater than 10 m. These 92 records have not needed to be filtered through a quality control since they were used in the WADIC project after recovery of 87% of the total initial raw data (including the data of the other devices). Only the condition of shallow water is used. This requirement causes the rejection of 23 records, precisely being the ones of the highest wave height in November 1985. Therefore, the total amount of analysed records is 69, representing approx. 9000 waves, with a maximum significant wave height of about 9 m.

3. STATISTICAL ANALYSIS

3.1. Introduction

The statistical analysis in the present study basically comprises of the calculation of representative parameters of the surface elevation and wave height (crest and trough) directly obtained from the surface buoy registration. The main objective is the comparison with the ones obtained from the spectral analysis in the framework of different theories. Moreover, the analysis of some parameters itself is useful to determine the applicability of the well-known linear theory, explained in Chapter 5.

3.2. Definitions of height, crest, trough and period

First of all, the concept of wave has to be defined. In the case of the linear theory or a sinusoidal wave, it is easy but when one looks at real ocean surface elevation it becomes more complicated. For a time record, two main methods are used to describe a wave: the downward zero-crossing (see Figure 3.1) and the upward zero-crossing. If the surface elevation is considered as a Gaussian process, the definition used does not matter for the statistics. Nevertheless, many prefer the downward zero-crossing because in visual estimates the height is taken between the crest height and the preceding trough. In addition, the steep front, which is relevant for the breaking process, is included in the downward zero-crossing definition. In fact, this criterion is recommended by the International Association for Hydraulic Research Working Group (IAHR, 1989). Attending to these recommendations, in the present study the downward zero-crossing method is used.

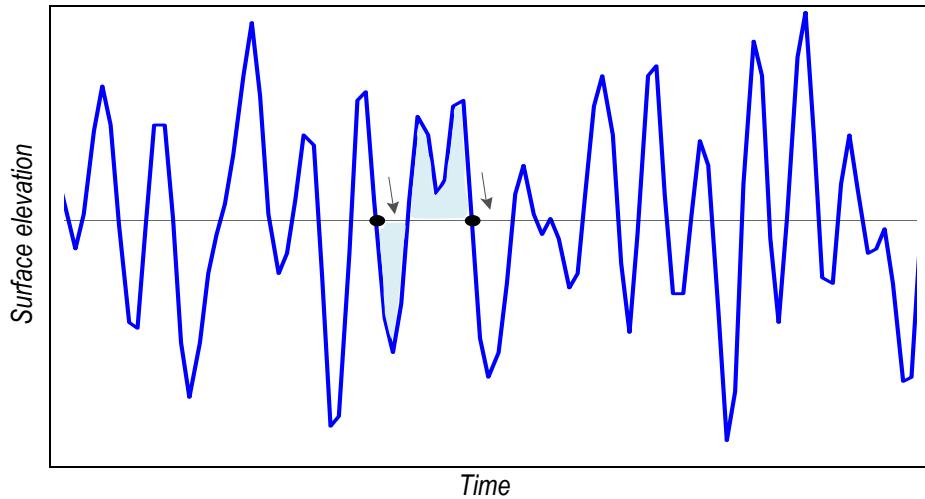


Figure 3.1 Sketch of the definition of a wave in a time record of the surface elevation with downward zero crossing

For each wave the following parameters are calculated: wave height, crest level, trough level and period (see Figure 3.2). For the calculation of the period a linear interpolation is made between the two points around the zero level. Note that the definition of crest/trough is not a local maximum/minimum surface elevation but the maximum/minimum per wave. As a consequence, the wave height is the difference between such a maximum and minimum. When reading studies in the literature, one must be careful because sometimes one is referring to the local maxima which, for

instance, would mean the possibility of having negative wave crests. For civil engineering purposes, the definition used in the present study is more useful, which does not allow negative values for wave crests.

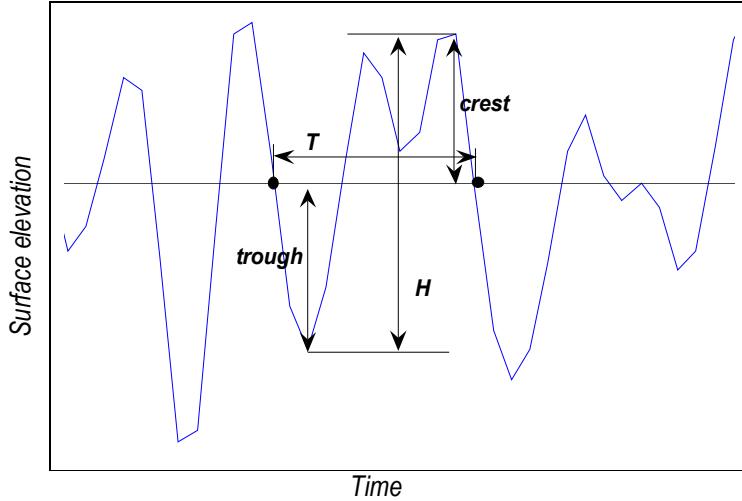


Figure 3.2 Definition of the main parameters of each wave

3.3. Parameters

Once the above mentioned calculations are made for all the waves, the following parameters are calculated per record:

- Mean wave height (H_{mean})

$$H_{mean} = \bar{H} = \frac{1}{N} \sum_{j=1}^N H_j \quad (3.1)$$

- Significant wave height ($H_{1/3}$): mean of the highest one third wave heights

$$H_{1/3} = \overline{H_i} = \frac{1}{N/3} \sum_{i=1}^{N/3} H_i, \quad H_i \text{ being highest one third wave heights} \quad (3.2)$$

- Root-mean-square wave height (H_{rms})

$$H_{rms} = \sqrt{\bar{H}^2} = \sqrt{\frac{1}{N} \sum_{j=1}^N H_j^2}, \quad (3.3)$$

- Maximum wave height (H_{max})

$$H_{max} = \max(H) \quad (3.4)$$

- Mean wave period (T_{mean})

$$T_{mean} = \bar{T} = \frac{1}{N} \sum_{j=1}^N T_j \quad (3.5)$$

- Significant wave period ($T_{1/3}$): mean of the period of the highest one third waves.

$$T_{1/3} = \bar{T}_i = \frac{1}{N} \sum_{i=1}^{N/3} T_i , \quad T_i \text{ being the periods of the highest one third wave heights} \quad (3.6)$$

The crest and the trough statistics are also calculated in an analogous procedure as in the wave height.

For quantification interval and noise considerations, waves with wave height smaller than 5 cm or crest heights smaller than 2.5 cm or wave periods smaller than twice the sampling interval, are not considered in the statistical analysis.

In addition, the standard deviation, skewness and kurtosis parameters are calculated. These are defined in Eq. (3.7), (3.8) and (3.9)

$$\text{Standard deviation: } \sigma = \sqrt{E\{(\eta - \mu_\eta)^2\}} \quad (3.7)$$

$$\text{Skewness: } s = \frac{E\{(\eta - \mu_\eta)^3\}}{\sigma^3} \quad (3.8)$$

$$\text{Kurtosis: } k = \frac{E\{(\eta - \mu_\eta)^4\}}{\sigma^4} \quad (3.9)$$

4. SPECTRAL ANALYSIS

4.1. Introduction

The spectral analysis is widely used in the analysis of noise-like signals because it provides a frequency decomposition in harmonics the behaviour of which can be studied separately. For that reason, it has become more important than the pure statistical analysis of the surface elevation itself and therefore is quite important to find a probability distribution which depends on spectral parameters which can be predicted.

Different methods exist in order to determine the spectral density function from a discrete time record. The Fast Fourier Transform (FFT), which is an algorithm for calculating the Discrete Fourier Transform (DFT), is the most used.

4.2. Fourier theory

4.2.1. Continuous function

Joseph Fourier (1768-1830) demonstrated that almost any function can be represented as a linear combination of an infinite number of harmonic oscillations. Therefore, the surface elevation, with a zero mean level, can be written as:

$$\eta(t) = \sum_{i=1}^{\infty} a_i \cos(2\pi f_i t + \alpha_i) \quad (4.1)$$

where $\eta(t)$ is the surface elevation, f_i the frequency of the i -harmonic and α the phase.

Using trigonometric identities the last expression can be written as:

$$\eta(t) = \sum_{i=1}^{\infty} [A_i \cos(2\pi f_i t) + B_i \sin(2\pi f_i t)] \quad (4.2)$$

$$\text{with } a_i = \sqrt{A_i^2 + B_i^2} \text{ and } \tan \alpha_i = -\frac{B_i}{A_i}$$

For a record of duration D , the amplitudes A_i and B_i can be determined with the *Fourier Integrals*:

$$A_i = \frac{2}{D} \int_D \eta(t) \cos(2\pi f_i t) dt \text{ with } f_i = \frac{i}{D} \quad (4.3)$$

$$B_i = \frac{2}{D} \int_D \eta(t) \sin(2\pi f_i t) dt \text{ with } f_i = \frac{i}{D} \quad (4.4)$$

The notion of a Fourier series can also be extended to a complex function:

$$\eta(t) = \sum_{n=-\infty}^{\infty} X_n \exp\left(\frac{i2\pi n t}{D}\right) \quad (4.5)$$

where X_n denotes the complex amplitude.

4.2.2. Discrete function

In practice, the surface elevation is recorded at discrete moments in time. Then, the *Fourier Integrals* of Eq. (4.3) and (4.4) become sums:

$$X_n = \frac{1}{N} \sum_{j=0}^{N-1} \eta_j \exp\left(-\frac{i2\pi j n}{N}\right) \quad (4.6)$$

where N is the length of the discrete time series.

From Eq. (4.6) the real amplitudes a_n can be calculated as:

$$a_n = 2\sqrt{\operatorname{Re}(X_n)^2 + \operatorname{Im}(X_n)^2} \quad (4.7)$$

4.3. The wave spectrum

4.3.1. Theoretical definition

From the amplitudes a_i and phases α_i associated with a certain f_i , the amplitude and phase spectrum can be determined. In deep water, linear theory is assumed. The phase is Uniformly distributed between 0 and 2π (see Chapter 5). The amplitude is considered as a random variable too. To remove the corresponding sample character of the estimated spectrum one option would be considering \bar{a}_i (from different records of surface elevation under statistically identical conditions). Instead, rather than the amplitude, the variance density is considered. The main reasons for this consideration are two. In the first place, $\frac{1}{2} \bar{a}_i^2$ is a measure of the variance

(see Chapter 5) and therefore is a more representative statistical parameter of the surface elevation and it is, according to linear theory, proportional to the energy of the waves. In the second place, the variance spectrum is discrete (only the frequencies f_i are present) but, in fact, all the frequencies are present in real waves. To account for this, the variance density $\frac{1}{2} \bar{a}_i^2 / \Delta f$ is defined for the interval $\Delta f = 1/D$. Finally, because of the jumps from one frequency band to the next, the following limit is taken (in which the averaged value is replaced by the expected value):

$$E(f) = \lim_{\Delta f \rightarrow 0} \frac{1}{\Delta f} E\left\{ \frac{1}{2} \bar{a}^2 \right\} \quad (4.8)$$

4.3.2. Practical limitations

First of all, the theoretical limit of Eq. (4.8) cannot be taken because of the finite duration of the record (taking the limit $\Delta f \rightarrow 0$ implies an infinite duration). That leads to a finite frequency resolution, removing details at a frequency scale $\Delta f = 1/D$. The frequency resolution can be improved by using a longer duration. However, the duration cannot be very long; otherwise, the record would not be stationary.

If one considers the amplitudes obtained from a single record (a_i^2 instead of $\overline{a_i^2}$), one makes an error of approximately 100% in the estimation of a spectrum (Holthuijsen, 2007) because of the estimation of the mean from a single value. Such an error is unacceptable. In fact, with the assumption of the linear theory (see Chapter 5), the estimated variance density is χ^2 distributed with 2 degrees of freedom. This is logical; the spectrum is proportional to the square amplitude, which, according to linear theory, is Rayleigh distributed.

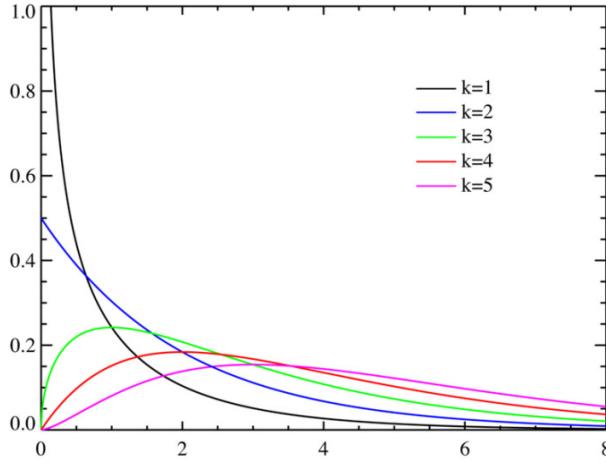


Figure 4.1 Chi-square probability density function for different degrees of freedom

A reasonable solution is dividing the initial record into p sub-records and to compute the variance density for each one. Such sub-records can be considered statistically identical and therefore one can calculate the average between their spectra. By this manipulation the variance density spectrum is χ^2 distributed with $2p$ degrees of freedom and the error is reduced by a factor \sqrt{p} :

$$\text{Error}(\%) \approx \frac{100}{\sqrt{p}} \quad (4.9)$$

The price paid for the reduction of this error is the decrease of the spectrum resolution because the new frequency interval is $\delta f = p\Delta f$. In conclusion, the choice of p may be made considering both the error and the resolution. It is recommendable to have a final resolution of about 0.01 Hz (for the Mediterranean data would imply $p=12$). However, a slightly higher factor has been considered in order to decrease the error.

Table 4.1 Parameters of the averaged spectra

| Parameter | Mediterranean Sea | North Sea |
|-----------------|----------------------|----------------------|
| p | 16 | 16 |
| Δf (Hz) | $8.33 \cdot 10^{-4}$ | $9.77 \cdot 10^{-4}$ |
| δf (Hz) | 0.013 | 0.016 |
| Error (%) | 25 | 25 |

Figure 4.2 qualitatively illustrates the influence of the variation of the parameter p . Logically, for smaller values of p , the spectrum looks more “grassy”. The chosen value $p = 16$ seems quite reasonable, avoiding such fluctuations.

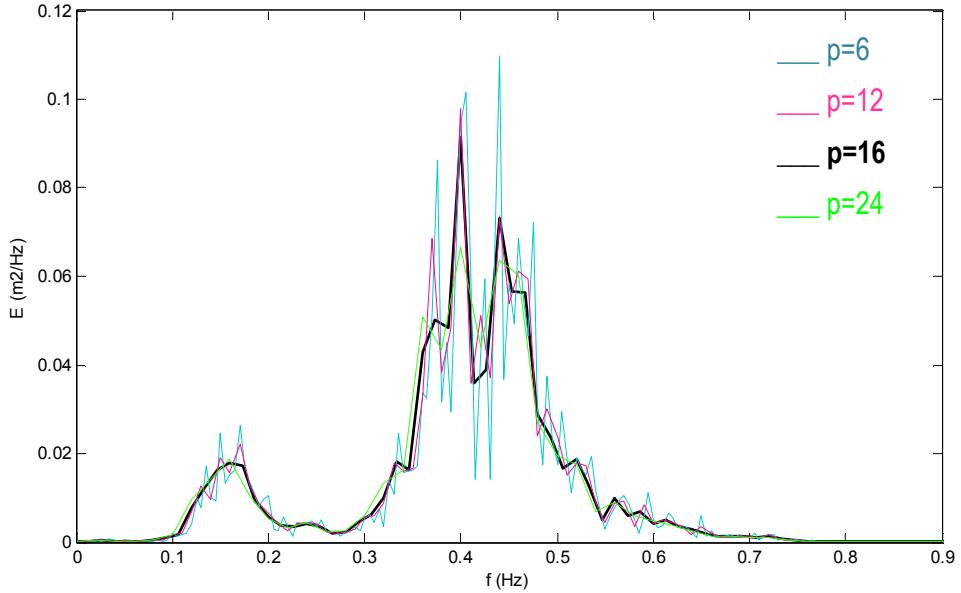


Figure 4.2 Comparison of the variance density spectrum with different values of p

In addition, the discrete character of the wave record introduces an error which is not so obvious. As illustrated in Section 4.2.2, the Fourier integrals are replaced by finite sums. The consequence is a phenomenon called aliasing, which consists of mirroring the energy of high frequencies around the so-called Nyquist frequency the value of which is:

$$f_N = \frac{1}{2\Delta t} \quad (4.10)$$

in which Δt is the sampling interval of the wave record. The “physical” reason is that in a discretized time record, two harmonic waves with different frequencies may pass through the same data points (see Figure 4.3). The aliasing phenomenon is caused by the inability to distinguish between them.

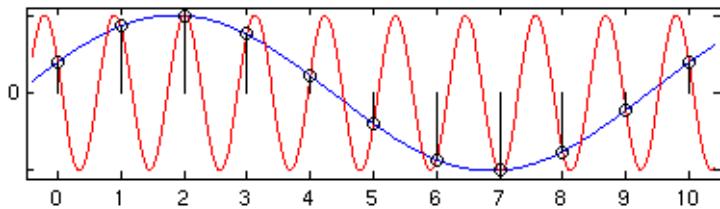


Figure 4.3 Two harmonic waves with frequencies f_1 and f_2 , given at a interval time of $\Delta t = 1 / (f_1 + f_2)$ are indistinguishable.

In outline, the derivation of expression of Eq. (4.10) can be explained by Figure 4.4: the discrete spectrum is not the Fourier transform of the surface elevation (understanding it as the continuous surface elevation) but the transform of the product of the surface elevation by an

equally spaced delta series. The spectrum is therefore the convolution product of the true spectrum by a delta series with $1/\Delta t$ frequency interval, which leads to a mirroring effect and the superposition of energy at the frequencies near the Nyquist frequency.

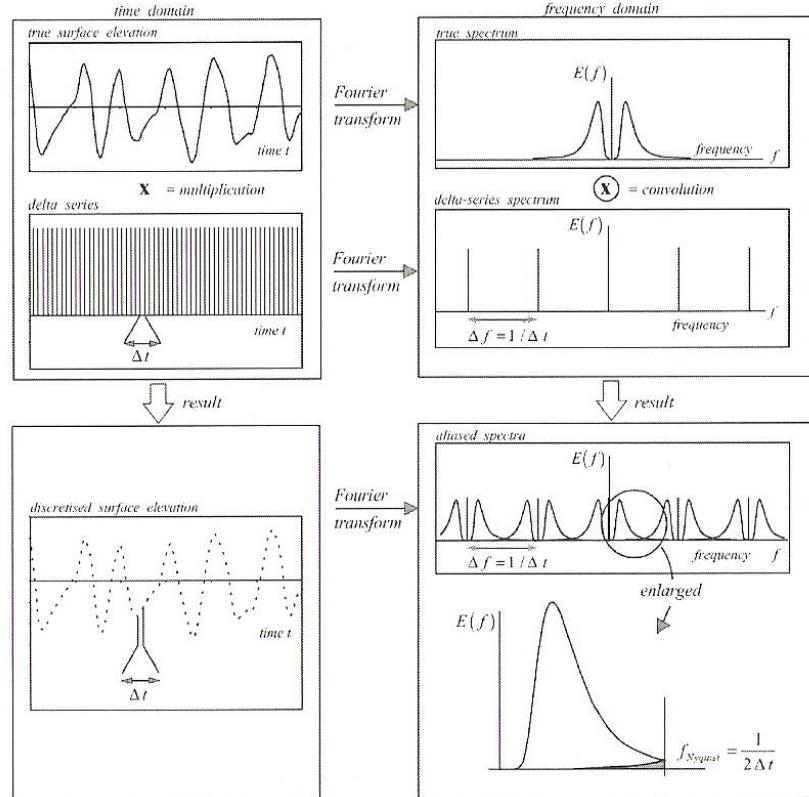


Figure 4.4 Sketch of the reason for the aliasing phenomenon (Holthuijsen, 2007)

The sampling interval of the Tortosa buoy is slightly higher than the commonly used value of 0.5 s. The reason for choosing, if possible, an interval of 0.5 s is that the Nyquist frequency becomes 1Hz. Therefore, for real sea waves, the aliasing effect on the main part of the spectrum is practically null. The problem arises when the spectrum is clearly bimodal due to the presence of sea and swell. In such a case, the energy of the higher frequencies of the storm waves may be added to the energy of the swell lower frequencies if the Nyquist frequency is similar to the one of the storm waves. In any case, the frequency domain of the calculated wave spectrum should be up-limited to the Nyquist frequency. For the analysed data the value of the Nyquist frequency is:

Table 4.2 Time interval and Nyquist frequency

| Parameter | Tortosa buoy | Other buoys | Lasers |
|----------------|--------------|-------------|--------|
| Δt (s) | 1/1.28 | 1/2.56 | 1 |
| f_N (Hz) | 0.64 | 1.28 | 0.5 |

Moreover, a minimum frequency has to be considered too. The sensor of the buoy is not capable of properly measuring the surface elevation below a certain value due to changes in temperature, pressure, etc. The used lower value for the frequency is 1/D (Rotés, 2004).

4.3.3. Windowing

It has been seen that a record of length D can be mathematically described as the superposition of a number of sinusoidal waves, each having an exact number of periods in duration D ($\Delta f = 1/D$). Nevertheless, the sea actually consists of a continuous spectrum of waves. The energy of a component wavelength which does not have the exact frequency of a harmonic of D , goes to the two nearest harmonics and a portion goes to more distant ones. This leaked energy is sometimes significant and may interfere with the results. This phenomenon can be explained by considering the finite signal as the product of the entire infinite signal and a rectangular function which equals one in the duration and zero outside the duration. Therefore, the spectrum of the finite signal is the convolution product of the real spectrum and the Fourier transform of the rectangular function. In Figure 4.5, the simple case of a sinus wave is illustrated. With a duration that is a multiple of the wave length, the estimated spectrum is not affected, contrary to the case in which the duration is not a multiple of the wave length. The transform of the rectangular function appears in both cases but in the first one it does not have any affect because of those points at which the spectrum is evaluated.

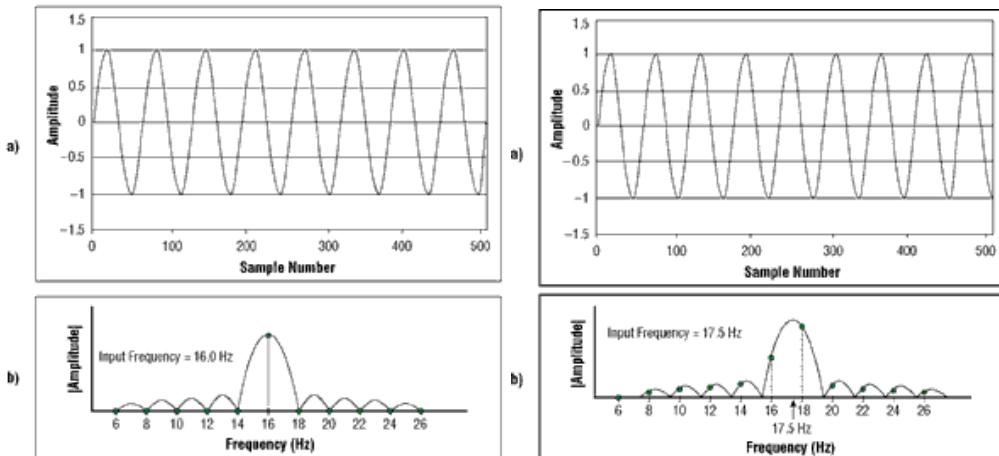


Figure 4.5 Comparison of the amplitude spectrum (logarithmic scale) between a sinus wave with a wave period a fraction of the duration (left) and any other period (right) (Lyons, 1998)

The purpose of window functions (also known as taper functions) is to reduce the above leakage of energy to other frequencies, often visible as side lobes (see Figure 4.6). They consist in multiplying the wave record by a certain function, attenuating the value of surface elevation at the ends of the record. Most of the common window functions, depending on the type of the used function, are Hamming, Hanning, Tukey (partial cosine taper), etc.

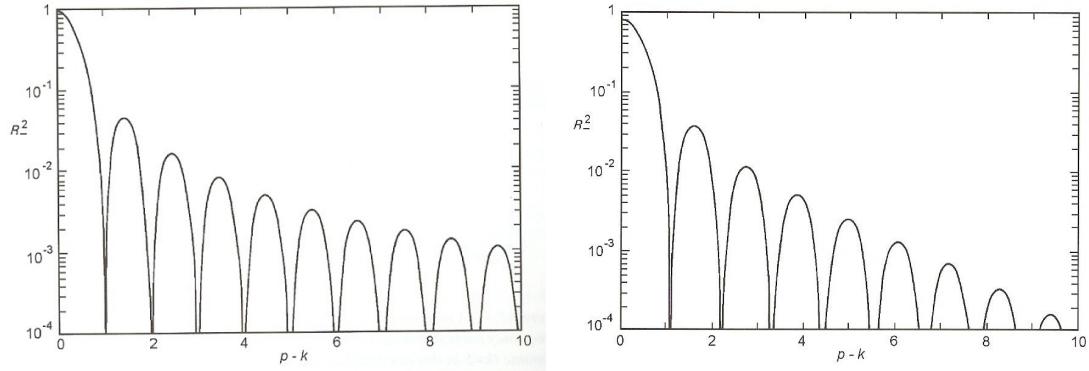


Figure 4.6 Amplitude response spectrum (logarithmic scale) of : rectangular function (left) and partial cosine taper carried out over the first and last 10% of the record (Tucker & Pitt, 2001)

One of the disadvantages of window functions is that the beginning and end of the signal is attenuated. For that reason, the Tukey window has been considered in the present study because it is the one that least affects the original data. It consists of multiplying the extremes of the signal by a cosine wave. One example is illustrated in Figure 4.7.

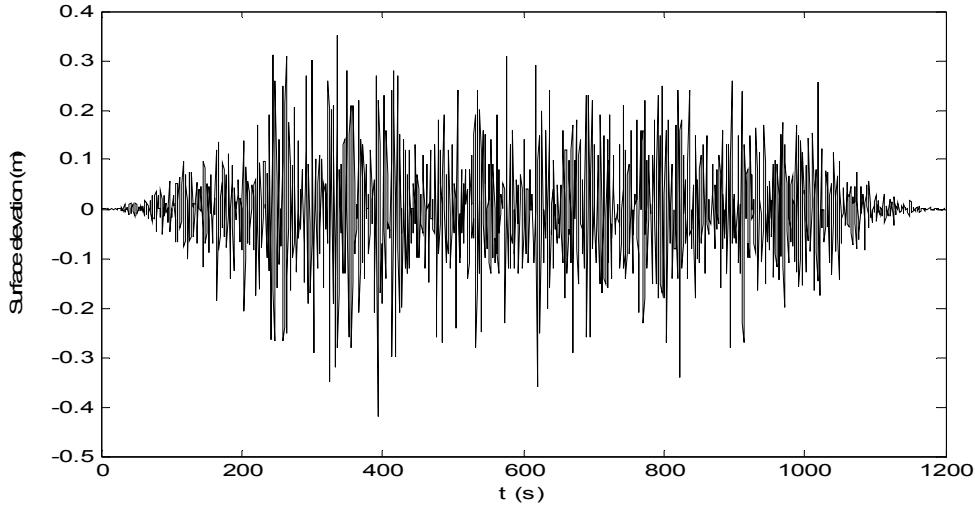


Figure 4.7 Example of a 50% cosine tapered record (25% at each end)

The total tapered length is 10%, 5% at each end. However, as one can appreciate from Figure 4.8, the spectra obtained by different percentages of tapering are practically equal.

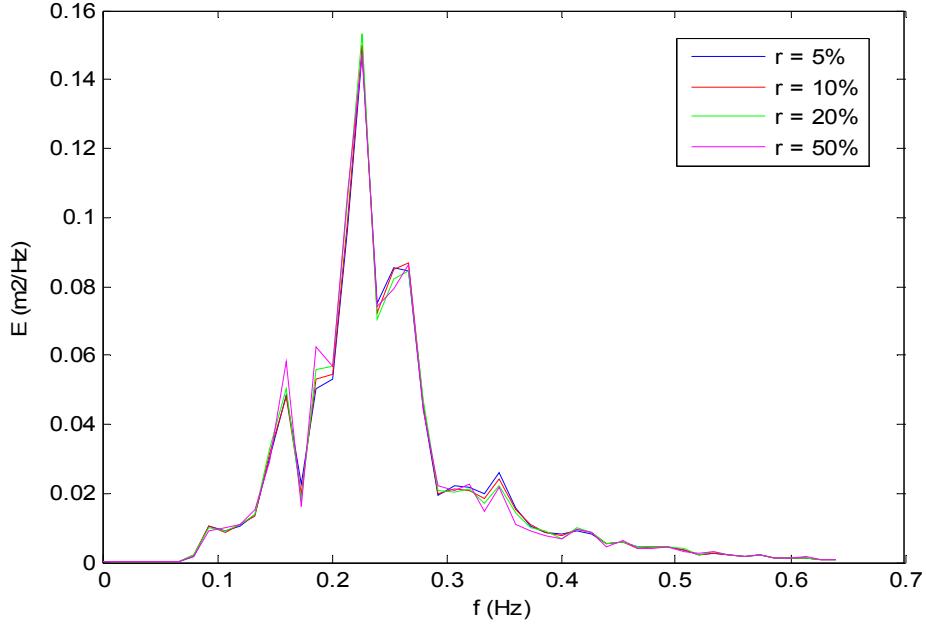


Figure 4.8 Spectrum of a partial cosine tapered record with different percentages of tapering.

The used function $f(k)$ in the Tukey window (implemented in MATLAB) is:

$$f(k) = \begin{cases} \frac{1}{2} \left[1 + \cos \left(\frac{2\pi}{r} \frac{(k-1)}{(N-1)} - \pi \right) \right] & k < \frac{r}{2}(N-1) + 1 \\ 1 & \frac{r}{2}(N-1) + 1 \leq k \leq N - \frac{r}{2}(N-1), \\ \frac{1}{2} \left[1 + \cos \left(\frac{2\pi}{r} - \frac{2\pi}{r} \frac{(k-1)}{(N-1)} - \pi \right) \right] & k > N - \frac{r}{2}(N-1) \end{cases} \quad (4.11)$$

for $k = 1 \dots N$

in which k is the index for each data point, r is the proportion of tapering (0.1, 0.05 at each end), and N the record length. The window multiplies the variance of the record by a factor:

$$G = 1 - 5r/8 \quad (4.12)$$

Therefore, the final spectrum has to be divided by the above mentioned factor of Eq. (4.12).

In Figure 4.9, the effects of both aliasing (due to discretisation of a signal) and leakage (due to truncation) are illustrated.

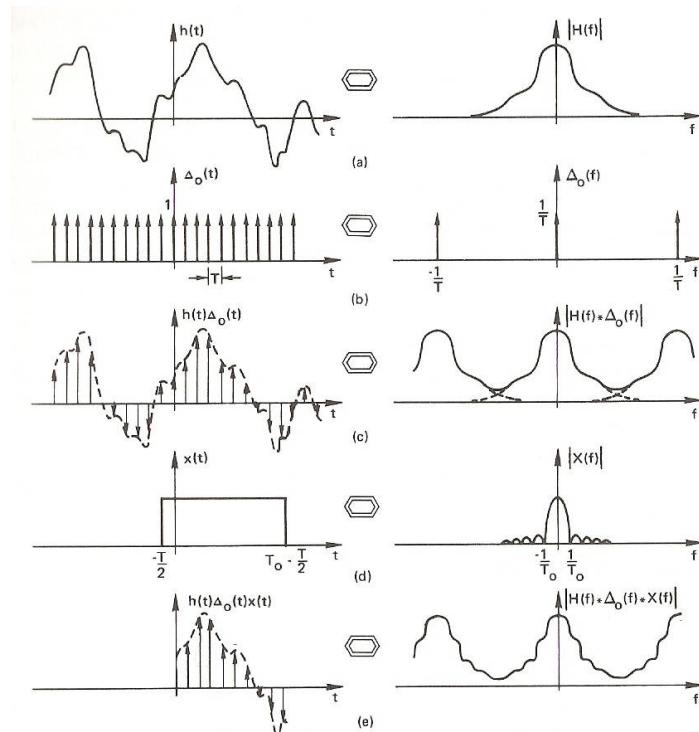


Figure 4.9 Aliasing (due to discrete record) and leakage (due to finite duration) (Brigham, 1988)

4.4. Wave spectrum calculation

In the present study, the FFT algorithm of MATLAB has been used, which implements the Fourier Transform of Eq. (4.6) (multiplied by a scale factor N). The summation differs slightly from Eq. (4.6) because the numeration begins with 1.

$$X_n = \sum_{j=1}^N \eta_j \exp\left(\frac{-i2\pi(j-1)(n-1)}{N}\right) \quad (4.13)$$

As previously said, the analysis is organized by year-buoy. Each one is organized into a matrix with dimensions: length of the record x number of records. For example, in the Mediterranean data, the number of the analysed matrix totals 27 (number of year-buoys). The FFT is computed for each matrix, in which the Fourier Transform is calculated for each column.

The calculation of Eq. (4.13) for all n -frequencies can be interpreted as the product of a matrix (the exponential part) multiplied by vector (the surface elevation). The efficiency of the FFT lies in the proper decomposition of such a matrix and other relevant permutations that considerably reduce the number of operations needed (for more detail information consult Brigham, 1988). The possible decompositions depend on the length of the vector. Therefore, the execution time and the required memory for the FFT (directly related to the required number of operations) depend on the length (in terms of number of data points) of η . The FFT is fastest for records in which the number of data points is a power of two and almost as fast for lengths that have only small prime factors. As the prime factors are smaller, the matrix can be further decomposed and a fewer number of operations are required.

This aspect may not seem relevant but it is. As the amount of data to be analysed increases, one has to consider these limitations. In fact, without taking into account the considerations explained below, in some "years-buoy" the FFT could not be computed in one step due to memory problems.

MATLAB offers the option of using a function called FFTW which optimises the FFT. This optimization is related to the chosen decomposition of the original length. Bear in mind that for the same length, the required number of operations varies, depending on the factorization used. For example, in a simple case of a length of 16, the base-4 algorithm requires approximately 30 % fewer multiplications than the base-2 algorithm (Brigham, 1988). The original lengths of the records are:

Table 4.3 Original length of the records

| Parameter | Tortosa (1991-2000) | Tortosa (2001-2006) | Other buoys | Lasers |
|-----------|------------------------|------------------------|-------------|--------|
| Length | 1536 | 1535 | 3072 | 1024 |

Except in the case of Tortosa (2001-2006), the prime factors involved are small: 1536 ($=2^9 \cdot 3$), 3072 ($=2^{10} \cdot 3$) and 1024 ($=2^{10}$). However, the length of 1535 ($5 \cdot 307$) cannot be decomposed into small prime numbers. For these records the zero padding technique is used. It consists of adding zeros until the desired length is achieved. In this case, only one zero is added in order to have a length of 1536. The only effect is the scaling of the results of Eq. (4.6). However, as the FFT implemented in MATLAB does not include the factor N (see Eq. (4.13)), the results are not scaled. The spectrum therefore is obtained by considering the original length and the original frequency interval for the amplitude calculation and no scaling factor is required.

Remark that by adding zeros the resolution is not increased (although the frequency band becomes smaller). Essentially, this is an interpolation procedure (Shiavi, 1991) (see Figure 4.10).

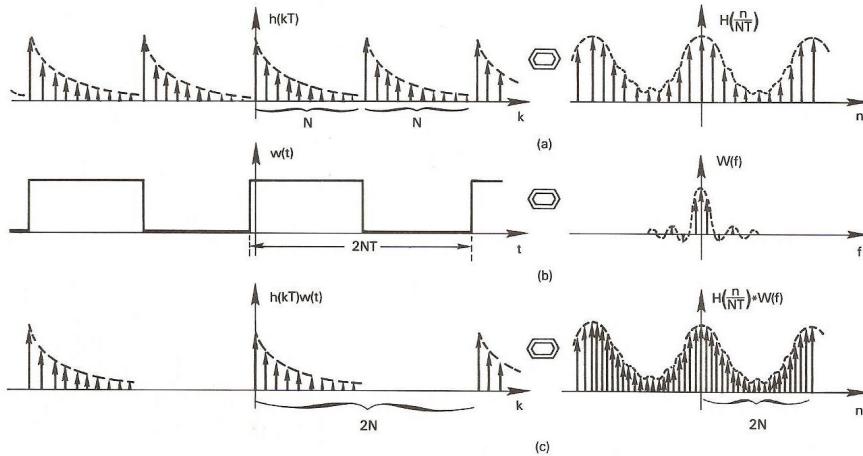


Figure 4.10 Example illustrating the false resolution enhancement by adding zeros (Brigham, 1988)

From the complex amplitude X_n provided by the FFT, the real part is considered because it is directly related to the amplitude associated to each frequency (see Eq. (4.7)). In addition, only the amplitudes associated with the interval frequency $1/D < f < f_N$ are considered (as explained in Section 4.3.2). The wave spectrum is calculated as:

$$E(f_i) = \frac{1}{\Delta f} \frac{1}{2} a_i^2 \quad (4.14)$$

However, this is not the final result. The final spectrum is found by averaging the amplitudes' values in the interval band $\delta f = p\Delta f$. This is mathematically the same operation as dividing the records in subrecords, computing the wave spectrum for each subrecord and taking the average (see Section 4.3.2). Figure 4.11 illustrates the difference between the initial and the final (i.e. averaged) wave spectrum.

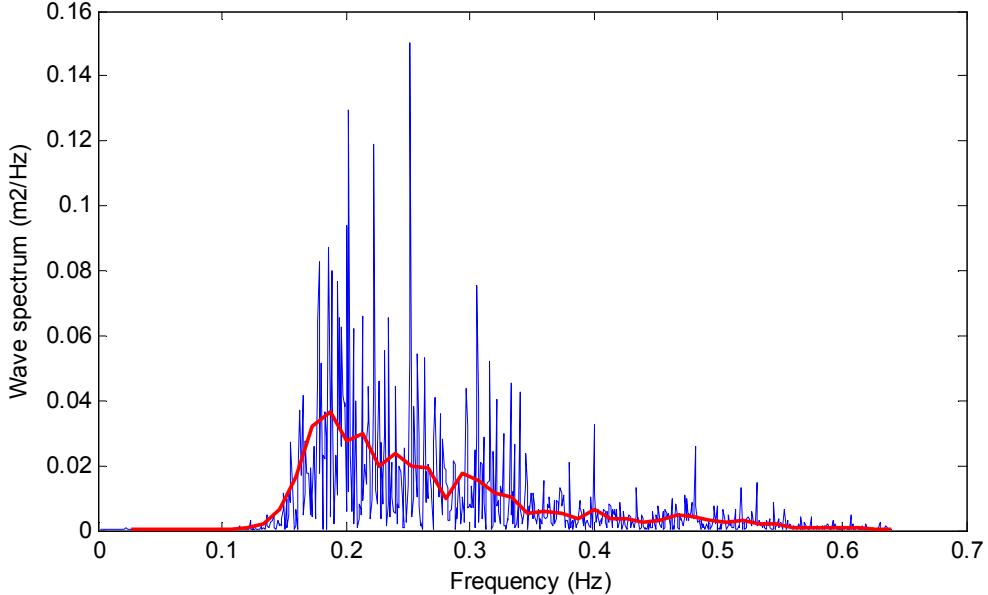


Figure 4.11 Averaged spectrum (red) compared to the initial one (blue) (Tortosa: 21/09/1991 00:16 h)

4.5. Spectral parameters

The spectral parameters can be split in two groups, depending on whether they characterize the surface elevation (H_{m_0}, T_m, T_0, T_c) or the spectral shape (ε, ν, Q_p). In addition, the *BFI* is computed, which is related to the occurrence of extreme waves such as, freak waves for example.

In most of parameter calculations, spectral moments are used. They are defined as:

$$m_n = \int_0^\infty f^n E(f) df \quad \text{for } n = \dots, -3, -2, -1, 0, 1, 2, 3, \dots \quad (4.15)$$

One of the most used is the zeroth-order moment, which is the variance of the surface elevation (see Chapter 5).

4.5.1. Parameters related with the surface elevation

First of all, note that the expressions shown bellow of the most important parameters are derived assuming linear theory.

- Mean wave height

$$H_{mean} = \sqrt{2\pi m_0} \quad (4.16)$$

- Significant wave height

$$H_s \approx 4\sqrt{m_0} \quad (4.17)$$

- Root-mean-square wave height

$$H_{rms} = \sqrt{8m_0} \quad (4.18)$$

- Expected maximum wave height:

$$E\{H_{max}\} = 2\left(1 + \frac{0.29}{\ln N}\right)\sqrt{2\ln N}\sqrt{m_0} \quad (4.19)$$

- Mean zero-crossing period

$$T_0 = \sqrt{\frac{m_0}{m_2}} \quad (4.20)$$

- Mean wave period (inverse of the mean frequency), which is less dependent on high-frequency noise than the previous one.

$$T_m = \frac{m_0}{m_1} \quad (4.21)$$

- Significant wave period

$$\begin{aligned} T_s &= T_{peak} && \text{for swell} \\ T_s &= 0.95T_{peak} && \text{for wind sea} \end{aligned} \quad (4.22)$$

where T_{peak} is the inverse of f_{peak}

4.5.2. Shape parameters

- Goda's parameter (Goda, 1970):

$$Q_p = \frac{2}{m_0^2} \int_0^\infty E^2(f) f df \quad (4.23)$$

- Longuet-Higgins's spectral width (1975):

$$\nu = \sqrt{\frac{m_0 m_2}{m_1^2} - 1} \quad (4.24)$$

- Cartwright and Longuet-Higgins (1956) defined another spectral width parameter:

$$\varepsilon = \sqrt{1 - \frac{m_2^2}{m_0 m_4}} \quad (4.25)$$

which is used in the probability density function of the wave crest defined as local maxima. In general, ν is preferable above ε because it depends on lower order moments and therefore less on the tail of the spectrum.

4.5.3. BFI

For a more detailed explanation of this parameter see Chapter 6. The calculation of BFI is made according to the following expression (Goda, 1970; Janssen, 2003, 2005):

$$BFI = Q_p k_{m_0} \sqrt{2\pi m_0} \quad (4.26)$$

where Q_p is the peakedness parameter of Goda (Eq. (4.23)), k_{m_0} the mean wave number:

$$k_{m_0} = \frac{(2\pi)^2}{g T_0^2} \quad (4.27)$$

and T_0 the mean zero crossing period.

5. LINEAR THEORY

5.1. Probability distributions

5.1.1. Surface elevation

Like many other processes in nature, the surface elevation, according to linear theory, is Gaussian distributed. A theoretical explanation of the applicability of this distribution is the *Central Limit Theorem*. Broadly speaking, it says that the sum of large independent random variables (not necessarily Gaussian distributed) is Gaussian distributed. In linear theory the surface elevation can be considered as the sum of a large number of independent harmonic waves and therefore the above mentioned theorem holds.

$$p(\eta) = \frac{1}{\sqrt{2\pi m_0}} \exp\left(-\frac{\eta^2}{2m_0}\right) \quad \text{for } E\{\eta\} = 0 \quad (5.1)$$

The spectral zero moment m_0 is the variance of the surface elevation:

$$\text{Var}(\eta) = E(\eta^2) = \sum_{i=1}^{\infty} a_i^2 E(\cos(\omega_i t + \alpha_i)) = \sum_{i=1}^{\infty} a_i^2 \frac{1}{2} = \sum_{i=1}^{\infty} 2E(f)\Delta f \frac{1}{2} = m_0 \quad (5.2)$$

in which the definition of wave spectrum has been used (Eq. (4.14)).

5.1.2. Wave crest

Once the surface elevation is considered to be Gaussian distributed, the distributions of the crest and the wave height can be found.

Firstly, one can derive the average time between successive up or down crossings through the level η in terms of the spectrum (Rice, 1945):

$$\bar{T}_\eta = \sqrt{\frac{m_0}{m_2}} \Bigg/ \exp\left(-\frac{\eta^2}{2m_0}\right) \quad (5.3)$$

From Eq. (5.3) the average time between zero up or down crossings directly follows:

$$\bar{T}_0 = \sqrt{\frac{m_0}{m_2}} \quad (5.4)$$

It is now necessary to make a distinction between narrow and wide spectrums. In Figure 5.1, one can appreciate that the wider the spectrum, the more irregular the character of the surface elevation. In the first case (narrow spectrum), associated to more regular waves, a crest height can be directly defined as a maximum surface elevation per wave.

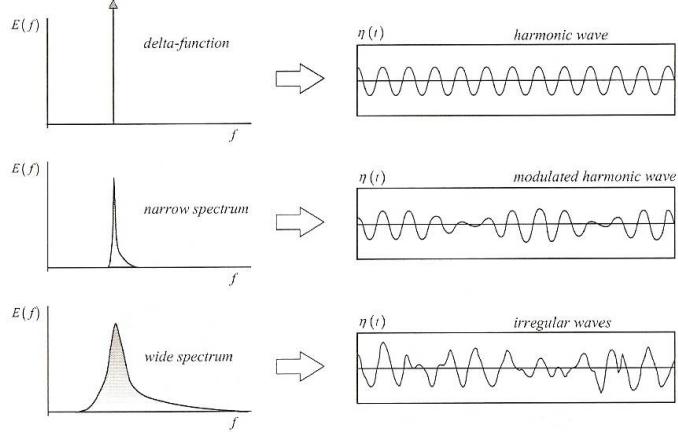


Figure 5.1 Surface elevation for different widths of the spectrum (Holthuijsen,2007)

The relative number of high crests exceeding a certain level η can be derived from Eq. (5.3):

$$\frac{\text{number of crests with } \underline{\eta}_{\text{crest}} > \eta \text{ in duration } D}{\text{total number of crests in duration } D} = \frac{D/\bar{T}_\eta}{D/\bar{T}_0} = \frac{\bar{T}_0}{\bar{T}_\eta} \quad (5.5)$$

This expression is, in fact, the exceedance probability of the crest height:

$$P(\underline{\eta}_{\text{crest}} > \eta) = \frac{\bar{T}_0}{\bar{T}_\eta} = \exp\left(-\frac{\eta^2}{2m_0}\right) \quad (5.6)$$

Therefore, the cumulative distribution function (cdf), that is to say, the non exceedance probability is:

$$F(\eta) = P(\underline{\eta}_{\text{crest}} < \eta) = 1 - \exp\left(-\frac{\eta^2}{2m_0}\right) \quad (5.7)$$

By deriving Eq. (5.7), the probability density function (pdf) is obtained:

$$p_{\underline{\eta}_{\text{crest}}}(\eta) = \frac{\eta}{m_0} \exp\left(-\frac{\eta^2}{2m_0}\right) \quad (5.8)$$

which is the well-known Rayleigh distribution.

Another way of deriving the expression of Eq. (5.8) is using the wave envelope theory. Using the peak frequency (ω_p), one can rewrite the surface elevation as follows:

$$\eta(t) = \operatorname{Re} \left[e^{i\omega_p t} \sum_{n=1}^{\infty} a_n e^{i[(\omega_n - \omega_p)t + \alpha_n]} \right] \quad (5.9)$$

which can be rewritten as:

$$\eta(t) = \eta_c(t) - \eta_s(t) = A_c(t) \cos(\omega_p t) - A_s(t) \sin(\omega_p t) \quad (5.10)$$

where:

$$\left. \begin{aligned} A_c(t) &= \sum_{i=1}^{\infty} a_i \cos[(\omega_i - \omega_p)t + \alpha_i] \\ A_s(t) &= \sum_{i=1}^{\infty} a_i \sin[(\omega_i - \omega_p)t + \alpha_i] \end{aligned} \right\} \quad (5.11)$$

Eq. (5.11) can be expressed in terms of the amplitude³ $A(t)$ and phase $\phi(t)$:

$$\eta(t) = A(t) \cos[\omega_p t + \phi(t)] \quad (5.12)$$

where:

$$A(t) = \sqrt{A_c^2(t) + A_s^2(t)}, \quad \phi(t) = \arctan[A_s(t) / A_c(t)]$$

or, equivalent:

(5.13)

$$A_c(t) = A(t) \cos \phi(t), \quad A_s(t) = A(t) \sin \phi(t)$$

Owing to the narrowness of the process, the amplitudes $A_c(t)$ and $A_s(t)$ vary gradually over time (Massel, 1996) as the spectrum is centred on ω_p . Therefore, $A(t)$, the wave envelope also varies slowly over time and practically equals the wave amplitude as illustrated in Figure 5.2.

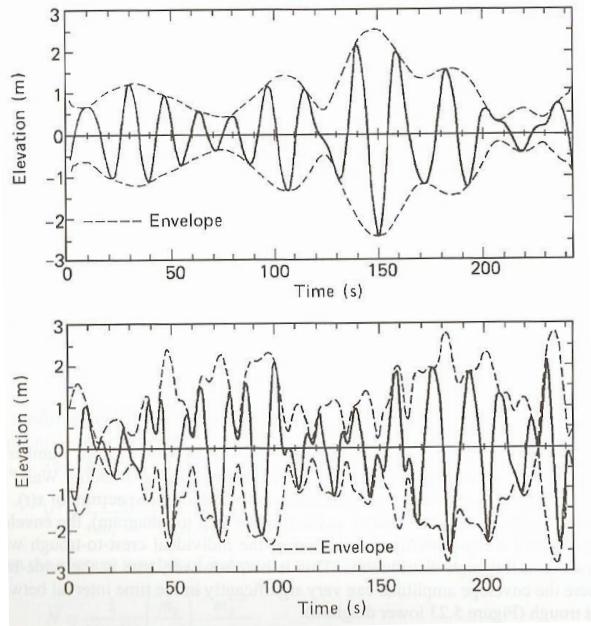


Figure 5.2 Wave envelope for: narrow spectrum (upper), wide spectrum simulated with a Pierson-Moskowitz spectrum (lower); from Carter et al., 1986 (Tucker & Pitt, 2001)

³ The term *amplitude* can have different meanings. In the spectral analysis, as in more general applications, it is usually associated to the amplitude of the harmonic waves in which the surface elevation is split. However, in most cases of the probability distribution's theories presented in this study, it is related to the amplitude of the theoretical wave envelope which can describe the crest and trough levels of surface elevation and therefore it is a function of time instead of a constant value.

$A_c(t)$ and $A_s(t)$ are independent and Gaussian distributed variables (the Central Limit theorem holds because they are the sum of independent cosine and sine waves respectively). Consequently, the two-dimensional probability density function is:

$$f_2(A_c, A_s) = f(A_c) \cdot f(A_s) = \frac{1}{2\pi\sigma_\eta^2} \exp\left(-\frac{A_c^2(t) + A_s^2(t)}{2\sigma_\eta^2}\right) \quad (5.14)$$

In order to transform $A_c(t)$ and $A_s(t)$ into the amplitude and phase variables, the Jacobian of the variable transformation is used:

$$J = \left| \frac{\partial(A_c, A_s)}{\partial(A, \phi)} \right| = A \cos^2 \phi + A \sin^2 \phi = A \quad (5.15)$$

and therefore:

$$f_2(A, \phi) = f_2[A_c(A, \phi), A_s(A, \phi)] J = \frac{A}{2\pi\sigma_\eta^2} \exp\left(-\frac{A^2}{2\sigma_\eta^2}\right) \quad (5.16)$$

Finally, by integrating the joint probability and substituting the variance of the surface elevation for the spectral zero moment, the one-dimensional pdf for the amplitude is found:

$$f(A) = \int_{-\pi}^{\pi} f_2(A, \phi) d\phi = \frac{A}{m_0} \exp\left(-\frac{A^2}{2m_0^2}\right) \quad (5.17)$$

Eq. (5.17) is the same as Eq. (5.8) (in which the amplitude is expressed in terms of the wave crest). Although not being of particular relevance, by integrating Eq. (5.18) one can find that the phase is Uniformly distributed between 0 and 2π :

$$f(\phi) = \int_0^\infty f_2(A, \phi) dA = \frac{1}{2\pi} \quad (5.18)$$

For a wide spectrum the total number of crests higher than a certain level η differs from the number of down-crossings through this level as one considers only the maximum wave crest per wave, not all the local maxima. From another viewpoint, the envelope does not slowly vary over time and differs from the wave amplitude (see Figure 5.2). Hence, the above mentioned results cannot theoretically be applied.

However, if we consider the crest height as the maximum crest height per wave, observations have shown that this is nearly Rayleigh distributed.

5.1.3. Wave height

From the engineering view point, the wave height is preferred to the crest height. In deep water and under the assumption of a narrow-band spectrum, the following approximation can be made:

$$\underline{H} \cong 2\underline{\eta}_{crest} \quad (5.19)$$

and therefore the cdf and pdf of the wave height are respectively:

$$F(H) = P(\underline{H} < H) = 1 - \exp\left(-\frac{H^2}{8m_0}\right) \quad (5.20)$$

$$p_{\underline{H}}(H) = \frac{H}{4m_0} \exp\left(-\frac{H^2}{8m_0}\right) \quad (5.21)$$

According to these results, the only parameter which one needs to know in order to describe the wave height distribution, is the zeroth-order spectral moment, that is to say, the variance of the surface elevation. In order to compare wave height exceedance probabilities of different records (and therefore with different variance) it is convenient to normalize the wave height as: $\hat{H} = H/\sqrt{m_0}$ which results in the following cdf:

$$F(\hat{H}) = 1 - \exp\left(-\frac{\hat{H}^2}{8}\right) \quad (5.22)$$

With the wide-band spectrum, which often occurs in reality, the wave height suffers from the same problem as the wave crest. The reason for this is that often there are different sources of wave energy as for example a distant storm or local winds. The wave crest probability is not theoretically derived and, therefore, the wave height probability. However, once more, observations more or less agree with the Rayleigh distribution, although a certain scaling factor is found: the real waves are slightly smaller than the predicted values (see Figure 5.3).

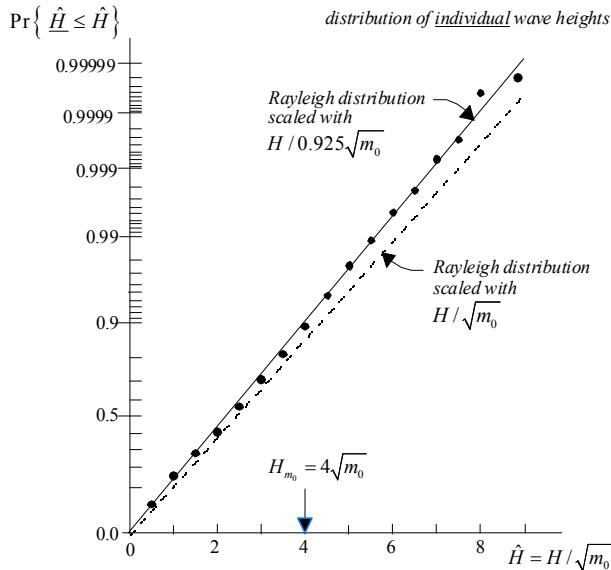


Figure 5.3 Scaling factor between Rayleigh distribution and observations from five hurricanes in the Gulf of Mexico (Forristall, 1978; from Holthuijsen, 2007)

5.1.4. Maximum wave crest/height

The theoretical derivation of extreme values will be done for the wave crest in the case of a narrow banded spectrum. To obtain the associated expression of the wave height only the relationship of Eq. (5.19) should be considered.

In the theory of maximum events, the non-exceedance probability of the maximum value of N , independent and identically distributed random variables, can be written as:

$$F_N(x_{\max}) = [F(x)]^N \quad (5.23)$$

Applying the expression of Eq. (5.23) on Eq. (5.7), the non-exceedance probability (cdf) of the maximum wave crest is found:

$$F_N(\eta_{\max}) = \left[1 - \exp\left(-\frac{\eta^2}{2m_0}\right) \right]^N \quad (5.24)$$

in which N is the number of crests. Then, the pdf is the derivative of Eq. (5.24). If one considers the standardized variable $x = \eta / \sqrt{m_0}$, the cdf becomes:

$$F_N(x_{\max}) = \left[1 - \exp\left(-\frac{x^2}{2}\right) \right]^N \quad (5.25)$$

The r th moment, provided it exists, is given by:

$$\begin{aligned} M_r(N) &= \int_{-\infty}^{\infty} x^r dF_N(x_{\max}) = \int_0^{\infty} rx^{r-1}(1 - F_N(x_{\max}))dx - \int_{-\infty}^0 rx^{r-1}F_N(x_{\max})dx \\ &= r \int_0^{\infty} x^{r-1} \left(1 - \left(1 - e^{-\frac{1}{2}x^2} \right)^N \right) dx \end{aligned} \quad (5.26)$$

By expanding the above expression binomially, and by integrating term by term, Eq. (5.26) becomes (Cartwright, 1958):

$$M_r(N) = 2^{\frac{1}{2}r} \left(\frac{r}{2} \right)! \left[N - \frac{N(N-1)}{2^{\frac{1}{2}r} \cdot 2!} + \frac{N(N-1)(N-2)}{3^{\frac{1}{2}r} \cdot 3!} \dots + \frac{(-1)^N}{(N-1)^{\frac{1}{2}r}} - \frac{(-1)^N}{N^{\frac{1}{2}r}} \right] \quad (5.27)$$

which is unsuitable for the calculation for large N . However, if N is large and the exceedance probability $(1 - F(x))$ is small, the following approximation can be made:

$$F(x_{\max}) = [F(x)]^N \approx \exp(-N \cdot (1 - F(x))) \quad (5.28)$$

which is called the asymptotic approximation. Cartwright (1958) quantified the difference between exact and asymptotic formulae for the non exceedance probability $F_N(\eta_{\max})$. Figure 5.4 illustrates such a difference, concluding that for $N > 128$, the difference is practically null.

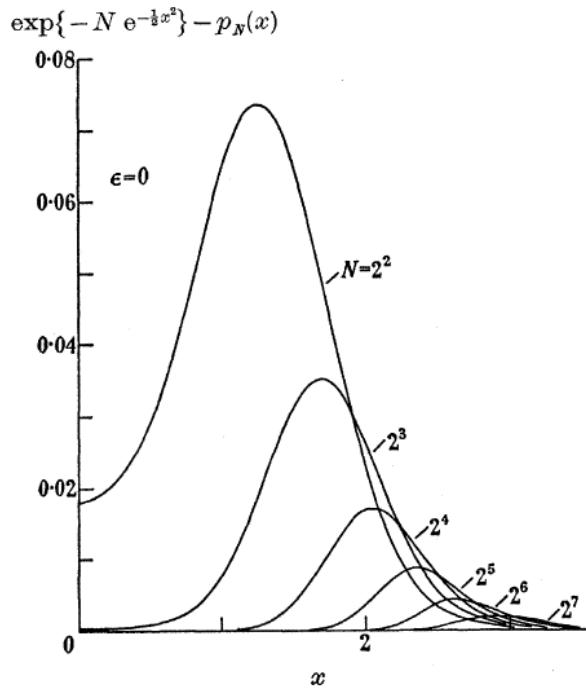


Figure 5.4 Difference between exact and asymptotic formulae for the non-exceedance probability density function (shown as $p_N(x)$) (Cartwright, 1958)

Using this principle the asymptotic formulae for all moments is (Cartwright, 1958):

$$M_r(N) \approx (2 \ln N)^{\frac{1}{2}r} \left[1 + \frac{1}{2} r A_1 (\ln N)^{-1} + \frac{\left(\frac{1}{2}r\right)\left(\frac{1}{2}r-1\right)}{2!} r A_2 (\ln N)^{-2} + \dots \right] \quad (5.29)$$

where:

$$\begin{aligned} A_1 &= 0.5772\dots, \\ A_2 &= 1.9781\dots, \end{aligned} \quad (5.30)$$

Eq. (5.29) is correct to an order $(\ln N)^{\frac{1}{2}r} N^{-1}$ and the accuracy therefore increases with increasing N and decreasing r . For the expected value of the normalized maximum crest height, $M_1(N)$, the first two terms of Eq. (5.30) are enough: the error is of only 2% for $N=20$ (Longuet-Higgins, 1952) and rapidly decreases with increasing N :⁴

$$E_N \{x_{\max}\} \approx \left(1 + \frac{0.29}{\ln N}\right) \sqrt{2 \ln N} \quad (5.31)$$

Another parameter that is sometimes used is the most probable value:

$$\text{mod}_N \{x_{\max}\} \approx \sqrt{2 \ln N} \quad (5.32)$$

which represents the value at the maximum in the pdf. For a sample of 200 waves the expected maximum is about 5% higher than the most probable maximum.

⁴ E here means expected value rather than E of spectral analysis

5.2. Parameters

In this Section the principal parameters which have been calculated in the study are cited. Most of them relate to the period and the crest/height, which are both calculated directly from the signal and the wave spectrum according to the relationships given in Table 5.1. In order not to be repetitive, the parameters related to the wave crest are not explicitly shown as they are half those of the wave height.

Table 5.1 Principal calculated parameters

| Parameter | Statistical | Spectral |
|------------|----------------------|---|
| T_{mean} | \bar{T} | $T_m = m_0/m_1$ $T_0 = \sqrt{m_0/m_2}$ |
| H_{mean} | \bar{H} | $\sqrt{2\pi m_0}$ |
| H_{rms} | $\sqrt{\bar{H}_i^2}$ | $\sqrt{8m_0}$ |
| H_s | $H_{1/3}$ | $4\sqrt{m_0}$ |

5.2.1. The significant wave height

Special attention is paid to the significant wave height (H_s) because it is frequently used in engineering. H_s is defined as the mean of the highest one third wave heights from the record. The reason for choosing one third lies in the fact that with this definition, H_s approximates the visual estimated wave height. Typically, this parameter can be calculated directly from the record or from the spectrum. In order to distinguish them, they are named respectively $H_{1/3}$ and H_{m_0} . The first one is merely the average of the highest one third wave heights. In linear theory, considering the wave height Rayleigh distributed, the second one can be calculated with the following expression:

$$H_{m_0} = 4.004\dots\sqrt{m_0} \approx 4\sqrt{m_0} \quad (5.33)$$

The derivation of Eq. (5.33) can be summarized in the following steps:

- Calculation of the threshold wave height H^* which has an exceedance probability of 1/3

$$P(H^*) = \exp\left(-\frac{H^{*2}}{8m_0}\right) = \frac{1}{3} \quad (5.34)$$

The threshold is approximately:

$$H^* \approx 2.96\sqrt{m_0} \approx 1.048H_{rms} \quad (5.35)$$

- Calculation of the mean of all wave heights larger than H^*

$$H_{m_0} = \frac{\int_{H^*}^{\infty} H p(H) dH}{\int_{H^*}^{\infty} p(H) dH} = 3 \int_{H^*}^{\infty} \frac{H^2}{4m_0} \exp\left(-\frac{H^2}{8m_0}\right) dH \quad (5.36)$$

In fact, considering $H_{1/p,R}$ as the mean of the highest $1/p$ waves and $H_{p,R}^*$ the corresponding threshold, generalisations of expressions of Eq. (5.33) and (5.35) are:

$$H_{1/p,R} = \left(2\sqrt{2 \ln(p)} + p\sqrt{2\pi} \operatorname{erfc}(\sqrt{\ln(p)}) \right) \sqrt{m_o} \quad (5.37)$$

$$H_{p,R} = \left(2\sqrt{2 \ln(p)} \right) \sqrt{m_o} \quad (5.38)$$

where erfc is the complementary error function:

$$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt \quad (5.39)$$

In the frame of linear theory, $H_{1/3}$ and H_{m_0} are, in mean terms, the same parameter. However, a perfect agreement between them is not found, which manifest the limitations of linear theory. In general, $H_{1/3}$ is 5%-10% lower than H_{m_0} (see Figure 5.5).

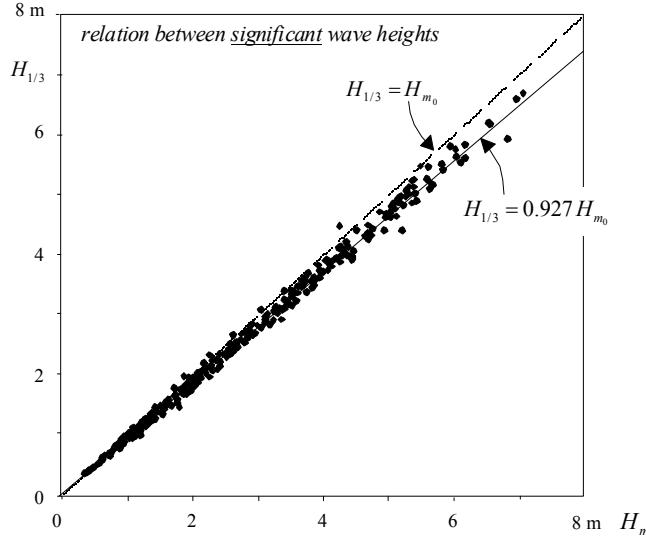


Figure 5.5 The observed and theoretically estimated significant wave height (Holthuijsen, 2007)

Longuet-Higgins (1980) suggested $H_{1/3} = 0.925H_{m_0}$ for the data of Forristall (1978) data and proposed that a rescaled Rayleigh distribution should be considered, with a significant wave height slightly smaller than the spectral one (see Figure 5.3). Another similar relation has been found by best-fit linear approximation with data from the North Sea: $H_{1/3} = 0.927H_{m_0}$ (Holthuijsen, 2007) and also by Forristall (1978): $H_{1/3} = 0.942H_{m_0}$

5.3. Limitations of Rayleigh theory

The Rayleigh distribution for the wave crest/trough assumes: linearity and a narrow-band process. These two basic assumptions are at the same time its limitations, as the reality is more complex than this.

Although considering that nonlinear effects did not occur and therefore the surface elevation had a Gaussian character, if the spectrum is wide, the Rayleigh theory could not strictly be applied. That may result in Rayleigh's overprediction. If one comes back to Figure 5.2, it is clear that the wave envelope is equal or higher than the wave amplitude and the wider the spectrum, the more difference there will be between them.

Moreover, in the derivation for the wave height distribution, another important assumption is made: the height is double the crest height. In the linear case, although this is true in the mean, this relationship applied to random variables is not correct. Making this assumption, the mean wave height is not affected but the statistics of higher wave heights are overpredicted. Actually, the wave height distribution should be derived from the sum of the crest and the trough (in absolute value). Taking double crest height, one assumes that crests and troughs are always paired as if they were symmetric. In other words, the possibility of having a low trough preceding a high crest, or viceversa, is not considered and therefore the probability of high wave heights is artificially enhanced.

In conclusion, the Rayleigh distribution may overpredict the crests/troughs due to the wide character of the spectrum.

Moreover, in reality, nonlinear effects can be also present which further complicates the problem. It is difficult to describe nonlinear effects or to attempt to find a distribution which includes most of them. In deep water, the most important are quadruplet wave-wave interactions and wave breaking due to white capping.

In Chapter 6, the Rayleigh-Edgeworth distribution is presented. It accounts for nonlinear effects related to the interaction between waves which involves the kurtosis of the surface elevation. Such nonlinearities enhance wave heights, especially higher wave heights. However, it does not consider the asymmetry of the surface profile which can take place in the nonlinear case: the crest becomes higher and more peaked and the trough more rounded and less deep. Such asymmetry is included in the theory of Tayfun (1994).

Wave breaking is a very complicated process which is not dealt with in the present study. This type of nonlinear effect produces the diminution of wave heights, possibly counteracting the effect of wave-wave interactions. Tayfun (1981a) suggested that the overprediction could be due to wave heights being steepness limited by the physical process of breaking. However, others (Chen & Borgman, 1979) have showed that measured wave heights and periods are rather far away from the breaking criteria.

In addition, for the calculation of the expected maxima, the asymptotic assumption is made. In Section 5.1.4, it was commented that for more than 128 waves, the error made is negligible. Moreover, independency between waves is assumed. If each wave is highly correlated to the adjacent waves, the effective number of independent waves is reduced. However, analysing 10.000 waves taken in the Bay of Biscay on 21 and 22 May, Cartwright (1958) found that this correlation was practically negligible.

To sum up, in the real sea, when all possible effects are combined, it is difficult to determine the reason for the discrepancy between observations and linear theory.

6. NONLINEAR THEORY

6.1. Introduction

In this Chapter, two different theories are presented. The first one includes the enhancement of wave heights, without considering any asymmetry between crests and troughs. Such enlargement of wave height seems to be closely related to the kurtosis parameter of surface elevation. In the second theory, the asymmetric character of waves is considered and therefore a distinction is made between the probability of crests and troughs. For the description of this behaviour, skewness has a key role.

In both cases the statistics of maximum wave crest/height are not mentioned since their explicit expressions become very complicated. So, in order to calculate them, an analogue procedure to linear theory should be applied here, considering the theory of extreme events.

6.2. Inclusion of kurtosis

6.2.1. Surface elevation

In linear theory, the sea surface elevation can be expressed as a combination of independent random variables symmetrically distributed around 0, with variance V_i :

$$\eta = \sum_{i=1}^N \alpha_i \eta_i \quad (6.1)$$

in which the η is Gaussian distributed with a variance $\sum V_i$.

However, higher order interactions between component waves may exist. Therefore, by adding these high order terms to Eq. (6.1), the following expression is obtained (Longuet-Higgins, 1963):

$$\eta = \alpha_i \eta_i + \alpha_{ij} \eta_i \eta_j + \alpha_{ijk} \eta_i \eta_j \eta_k + \dots \quad (6.2)$$

where α_i , α_{ij} , α_{ijk} , etc. are constants (and the summation convention is used). Assuming, as in linear theory, narrow banded spectra, the pdf, with zero mean value, is (Longuet-Higgins, 1963):

$$p(\eta) = \frac{1}{\sqrt{2\pi m_0}} e^{-\frac{1}{2m_0}\eta^2} \left(1 + \frac{1}{6} \lambda_3 H_3 + \left(\frac{1}{24} \lambda_4 H_4 + \frac{1}{72} \lambda_3^2 H_6 \right) + \dots \right) \quad (6.3)$$

where λ_3 and λ_4 are respectively the coefficients of skewness and kurtosis (see Chapter 3) and H_i the Hermite polynomials:

$$\begin{aligned} H_3 &= f^3 - 3f, & H_4 &= f^4 - 6f^2 + 3, \\ H_5 &= f^5 - 10f^3 + 15f, & H_6 &= f^6 - 15f^4 + 45f^2 - 15, \end{aligned} \quad (6.4)$$

being $f = \eta / \sqrt{m_0}$; or

$$\begin{aligned} H_n(x) &= (-1)^n \exp\left(-\frac{1}{2}x^2\right) \frac{d^n}{dx^n} \exp\left(\frac{1}{2}x^2\right) \equiv \\ &\equiv x^n - \frac{n(n-1)}{1!} \frac{f^{n-2}}{2} + \frac{n(n-1)(n-2)(n-3)}{2!} \frac{x^{n-4}}{2^2} \end{aligned} \quad (6.5)$$

The distribution given by Eq. (6.3) is a Gram-Charlier series, in which the standard deviation has been approximated by the zero-order spectral moment. Without the high order terms, it becomes the Normal distribution, which agrees with linear theory. In fact, the result for nonlinear theory is the product of Normal distribution by a factor which depends on skewness and kurtosis. Therefore, both these parameters are related with nonlinearities.

6.2.2. Wave crest

In an analogous way to the linear theory, in nonlinear theory Eq. (5.14) becomes (the variables have been normalized) (Mori & Yasuda, 2001):

$$\begin{aligned} p(A_c', A_s') &= \frac{1}{2\pi} e^{-\frac{1}{2}(A_c' + A_s')^2} \left[1 + \frac{1}{6} k_{30} H_3(A_c') + \frac{1}{24} k_{40} H_4(A_c') + \frac{1}{24} k_{30}^2 H_6(A_c') \right] \\ &\quad \left[1 + \frac{1}{6} k_{03} H_3(A_s') + \frac{1}{24} k_{04} H_4(A_s') + \frac{1}{24} k_{03}^2 H_6(A_s') \right] \end{aligned} \quad (6.6)$$

Eq. (6.6) is the truncation of a longer summation. H_i are the Hermite polynomials (see Eq. (6.5)) and k_{ij} are as follows:

$$\begin{aligned} k_{ij} &= k_{ji}, \\ k_{30} &= \frac{E(\eta^3)}{\sigma^3}, \\ k_{40} &= \frac{E(\eta^4)}{\sigma^4} - 3, \end{aligned} \quad (6.7)$$

Note that k_{40} is the common kurtosis (Eq. (3.9)) minus 3 (i. e. For the Gaussian distribution). Ignoring the third term of Eq. (6.6), the pdf of envelope A (normalized) follows from the integration of the joint probability of Eq. (6.6):

$$p(A, \phi) = A p(A_c, A_s) \quad (6.8)$$

$$p(A) = \int_0^{2\pi} p(A, \phi) d\phi = \int_0^{2\pi} A p(A_c, A_s) d\phi \quad (6.9)$$

$$p(A) = A e^{-\frac{1}{2}A^2} \left[1 + \frac{1}{3} \kappa_{40} \left(1 - A^2 + \frac{1}{8} A^4 \right) \right] \quad (6.10)$$

If one does not consider the higher order terms, Eq. (6.10) gives the Rayleigh distribution.

In Eq. (6.10), the skewness κ_{30} does not appear. The reason is that the integral of the 2nd term of the equation becomes zero because they are odd functions of ϕ . An extension of the

Edgeworth distribution to sixth order, including the third term of Eq. (6.6) (proportional to the squared skewness), would account for the contribution of the skewness. Nevertheless, such a contribution to the pdf of narrowband wave trains is small (Mori & Janssen, 2005).

6.2.3. Wave height

In the narrowband approximation, the wave height is double the amplitude. Hence, from Eq. (6.10) the pdf of the wave height becomes:

$$p(H) = \frac{1}{4} H e^{-\frac{1}{8} H^2} \left[1 + \frac{1}{384} \kappa_{40} (H^4 - 32H^2 + 128) \right] \quad (6.11)$$

By integrating Eq. (6.11), the exceedance probability function of wave height can be immediately found:

$$P_H(H) = e^{-\frac{1}{8} H^2} \left[1 + \frac{1}{384} \kappa_{40} H^2 (H^2 - 16) \right] \quad (6.12)$$

6.2.4. Estimation of kurtosis from the spectrum: the BFI

To measure the relative importance of nonlinearity and dispersion, the Benjamin-Feir Index was introduced and defined for the narrow case as (Janssen, 2003):

$$BFI = \frac{\epsilon \sqrt{2}}{\sigma_\omega'} \quad (6.13)$$

where ϵ is the slope parameter and σ_ω' the relative width of the frequency spectrum:

$$\epsilon = k_0 \sqrt{m_0} \quad (6.14)$$

$$\sigma_\omega' = \frac{\sigma_\omega}{\omega_0} \quad (6.15)$$

where k_0 is the characteristic wave number, ω_0 the characteristic frequency, m_0 the variance and σ_ω a measure of spectral width. The introduction of BFI was closely related to the study of the probability of freak wave occurrence, also called rogue waves, which are extreme wave events in a moderate sea state. In a manner, the BFI quantifies the increased probability of encountering a freak wave and characterises the process of self-focusing. Therefore, higher BFI are theoretically associated to more probability of occurrence of giant waves.

For the computations, the mean wave number has been chosen as the characteristic wave number and the relative width of the spectrum as the inverse of the peakedness parameter of Goda (Goda, 1970; Janssen, 2005):

$$\epsilon = k_{m0} \sqrt{m_0}, \quad \sigma_\omega' = \sqrt{\pi} / Q_p \quad (6.16)$$

$$Q_p = 2 \frac{\int \omega E^2(\omega) d\omega}{m_0^2} \quad (6.17)$$

in which $E(\omega)$ is the wave spectrum. k_0 is estimated at ECMWF (European Centre of Medium-range weather forecast) with the peak wave number of the spectrum, but this would lead to a discredited variation of the BFI of 10% or more, due to the discredited character of the spectrum. According to Lanssen et al. (2006), the peak wave number is replaced by the mean wave number. Others (e.g. Rotés, 2004) calculate σ_ω as the ratio:

$$\sigma_\omega' = \Delta f / f_p \quad (6.18)$$

Δf being half the spectral band width at half the maximum spectral peak energy and f_p the peak frequency.

Mori and Janssen (2006) found that for unidirectional waves the (normalized) kurtosis κ_{40} (see Eq. (6.7)) depends on the square of BFI:

$$\kappa_{40} = \frac{\pi}{\sqrt{3}} BFI^2 \quad (6.19)$$

They concluded that the kurtosis depends on the wave steepness and the relative width of the frequency spectrum. Benjamin and Feir (1967) and Alber and Saffman (1978) showed that these parameters play a key role in the evolution of the spectrum of deep-water gravity waves. Nonlinearity may counteract dispersion in such a way that focusing of the wave energy may occur, resulting in extreme wave events. Therefore, to measure the relative importance of nonlinearity, Janssen (2003) introduced the BFI. The factor of $\sqrt{2}$ is included for historical reasons as according to Alber and Saffman (1978) a random wave train becomes unstable if $BFI > 1$.

However, Alber (1978) demonstrated that for random sea states, the BFI vanishes if the wave spectrum is sufficiently wide in frequency space. In fact, Olagnon & Magnusson (2004) found after analysing 200.000 waves from the Frigg storm database, that although BFI is biased towards higher values when freak waves occur, a small BFI is no guarantee that freak waves will not occur. They concluded that BFI has a little practical value because it does not seem to be a stable estimator.

By considering $n = H / H_{m0}$ and using (6.19), Eq. (6.12) becomes the Rayleigh-Edgeworth distribution in terms of the BFI:

$$F(n) = e^{-2n^2} \left[1 + 2\pi n^2 (n^2 - 1) \frac{BFI^2}{3\sqrt{3}} \right] \quad (6.20)$$

Using Eq. (6.20), it becomes evident that for large values of BFI, the wave train is unstable. This physical phenomenon is translated to the probability space by having negative probabilities. In Figure 6.1 is shown, as an example, the presence of negative probabilities in the case of $BFI = 2$. However, the breakpoint in which negative probabilities appear is about 1.286 instead of 1.

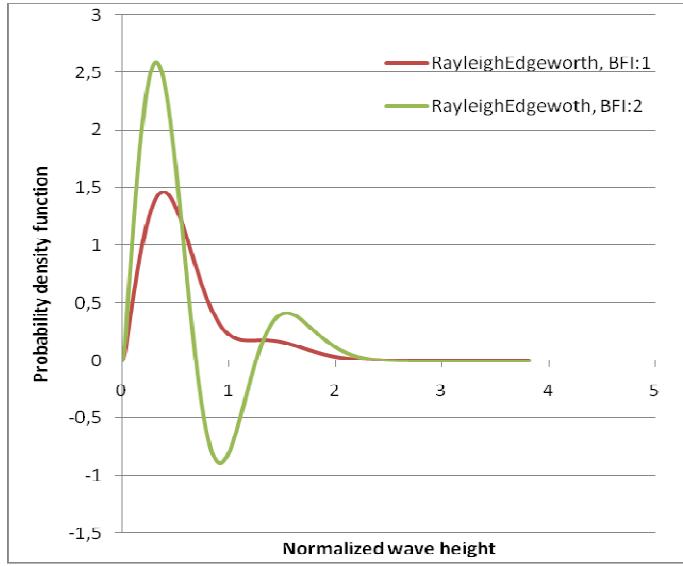


Figure 6.1 Comparison of two Rayleigh-Edgeworth distribution with different values of BFI. The wave height is normalized dividing by the linear significant wave height

6.3. Inclusion of skewness

6.3.1. Surface elevation

In a analogous manner to Longuet-Higgins (1963), Tayfun (1994) decomposed the surface elevation in the sum of two terms (second order approximation):

$$\eta = \eta_1 + \eta_2 \quad (6.21)$$

corresponding to the first two terms of Eq. (6.1). Physically, any two components of η_1 (linear combination of independent cosine waves) width frequencies ω_m and ω_n generate two nonlinear components of frequencies $\omega_m \pm \omega_n$. Therefore, considering the results of Longuet-Higgins (1963) (see Section 6.2.1), the pdf of the surface elevation is:

$$p(\eta) = \frac{1}{\sqrt{2\pi m_0}} e^{-\frac{1}{2m_0} \eta^2} \left(1 + \frac{1}{6} \lambda_3 \eta (\eta^2 - 3) \right) \quad (6.22)$$

which includes the first two terms of Eq. (6.3), λ_3 being the skewness of the surface elevation. The kurtosis does not appear. Actually, Jha & Winterstein (2000) pointed out that the skewness varies linearly with steepness whereas the kurtosis is predicted to vary quadratically with the steepness. Since steepness is far less than unity, squared steepness is even smaller, suggesting that nonlinear effects will be almost displayed by the skewness.

6.3.2. Wave crest/trough

Within the context of a second-order approximation (see Eq. (6.21)), the joint probability density functions of the normalized wave amplitude $\xi = A / \sqrt{2m_0}$ and the phase ϕ is (Tayfun, 1994):

$$p(\xi, \phi) = 2\xi e^{-\xi^2} \cdot \frac{1}{2\pi} \left[1 + \frac{\sqrt{2}}{3} \lambda_3 \xi (\xi^2 - 2) \cos \phi \right] \quad (6.23)$$

where λ_3 is the skewness of the surface elevation. By integrating Eq. (6.23), it is found that A is Rayleigh distributed but the phase is not Uniform distributed (unless the skewness becomes zero):

$$p(\phi) = \frac{1}{2\pi} \left[1 - \frac{1}{6} \sqrt{\frac{\pi}{2}} \lambda_3 \cos \phi \right] \quad (6.24)$$

Bear in mind that this derivation of the theory has been done on the basis of cosine waves. The time during which the profile stays above the mean-zero level (+) is conditioned by $0 \leq \phi \leq \pi / 2$ and $3\pi / 2 \leq \phi \leq 2\pi$. For positive skewness, the probability of surface elevation being above the mean zero level is slightly lower than for being below. The corresponding probabilities are:

$$P(\pm) = \frac{1}{2} \left[1 \mp \frac{\lambda_3}{3\sqrt{2\pi}} \right] \quad (6.25)$$

where:

$$\begin{aligned} P(+) &= P(\eta > 0) \text{ for which } 0 \leq \phi \leq \pi / 2 \text{ and } 3\pi / 2 \leq \phi \leq 2\pi \\ P(-) &= P(\eta < 0) \text{ for which } \pi / 2 \leq \phi \leq 3\pi / 2 \end{aligned} \quad (6.26)$$

This is reasonable because a positive skewness implies a pdf with the peak translated to negative values but a higher probability of high crests and a lower probability of deep troughs (see Figure 6.2). In fact, according to Tayfun (2001), $0 \leq \lambda \leq 0.3$ is a typical range of values in deep water.

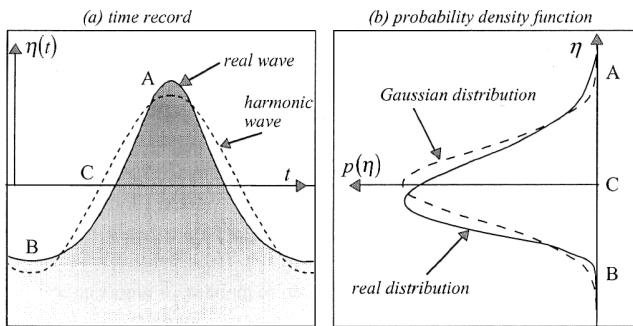


Figure 6.2 Comparison between a harmonic wave with a nonlinear wave with higher peaked crest and rounded trough and their probability density function (Holthuijsen, 2007)

Therefore, according to this nonlinear theory, the wave amplitude in general is Rayleigh distributed but if we look at crests and troughs separately the probability density function is skewed because they are the conditional density function of being, respectively, positive or negative. Therefore, if we integrate expression of Eq. (6.23) from 0 to 2π , the probability density function of wave amplitude becomes the Rayleigh distribution because the term involving the skewness

vanishes (as explained in Section 6.2). However, the probability of crests and troughs are, respectively, the following conditional probabilities:

$$p^+(\xi) = \frac{\int_0^{\pi/2} p_{\xi\phi} d\phi}{\int_0^{\pi/2} p_\phi d\phi}, \quad p^-(\xi) = \frac{\int_{\pi/2}^\pi p_{\xi\phi} d\phi}{\int_{\pi/2}^\pi p_\phi d\phi} \quad (6.27)$$

Tayfun (1994) found the mathematical expression for them:

$$p^\pm(\xi) = \frac{1 \pm c_1 \lambda_3 \xi (\xi^2 - 2)}{1 \mp c_0 \lambda_3} p(\xi) \quad (6.28)$$

where $p(\xi)$ is the Rayleigh probability density function and the constant parameters (Al-Humoud et al., 2001):

$$c_0 = \frac{1}{3\sqrt{2\pi}}, \quad c_1 = \frac{2\sqrt{2}}{3\pi} \quad (6.29)$$

Figure 6.4 and 6.4, for a typical range of values of λ_3 in deep water, shows that the probability deviates from the Rayleigh distribution noticeably, the crest showing an excess toward large values and a deficiency in mid-range, and for the trough the effect is the opposite. Therefore, according to this theory, the Rayleigh distribution of the linear theory would underpredict the extreme values of wave crests and overpredict the extreme values of wave troughs.

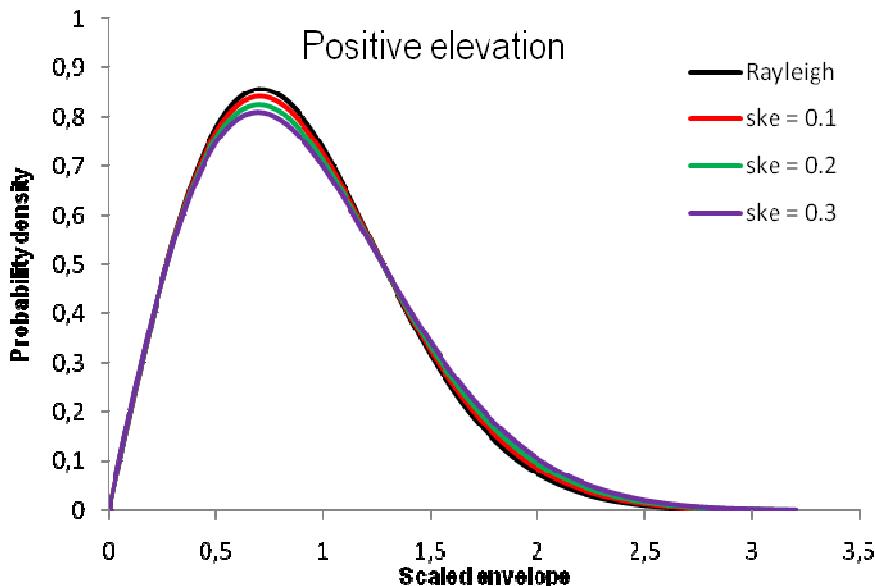


Figure 6.3 Theoretical probability density function given by Eq. (6.28) for positive (crest) surface elevation ($ske = \lambda_3$).

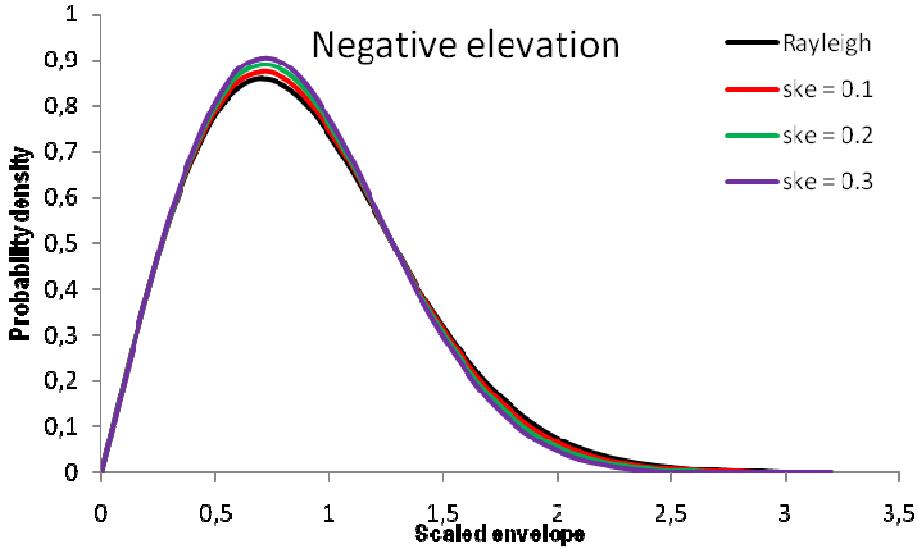


Figure 6.4 Theoretical probability density function given by Eq. (6.28) for negative (trough) surface elevation ($ske = \lambda_3$).

The mean and root-mean-square values associated with p^\pm are:

$$\xi_{mean}^\pm = \frac{1}{1 \mp c_0 \lambda_3} \sqrt{\frac{\pi}{2}} \quad (6.30)$$

$$\xi_{rms}^\pm = \sqrt{\frac{1 \pm c_2 \lambda_3}{1 \mp c_0 \lambda_3}} \quad (6.31)$$

in which c_0 and c_1 are those of Eq. (6.29) and $c_2 = 1/8\sqrt{2}$. By integration of Eq. (6.25), the exceedance distribution is (Al-Humoud et al., 2001):

$$P_{exceed}^\pm(\xi) = \frac{1 \pm c_1 \lambda_3 \xi (\xi^2 - 1/2) \mp c_0 \lambda_3 erfc(\xi) \exp(-\xi^2)}{1 \mp c_0 \lambda_3} P_{exceed}(\xi) \quad (6.32)$$

where $erfc$ is the standard complimentary error function (see Eq. (5.39)), and $P_{exceed}(\xi)$ the Rayleigh exceedance probability.

Many of the above mentioned expressions can be rewritten in slightly simplified forms, using the expansion:

$$\frac{1}{1 \mp c_0 \lambda_3} = 1 \pm c_0 \lambda_3 + o(\lambda_3^2) \quad (6.33)$$

In particular,

$$p^\pm(\xi) = [1 \pm c_0 \lambda_3 \pm c_1 \lambda_3 \xi (\xi^2 - 2)] p(\xi) \quad (6.34)$$

$$P_{exceed}^\pm(\xi) = [1 \pm c_0 \lambda_3 \pm c_1 \lambda_3 \xi (\xi^2 - 1/2) \mp c_0 \lambda_3 erfc(\xi) \exp(-\xi^2)] P_{exceed}(\xi) \quad (6.35)$$

Furthermore, the conditional mean of the highest $1/p$ th crests / troughs is (Tayfun, 2006):

$$\xi_{1/p}^{\pm} = \xi_{1/p,R} \pm \frac{\lambda_3}{3\sqrt{2}}(1 + \ln p) \quad (6.36)$$

where $\xi_{1/p,R}$ is the value of the Rayleigh distribution. For $p = 3$:

$$\xi_{1/p}^{\pm} / \xi_{1/p,R} = 1 \pm 0.3494\lambda_3 \quad (6.37)$$

6.3.3. Wave height

According to the theory of Tayfun (1994, 2006), including the skewness, and assuming the assumption of the wave height being twice the wave amplitude, the wave height is described by the Rayleigh distribution.

6.3.4. Estimation of skewness from the spectrum

The skewness derived from the wave record tends to be an unstable statistic because of its sensitivity to local trends and possible presence of exceptionally large wave heights. Actually, Tayfun (2005) suggested that the skewness would have to be estimated from the surface time history representing relatively high waves. He found that the lower-upper bounds of skewness are:

$$3 \in \left(1 - \frac{\nu\sqrt{2}}{1+\nu^2} \right) \leq \lambda_3 \leq 3 \frac{\epsilon}{1+\nu^2} \quad (6.38)$$

in which ν is the spectral bandwidth defined in Chapter 4 and ϵ the same slope parameter used in the BFI definition (see Section 6.2.4).

7. CREST-TO-TROUGH WAVE HEIGHT DISTRIBUTION

7.1. Introduction

In Chapter 5, it has been mentioned that, in the wide spectrum case, the assumption of $H \approx 2\eta_{crest}$ does not hold even in the narrow case (considering that the wave envelope does not overpredict the crests and troughs). The crest-to-trough wave height distribution, instead of considering the height as double the crest, considers the wave height as the sum of the wave crest and the wave trough, taking into account a certain time lag between them (τ). More specifically, the time lag is approximated by half the mean wave period. In Figure 7.1, this fact is illustrated. The wave height is different from the value of twice the wave amplitude. This theory was firstly derived by Tayfun (1981b), involving very complex expressions. Later, Tayfun (1990) simplified the results for large wave heights.

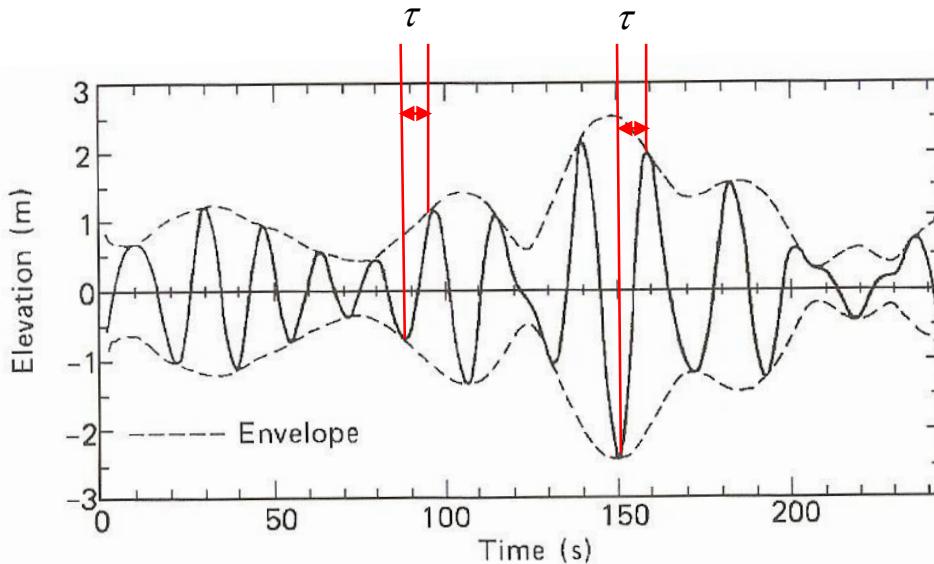


Figure 7.1 Illustration of time lag between wave crest and trough (from Tucker & Pitt, 2001)

In conclusion, according to this theory, the crest and trough, separately, are Rayleigh distributed except for the wave height where the above mentioned considerations are taken into account. Therefore, only the pdf of the wave height is presented. The derivation of an explicit expression for the maximum wave height has not been done but the probability of such maximum can be calculated by numerical integration using the theory of extremes (explained in Chapter 5).

7.2. Wave height

Defining the normalized wave height as the sum of the wave amplitude at time t and the wave amplitude at time $t + \tau$ (the same wave amplitude as in linear theory):

$$\xi = \frac{H}{\sqrt{m_0}} = \xi_1 + \xi_2 = \frac{A(t)}{\sqrt{m_0}} + \frac{A(t + T_m / 2)}{\sqrt{m_0}} \quad (7.1)$$

the probability density function of ξ is given by the convolution integral:

$$f(\xi) = \int_0^\infty f_{\xi_1 \xi_2}(\xi - y, y) dy \quad (7.2)$$

The probability density function of two normalized amplitudes separated by time τ was derived by Rice (1945):

$$f_{\xi_1 \xi_2}(\xi_1, \xi_2) = \frac{\xi_1 \xi_2}{1-r^2} I_0\left(\frac{\xi_1 \xi_2 r_0}{1-r^2}\right) \exp\left[-\frac{\xi_1^2 + \xi_2^2}{2(1-r^2)}\right] \quad (7.3)$$

Where I_0 is the zero-order modified Bessel function of the first kind, which is a fundamental solution for the zero-order modified Bessel's equation:

$$z^2 \frac{d^2 y}{dz^2} + z \frac{dy}{dz} - z^2 y = 0 \quad (7.4)$$

The coefficient $r^2 = r^2(\tau = T/2)$ is:

$$r^2(\tau) = \rho^2(\tau) + \lambda^2(\tau), \quad (7.5)$$

$$\rho(\tau) = \frac{1}{m_0} \int_0^\infty E(\omega) \cos[\omega\tau] d\omega, \quad (7.6)$$

$$\lambda(\tau) = \frac{1}{m_0} \int_0^\infty E(\omega) \sin[\omega\tau] d\omega \quad (7.7)$$

Theoretically, r^2 is the correlation coefficient between the wave amplitudes squared separated by half the wave period but in practice also approximate the correlation factor of the wave amplitudes (without the square).

Therefore, the pdf of ξ results becomes (Tayfun, 1981b):

$$f(\xi) = \frac{1}{1-r^2} \int_0^\xi (\xi - x) x I_0\left(\frac{(\xi - x)x r}{1-r^2}\right) \exp\left[-\frac{(\xi - x)^2 + x^2}{2(1-r^2)}\right] dx \quad (7.8)$$

In a more general case, considering the random character of the wave period Eq. (7.8) becomes:

$$f(\xi) = \int_0^\infty \int_0^\xi f(T) f_2(\xi - \xi_2, \xi_2; T/2) d\xi_2 dT \quad (7.9)$$

However, as $f(T)$ is not well known and it would complicate the solution, the above mentioned approximation of considering half the wave period is taken into account in the following expressions.

According to Eq. (7.8), the first two moments are:

$$E[\xi] = \sqrt{2\pi} \quad (7.10)$$

$$E[\xi^2] = 4 \left[1 + \frac{\pi}{4} \left(1 + \mu + \frac{\mu^2}{4} + \frac{\mu^3}{4} + \frac{\mu^4}{64} + \frac{49\mu^5}{2.304} + \dots \right) \right] \quad (7.11)$$

where $\mu = r^2 / 4$. The mean wave height is identical to the one of the Rayleigh distribution. This is reasonable because, as it has been stated, in mean terms (and surface elevation Gaussian distributed) the wave height is twice the wave amplitude. On the other hand, the root-mean-square height is, in general, lower. The use of the truncated series of Eq. (7.11) involves errors less than 0.3%. For $0 \leq r \leq 1$, $2.672 \leq H_{rms} / \sigma_\eta \leq 2\sqrt{2}$. $r=1$ implies the Rayleigh limit since the crest and trough are 100% correlated. On the contrary, $r=0$ implies the hypothetical case of statistical independency. In general, the distribution of crest-to-trough wave heights results in a diminution of the probabilities of low and high wave height whereas the heights around the mean value are enhanced (see Figure 7.2). Massel (1996) found from observations a mean value of $r=0.58$ (in the range $0.57 \leq r \leq 0.60$).

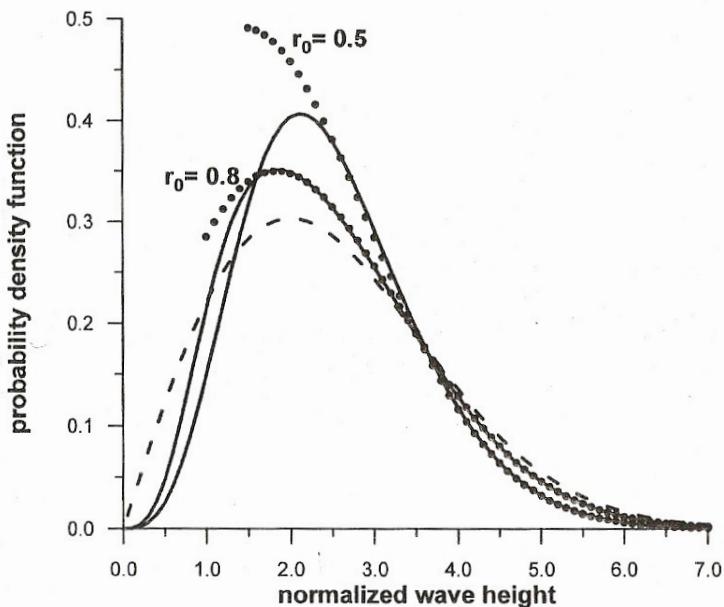


Figure 7.2 Comparison of various theoretical probability density function for large wave heights: ---- Rayleigh distribution; ____ crest-to-trough distribution for $r_0=0.5$ and $r_0=0.8$; Tayfun's distribution for large wave heights for $r_0=0.5$ and $r_0=0.8$ (Massel, 1996)

Though the simplification made for the wave period, Eq. (7.8) is still quite complicated and impractical (requiring numerical integration). Restricting the attention to higher waves which in fact are of special interest for engineers, the asymptotic expansion of the Bessel function can be used (Abramowitz and Stegun, 1965):

$$I_0(z) = \frac{e^z}{\sqrt{2\pi z}} \left(1 + \frac{1}{8z} + \dots \right) \quad (7.12)$$

By considering the first and second term of the above expansion, two upper and lower bounds of the pdf are found (Tayfun, 1990):

$$\text{From first term: } f_l = \frac{\xi}{2\sqrt{2r(1+r)}} \exp\left(-\frac{\xi^2}{4(1+r)}\right) \quad (7.13)$$

$$\text{From second term: } f_u = \left(1 + \frac{1-r^2}{2r\xi^2}\right) f_l \quad (7.14)$$

The approximation of the pdf is therefore the algebraic average of Eq. (7.13) and (7.14):

$$f(\xi) = \frac{\xi}{2\sqrt{2r(1+r)}} \left(1 + \frac{1-r^2}{4r\xi^2}\right) \exp\left[-\frac{\xi^2}{4(1+r)}\right] \quad \text{for } \xi > E[\xi] = \sqrt{2\pi} \quad (7.15)$$

By integrating Eq. (7.15), the corresponding probability of exceedance is:

$$P(\xi) = \left(\frac{1+r^2}{2r}\right)^{1/2} \left(1 + \frac{1-r^2}{4r\xi^2}\right) \exp\left[-\frac{\xi^2}{4(1+r)}\right] \quad \text{for } \xi > E[\xi] = \sqrt{2\pi} \quad (7.16)$$

In Figure 7.2, the agreement of such approximation with the more complicated initial expression is shown for $r = 0.5$ and $r = 0.8$. Tayfun (1990) also gives an approximation for the significant wave height. Actually, he derived a more general expression for $\xi_{1/p}$, the mean of highest $1/p$ th normalized wave heights:

$$\xi_{1/p} = (\xi_{1/p})_l + p \left(\frac{1-r^2}{8r}\right) \left(\frac{\pi}{2r}\right)^{1/2} \operatorname{erfc}\left(\frac{(\xi_p)_l}{2(1+r)^{1/2}}\right) \quad (7.17)$$

where $(\xi_{1/p})_l$ and $(\xi_p)_l$ are the solutions only considering f_l :

$$(\xi_{1/p})_l = (\xi_p)_l + p(1+r) \left(\frac{\pi}{2r}\right)^{1/2} \operatorname{erfc}\left(\frac{(\xi_p)_l}{2(1+r)^{1/2}}\right) \quad (7.18)$$

$$(\xi_p)_l = 2 \left((1+r) \ln \left[p \left(\frac{1+r}{2r} \right)^{1/2} \right] \right)^{1/2} \quad (7.19)$$

In Figure 7.3, variations of ξ_{rms} and $\xi_{1/p}$ with r are illustrated. The upper bounds are the values derived from Rayleigh theory. Note that the higher p , the difference between Rayleigh and crest-to-trough theory increases. In particular, for $r = 0.5$, the significant wave height ($p = 3$) is $3.75\sqrt{m_0}$, 6.3% lower than in the case of linear theory ($r = 1$).

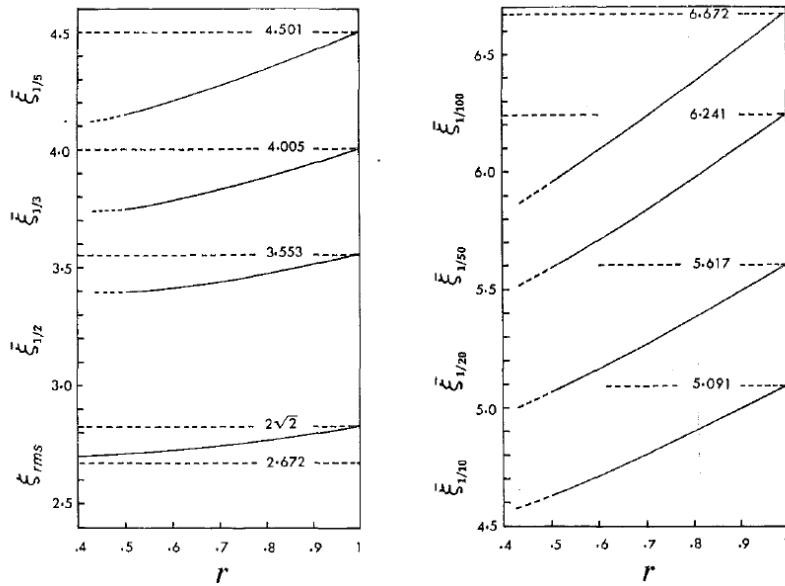


Figure 7.3 Variation of ξ_{rms} and $\xi_{1/p}$ with r , compared to the upper bounds values derived from Rayleigh theory
(Tayfun, 1990)

Tayfun (1990) found from observations, using the mean $r = 0.65$, that the crest-to-trough theory overestimates the mean wave height by about 4%, just as the Rayleigh theory does. For $p = 2$ he found that the discrepancy is reduced to 1.6 %, and, for $p = 3$ (significant wave height), even less: 1 %. For ξ_{rms} it is 1.67%. A further comparison provided by Osborne's (1982) simulation results, gives a discrepancy for the significant wave height of 0.53%. Tayfun (1990) pointed out that, in general, observations, vs. simulations, show less agreement because they come from different sea states at different times, each specified by a different spectrum and a different value of r . Actually, he suggested that the random character of r should be strictly taken into account:

$$\overline{\xi_{1/p}} = \int_0^r \xi_{1/p}(r) f_r(r) dr \quad (7.20)$$

$f_r(r)$ being the pdf of r . However, he also concludes that, although there is an apparent asymmetry of $f_r(r)$ (the median is slightly larger than the mean), using the mean value of r , the results are quite accurate.

As for the maximum wave height, Massel (1996) found that for a typical value of $r = 0.68$, the expected maximum wave height was reduced by approx. 7% compared to the Rayleigh distribution (for N , number of waves, from 200 to 10^5).

8. IMPROVEMENTS OF RAYLEIGH DISTRIBUTION

8.1. Introduction

In this Chapter, two different theories are presented. They are based on the Rayleigh theory of Longuet-Higgins (1952) with a modified root-mean-square wave amplitude. In both cases the Rayleigh distribution is scaled in such a way that the probability distribution for higher wave heights is reduced, to try to solve the often found overprediction problem of linear theory: Thompson (1974), Haring et al. (1976), Forristall (1978), etc. The probabilities density functions are presented in terms of the wave height variable. In order to calculate maximum values, the theory of maximum events can be applied as in the linear theory. The difference is only the scaling factor, different for each theory.

8.2. Modified Rayleigh distribution I (Longuet-Higgins)

Forristall (1978) fitted a Weibull distribution to 116 hours of hurricane-generated waves in the Gulf of Mexico. Later, Longuet-Higgins (1980) found that the same data could be adjusted to a Rayleigh distribution with a root-mean square wave height 7.5% lower than in the case of the classical linear theory. Supported by these results, Longuet-Higgins (1980) suggested that the Rayleigh distribution is still applicable with a properly chosen root-mean-square wave height. Actually, he derived an expression for such a parameter:

$$H_{rms}^2 = 8m_0 \left[1 - \left(\frac{\pi^2}{8} - \frac{1}{2} \right) \nu^2 \right] = 8m_0 [1 - 0.734\nu^2] < 8m_0 \quad (8.1)$$

in which ν is the bandwidth, shown in Eq. (4.24). By considering the above mentioned expression, the root-mean-square height is reduced and therefore the Rayleigh distribution is scaled, in principle, reducing its overprediction.

The derivation of the above mentioned formula follows from the assumption of the presence of background “noise” in the spectrum, outside the dominant peak. Firstly, Longuet-Higgins (1980) defined the surface elevation as a perturbed cosine wave:

$$y = \bar{b} \cos \bar{\sigma}t + \sum_n a_n \cos(\sigma_n t + \theta) \quad (8.2)$$

where all phases are random and the first term has a narrow spectrum in such a way that:

$$a_n \ll \bar{b} \quad \sigma_n^2 a_n \ll \bar{\sigma}^2 \bar{b} \quad (8.3)$$

After approximating the expression of Eq. (8.2) to second order (with Taylor series), Longuet-Higgins (1980) found that the mean-square value of a is:

$$\bar{a}^2 = 2m_0 + \int_0^\infty \left\{ (\sigma/\bar{\sigma}^2 - 1) - \cos^2 \left(\frac{\pi\sigma}{2\bar{\sigma}} \right) \right\} E'(\sigma) d\sigma \quad (8.4)$$

in which the first term is that linear theory and $E'(\sigma)$ is the spectrum of the perturbation. Assuming that the perturbation is considered near $\bar{\sigma}$ itself:

$$\overline{a^2}/2m_0 = 1 - \left(\frac{\pi^2}{8} - \frac{1}{2} \right) \nu^2 = 1 - 0.734\nu^2 \quad (8.5)$$

From Eq. (8.5), the expression of Eq. (8.1) follows directly assuming the relation $\underline{H} \approx 2\underline{\eta}_{crest}$.

8.3. Modified Rayleigh distribution II (Naess)

Naess (1985) found that, considering that crests and troughs are separated in time by a certain time lag Δt , similarly to the assumption of crest-to-trough distribution (see Chapter 7), the wave height can be described by the Rayleigh distribution with the root-mean-square wave height defined by Eq. (8.7). The difference with the crest-to-trough distribution is the following: instead of considering the joint probability distribution of two amplitudes separated by Δt , he defines the Gaussian stochastic process as the difference between two Gaussian process:

$$X(t) = \underline{\eta}(t) - \underline{\eta}(t + \Delta t) \quad (8.6)$$

and therefore H , the maximum of the Gaussian process, is (approximately) Rayleigh distributed.

$$H_{rms}^2 = 4 \left[K_\eta(0) - K_\eta(\Delta t) \right] \quad (8.7)$$

where $K_\eta(\tau)$ is the autocorrelation function of the surface elevation:

$$K_\eta(\tau) = E \{ \underline{\eta}(t) \underline{\eta}(t + \tau) \} \quad \text{for} \quad E \{ \underline{\eta}(t) \} = 0 \quad (8.8)$$

and E here means the expected value. The autocorrelation function can be found from the spectrum (\tilde{E}) with:

$$K_\eta(\tau) = \int_0^\infty \tilde{E}(f) \cos(2\pi f \tau) df \quad (8.9)$$

Eq. (8.9) shows that the variance is the zero-order spectral moment:

$$\overline{\eta^2} = K_\eta(0) = \int_{-\infty}^\infty \tilde{E}(f) df = m_0 \quad (8.10)$$

The value of $K_\eta(\Delta t)$ is approximated by the global minimum of $K_\eta(\tau)$. For the limit case of a sine wave, the time Δt becomes exactly half the period, in which the autocorrelation has a minimum (see Figure 8.1). However, in more general case, Δt can be approximated by half the mean period (T_m , see Eq. (4.21)).

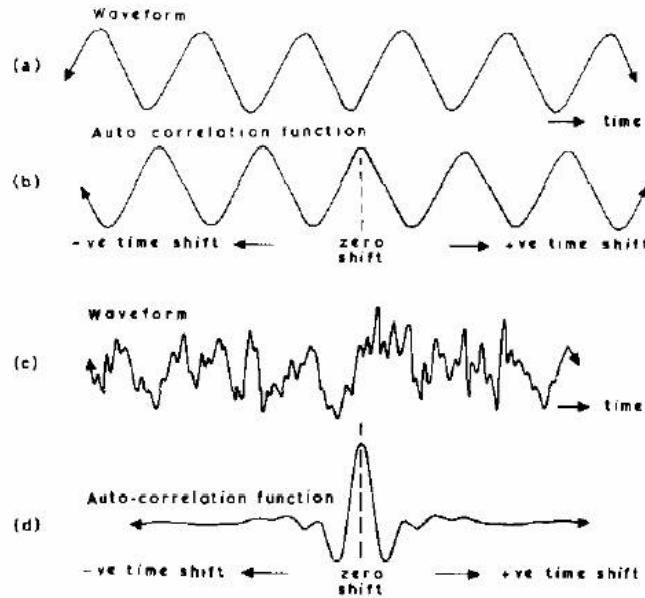


Figure 8.1 Relation between waveform and autocorrelation function: first, a perfect sine wave and, secondly, a real ocean wave form.

Therefore, Eq. (8.7) becomes:

$$H_{rms}^2 = 4m_0 [1 - r_{min}] \quad (8.11)$$

in which r_{min} is:

$$r_{min} = \frac{K_\eta(\tau)_{min}}{m_0} \approx \frac{K_\eta(T_m/2)}{m_0} \quad (8.12)$$

For the very narrow-banded processes, r_{min} becomes -1 and therefore the probability density function becomes the Rayleigh distribution.

8.4. Comparison

For a Pierson-Moskovitz spectrum, Longuet-Higgins (1980), with Eq. (8.1), found $H_{rms} = 1.86\sqrt{2}\sigma_\eta$. For the same spectrum, $r_{min} = -0.653$ (see Eq. (8.12)) which corresponds to $H_{rms} = 1.82\sqrt{2}\sigma_\eta$ (Massel, 1996). For a common value of $\gamma = 3.3$ of a Jonswap spectrum, $r_{min} = -0.73$, which is equivalent to say $H_{rms} = 1.86\sqrt{2}\sigma_\eta$ (see Eq. (8.12)). Using experimental data, Massel (1996) found a slightly lower value for the root-mean-square height: $H_{rms} = 1.78\sqrt{2}\sigma_\eta$. By linear fitting, Holthuijsen (2007) found from other observations that $H_s = 0.927H_{m_0}$ (see Chapter 5) which, assuming linear theory, is equal to say $H_{rms} = 1.85\sqrt{2}\sigma_\eta$, similar to above values.

9. CONSIDERATIONS FOR EXPECTED MAXIMA

Cartwright (1958) made a detailed theoretical analysis of the probability distribution of maximum crests. The results were applied to measurements containing some 10.000 waves. He started with linear theory in which he concluded that the asymptotic assumption produces a negligible error when $N > 128$. He concluded that the expression for the maximum crest for N waves can be calculated with the linear theory expression (Eq. (5.31)) but using, instead of N , an effective value of N :

$$N_{ef} = N(1 - \varepsilon^2)^{1/2} (1 - \alpha) \quad (9.1)$$

in which ε is the spectral bandwidth introduced in Chapter 4 (Eq. (4.25)) and α an autocorrelation parameter between waves. The explanation of the two factors in Eq. (9.1) multiplying N are explained separately:

- Inclusion of bandwidth parameter: $(1 - \varepsilon^2)^{1/2}$

Cartwright does not actually consider the Rayleigh distribution of linear theory explained in Chapter 5. He uses the most general formula for the exceedance probability of crest height (Cartwright & Longuet-Higgins, 1956):

$$P_{exceed}(\eta_{crest}) = \frac{1}{\sqrt{2\pi}} \left[\int_{\eta/\varepsilon}^{\infty} \exp\left(-\frac{t^2}{2}\right) dt + \sqrt{1 - \varepsilon^2} \exp\left(-\frac{\eta^2}{2}\right) \int_{-\infty}^{\eta\sqrt{1-\varepsilon^2}/\varepsilon} \exp\left(-\frac{t^2}{2}\right) dt \right] \quad (9.2)$$

As the spectral width of the spectrum decreases, ε tends to zero and Eq. (9.2) becomes the Rayleigh exceedance probability. In the limit case of $\varepsilon = 1$, Eq. (9.2) becomes the Gaussian probability of exceedance. According to Eq. (9.2), the non exceedance probability for the maximum crest is, using the asymptotic expression:

$$F(\eta_{max\ crest}) \approx \exp\left[-\left(N(1 - \varepsilon)^{1/2} \exp\left(-\frac{\eta^2}{2}\right)\right)\right] \quad (9.3)$$

Eq. (9.2) accounts for the spectral width of the spectrum, apparently solving the overprediction problem of linear theory for the wave crest/trough. However, the definition of the crest in above equations is not the maximum crest per wave; it includes all local maxima. Therefore, the distribution of Cartwright & Longuet-Higgins (1956) has not been studied and used in the present study for calculations related to the crest since the definition of the crest variable is different. Moreover, note that the spectral bandwidth parameter ε depends on high-order spectral moments, which are very sensitive to the tail of the spectrum and therefore more susceptible to the presence of noise.

Nevertheless, when one looks to the maximum crest per N waves, it does not matter if the original definition of the crest is one of local maxima or only the maximum per wave. The maximum for N waves is still the same. Therefore, for the calculation of the expected maximum crest, it is important to have in mind the $(1 - \varepsilon)^{1/2}$ factor.

- Inclusion of autocorrelation parameter: $(1 - \alpha)$

The expected maximum in linear theory has been developed on the assumption that the N waves are mutually independent. In practise, each 3-4 adjacent waves are correlated (Cartwright, 1958). Then, the effective number of waves is reduced. Actually, the assumption of independent waves is inconsistent with the assumption of narrow-band spectrum and it becomes more acceptable for the wide-band case. In the present study, the detailed explanation of $(1 - \alpha)$ is not done because Cartwright (1958) found that such affection is practically null.

Cartwright (1958) found empirical agreement between the observed maximum crest and the expected value using linear theory but with N_{ef} instead of N . Figure 9.1 qualitatively illustrates, exaggerating, the reduction of the expected value compared to classical linear theory. From observations, he found $\varepsilon = 0.52$, $\alpha = 0.08$ for $N = 50$ and $\alpha = 0.01$ for $N = 10000$. Therefore, the effect of the autocorrelation is little, implying that the assumption of mutual independence is acceptable. Naess (1985) found that, considering that the correlation between waves is the same as between wave heights, it affects the expected maximum wave height less than 1%.

Figure 9.2 shows the reduction of the expected maximum crest due to the factor $(1 - \varepsilon^2)^{1/2}$, considering $\varepsilon = 0.5$, $\varepsilon = 0.7$, $\varepsilon = 0.9$. The discrepancy is high at low values of N and becomes smaller for larger N since, at this stage, the linear theory has little variation with N .

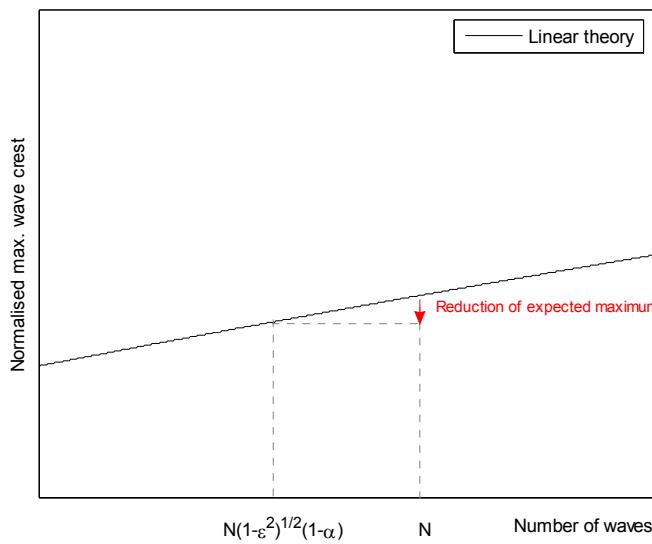


Figure 9.1 Illustration of the (exaggerated) reduction of the expected maximum due to the band width character and the autocorrelation between adjacent waves

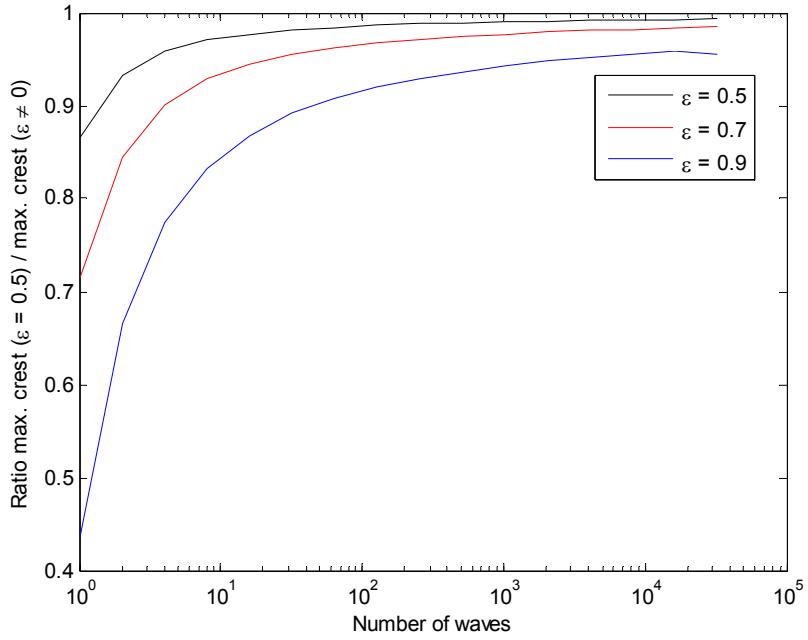


Figure 9.2 Reduction of the expected maximum crest compared to the linear theory. For this picture the values in the Appendix 1 made by Cartwright (1958) has been used.

For the prediction of highest waves, Vinje (1989) proposed to correct the probability distribution of Naess (1985), see Chapter 8, by adding a correction factor multiplying the distribution. However, it presents the drawback of assigning probabilities greater than 1 for small normalised wave heights and this does not make sense.

10. SUMMARY OF PROBABILITY DISTRIBUTIONS

Previous to the analysis of the data, in the present Chapter a summary of the explained theories in the present study is made. Figure 10.1 tries to clearly summarise these distributions, also reminding us what their assumptions are.

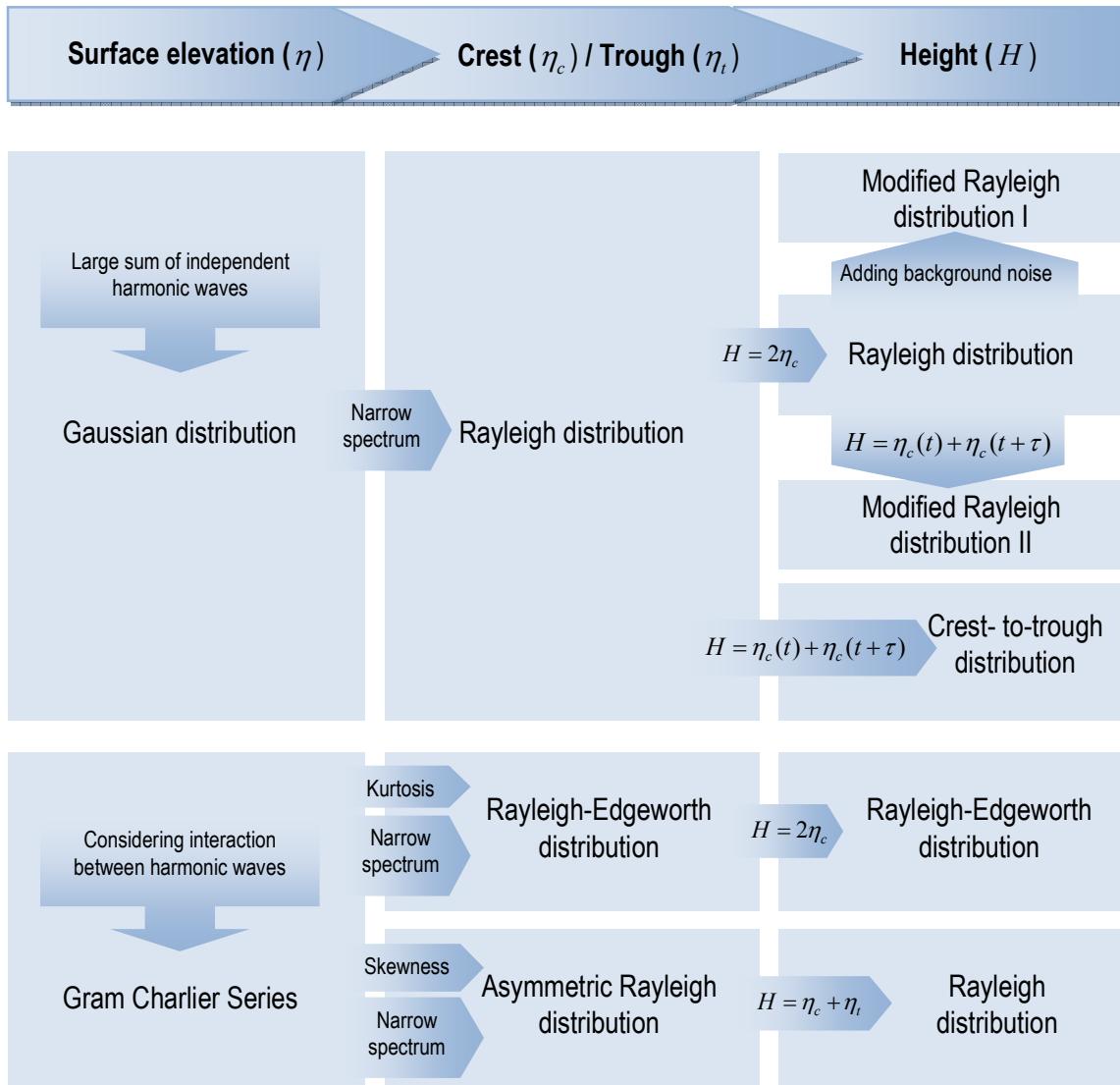


Figure 10.1 Overview of probability distributions of different ocean wave's parameters

In addition, as explained in Chapter 9, for the maximum variable a reduction of the number of waves should be made by the inclusion of the band-width parameter.

Below are listed the main expressions for the probability of exceedance for each distribution (for further details, see the previous chapters) in which $h = H / \sqrt{m_0}$, $\xi^+ = \eta_{crest} / \sqrt{m_0}$ and $\xi^- = \eta_{trough} / \sqrt{m_0}$:

- Rayleigh

$$P(h) = \exp\left(-\frac{h^2}{8}\right) \quad (10.1)$$

- Rayleigh-Edgeworth

$$P(h) = \exp\left(-\frac{h^2}{8}\right) \left[1 + \frac{1}{384} \kappa_{40} h^2 (h^2 - 16) \right] \quad (10.2)$$

- Asymmetric Rayleigh

$$P(\xi^\pm) = \exp\left(-\frac{\xi^{\pm 2}}{2}\right) \left[1 \pm c_0 \lambda_3 \pm \frac{c_1 \lambda_3 \xi^\pm}{\sqrt{2}} \left(\frac{\xi^{\pm 2}}{2} - \frac{1}{2} \right) \mp c_0 \lambda_3 \operatorname{erfc}\left(\frac{\xi^\pm}{\sqrt{2}}\right) \exp\left(\frac{\xi^{\pm 2}}{2}\right) \right] \quad (10.3)$$

- Crest-to-trough

$$P(h) = \exp\left(-\frac{h^2}{4(1+r)}\right) \left(1 + \frac{1-r^2}{4rh^2} \right) \left(\frac{1+r^2}{2r} \right)^{1/2} \quad (10.4)$$

- Modified Rayleigh I

$$P(h) = \exp\left(-\frac{h^2}{8} \left[1 - \left(\frac{\pi^2}{8} - \frac{1}{2} \right) r^2 \right]^{-1}\right) \quad (10.5)$$

- Modified Rayleigh II

$$P(h) = \exp\left(-\frac{h^2}{4(1-r_{\min})}\right) \quad (10.6)$$

11. THE MAXIMUM WAVE HEIGHT PARADOX

11.1. Introduction

As explained in Chapter 5, the observations of Longuet-Higgins (1980) and Holthuijsen (2007) show that the theoretically estimated significant wave height from the spectrum should be reduced by a factor of 5%-10% in order to properly represent the real value. Moreover, some observations also show that the scaled Rayleigh distribution agree with the observed one (see Figure 5.3). This discrepancy can be expected since the Rayleigh distribution (R), and therefore H_{m_0} (calculated as $4\sqrt{m_o}$), is derived with some assumptions (linear theory, narrow-band spectrum...) which in reality do not hold.

What it was surprising is that in some observations the estimation of the maximum wave crest agrees with the real value (see Figure 11.1Figure 11.2). In this estimation, instead of the Rayleigh distribution, the distribution of Cartwright & Longuet-Higgins (1956) is used, which considers the maxima of all local maximum crest (see Chapter 9). However, the difference with the Rayleigh distribution is less than 2% for $N > 30$ approx., which is less than the 5-10% found discrepancy in the significant wave height. Therefore, at first sight, it seems that the extreme crest heights do not seem to suffer from the same scale discrepancy as the wave heights.

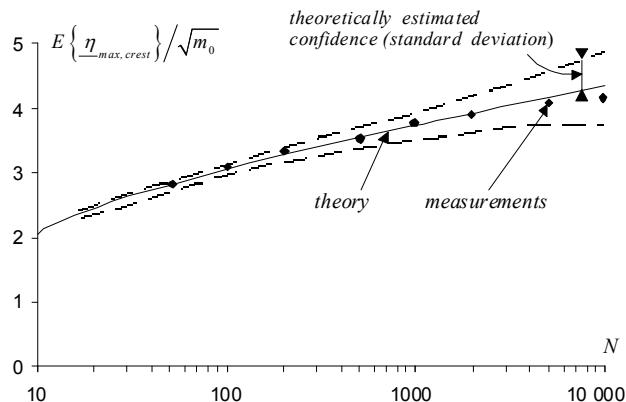


Figure 11.1. The observed and theoretically estimated expected value of the maximum crest height in a duration of N waves (Cartwright, 1958)

11.2. Possible reasoning

The paradox was meant to be resolved by considering the nonlinear effects and more exactly the Rayleigh- Edgeworth distribution (RE) with includes a parameter named BFI (see Chapter 6). Figure 11.2 compares the R and RE pdf (the wave height has been standardised by the significant wave height calculated as $4\sqrt{m_o}$). The expectations were that the wave height was overpredicted by the linear theory due to the wide-band spectrum, including the assumption $H \approx 2\eta_{crest}$. Nonlinearities presumably enhanced the extreme waves, counteracting this scaling

(as previously mentioned, the Rayleigh-Edgeworth distribution also assumes a narrow band spectrum).

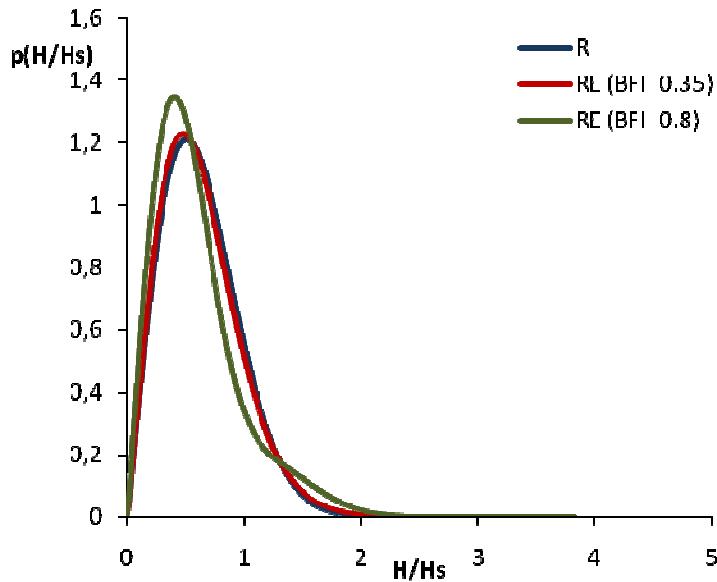


Figure 11.2 Comparison of Rayleigh distribution (R) with Rayleigh Edgeworth (RE) for two different values of BFI.

With the RE distribution, the probability of higher and low wave heights is enhanced whereas in the mid range wave height it is reduced. The estimated maximum wave height becomes higher and the significant wave height practically remains the same (it is slightly higher). In addition, the higher the BFI, the more pronounced such effects are, but with a larger number of waves, the estimated maximum wave heights for R and RE become closer (see Table 11.1). Therefore, by using the RE distribution, it is expected to find more or less the same discrepancy between estimations and observations in both significant wave height and maximum wave height. In such a case, there would be overprediction but it would be consistent in the sense of being present in both parameters and perhaps only a scale factor should be applied.

Table 11.1 Comparison R-RE

| Relation RE/R | BFI=0.35 | BFI=0.8 |
|---------------|----------|---------|
| H_{mean} | 0.99 | 0.95 |
| H_s | 1.015 | 1.026 |
| $N=1,000$ | 1.072 | 1.175 |
| $E(H_{max})$ | 1.078 | 1.160 |
| $N=10,000$ | 1.078 | 1.145 |
| $N=100,000$ | 1.078 | 1.145 |

For BFI=0.35, the percentage of enhancement of the maximum wave height by the RE distribution (7%) is more or less the same as the discrepancy found in the significant wave height (5-10%) between observations and the Rayleigh distribution.

However, after the analysis of the data in the following chapters, the scope of this study has been changed and it seems that the discrepancy is between wave height and wave crest and not between significant and maximum wave height.

12. ANALYSIS OF THE MEDITERRANEAN DATA

12.1. Introduction

After having presented some of the most important current theories, different aspects of the data are analysed.

As will be illustrated in the current chapter (Section 12.4), the nonlinearities are weak. Therefore, in the first part of the presentation of the analysis, the wave height and wave crest/trough are analysed, omitting the nonlinear theories explained in Chapter 6. For each variable, the significant and maximum height/crest/trough is calculated and compared to the linear theory expectations and, if necessary, to other theories (Chapters 7, 8 and 9).

Finally, the nonlinearities are analysed with the kurtosis and skewness parameter as well as the spectral estimator of the kurtosis: BFI.

Note that often the variables such wave heights or crests are normalized. It always means that they are divided by the standard deviation.

12.2. Wave height

12.2.1. Significant wave height

The observed and estimated normalized significant wave heights, according to the linear theory ($H_{s,m_0} = 4\sqrt{m_0}$), are compared for all records in Figure 2.1. The calculated slope is from the straight line which fits the data best (in red), using the least square method. In the calculation of such a line, the condition of passing through the origin has been established; otherwise it would be nonsense. The black line represents the perfect agreement.

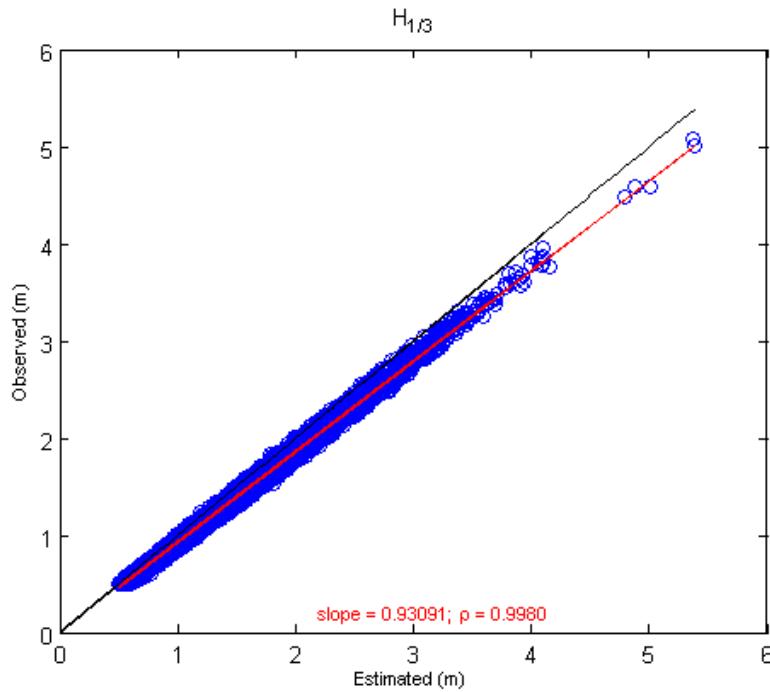


Figure 12.1 Comparison between observed and estimated significant wave height (Rayleigh theory)

As illustrated in Figure 12.1, according to the fitted straight line, $H_{s,obs} \approx 0.931H_{s,m_0}$. The observations are 7% lower than the estimations which agrees with the values found by Longuet-Higgins (1980), Holthuijsen (2007) and Forristall (1978) (see Chapter 5). Apart from this result it is important to note that there is little deviation from the straight line, meaning that the scaling factor does not seem to depend on the magnitude of the significant wave height. That may lead to thinking that the use of a constant scale factor for the Rayleigh distribution is quite acceptable:

$$p(H) = \frac{H}{4f^2 m_0} \exp\left(-\frac{H^2}{8f^2 m_0}\right) \quad (12.1)$$

in which $f = 0.931$ for this data set. In Figure 12.2, the Rayleigh distribution is compared with a hypothetical scaled one, using the discrepancy found in the significant wave height between estimations and observations, Eq. (12.1).

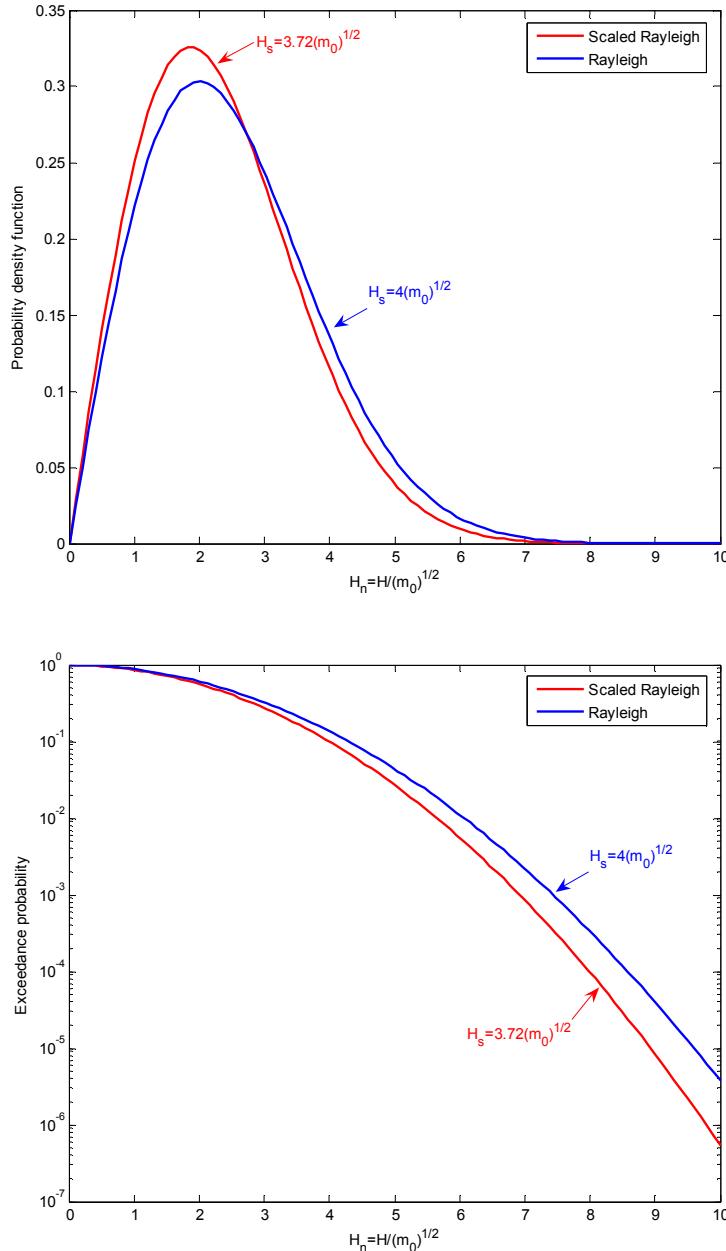


Figure 12.2 Comparison between scaled Rayleigh distribution (according to the found discrepancy in the observed significant wave height) and the original one. Up: probability density function; down: exceedance probability

Although the value of the scaling factor found for this data set (0.931) seems to be a more or less “universal” constant compared to Longuet-Higgins (1980), Holthuijsen (2007) and others, it would be better to try to find such a scaling factor from a theoretical background instead of empirical. Remark that the standard deviation of the ratio $H_{1/3}/H_{m0}$ has a standard deviation of 0.0262. The mean of the same ratio is 0.929, very close to the slope of the fitted straight line.

Therefore, it is interesting to compare the observations with the theories explained in Chapter 8, the first accounts for a certain background noise (Modified Rayleigh I; Longuet-Higgins, 1980) and the second accounts for a certain time lag between crest and trough (Modified Rayleigh II; Naess, 1985). In the Figure 12.3 and 12.4, the estimated value for the significant wave height is slightly modified according to such theories.

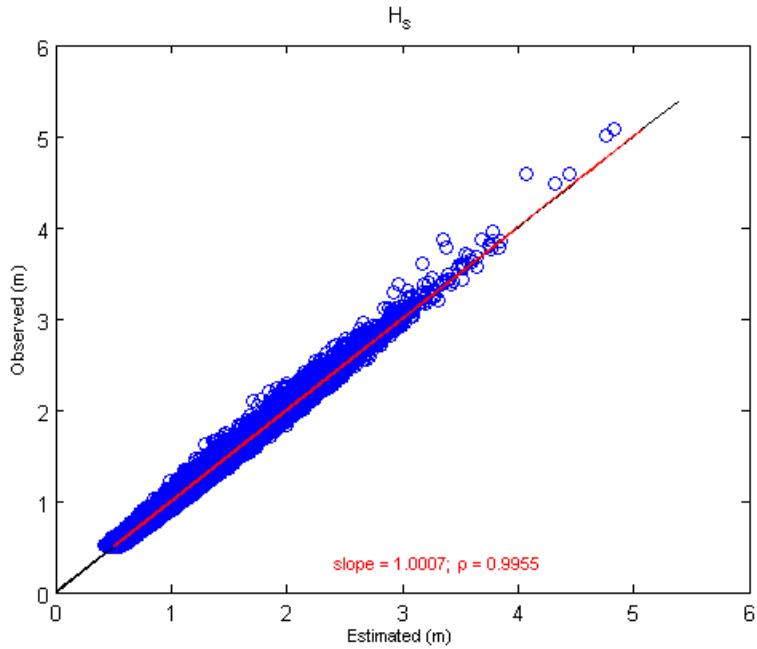


Figure 12.3 Observed and estimated significant wave height with the Modified I theory

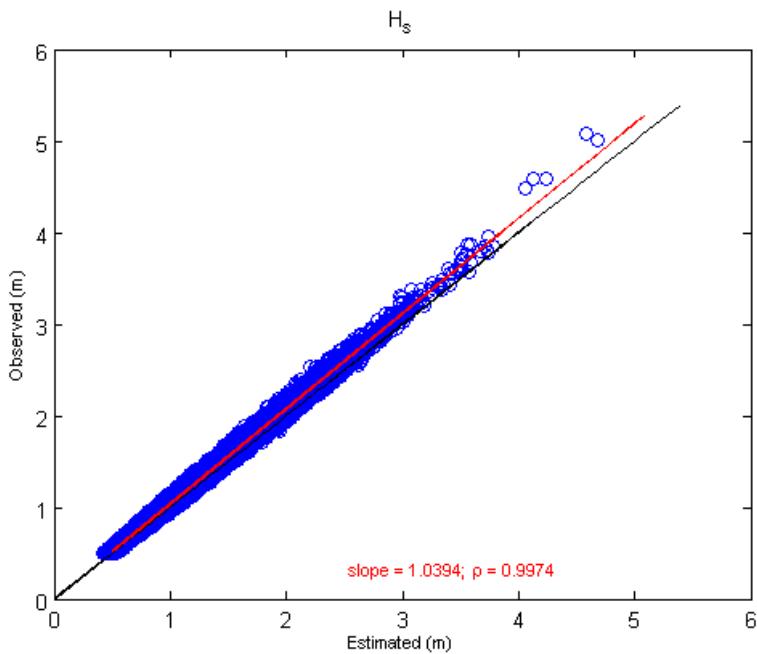


Figure 12.4 Observed and estimated significant wave height with Modified II theory

In the second theory, the minimum of the autocorrelation function has been calculated but if we calculate the autocorrelation function for half mean period (Eq. (4.21)), the results do not differ significantly; the fitted slope is 1.0448 instead of 1.0394, reducing the computational cost.

Both theories, concerning the significant wave height, are better adjusted than the linear theory but the first one is clearly the best (1.00 vs. 1.04) although the scatter is higher. Note that in both cases the involved parameters (respectively, ν and r_{min}) have been calculated for each record instead of considering the mean values, which are: $\nu = 0.41$ and $r_{min} = -0.57$.

The scaling produces an enhancement of the probability of the low-mid range whereas the higher waves are reduced (see Figure 12.2). Other parameters than the significant wave height, like mean wave height, root-mean-square wave height and estimated maximum wave height are reduced by the same scaling factor. Therefore, if the scaled Rayleigh distribution was the solution, the same discrepancy would be found in the other observed parameters. Figure 12.5 shows the discrepancy found for H_{mean} and H_{rms} compared to the linear theory. Despite not being exactly the same, the differences are not very large. However, there is a certain tendency in which the more the parameter is related to higher waves, the larger the scaling factor becomes. Actually, Massel (1996) stated that the Rayleigh distribution overpredicts the higher waves in a record and the error increases toward the low-probability tail of the distribution. Such a tendency is confirmed in Section 12.2.2, in which it is found that the scaling factor associated to the maximum wave height is even higher.

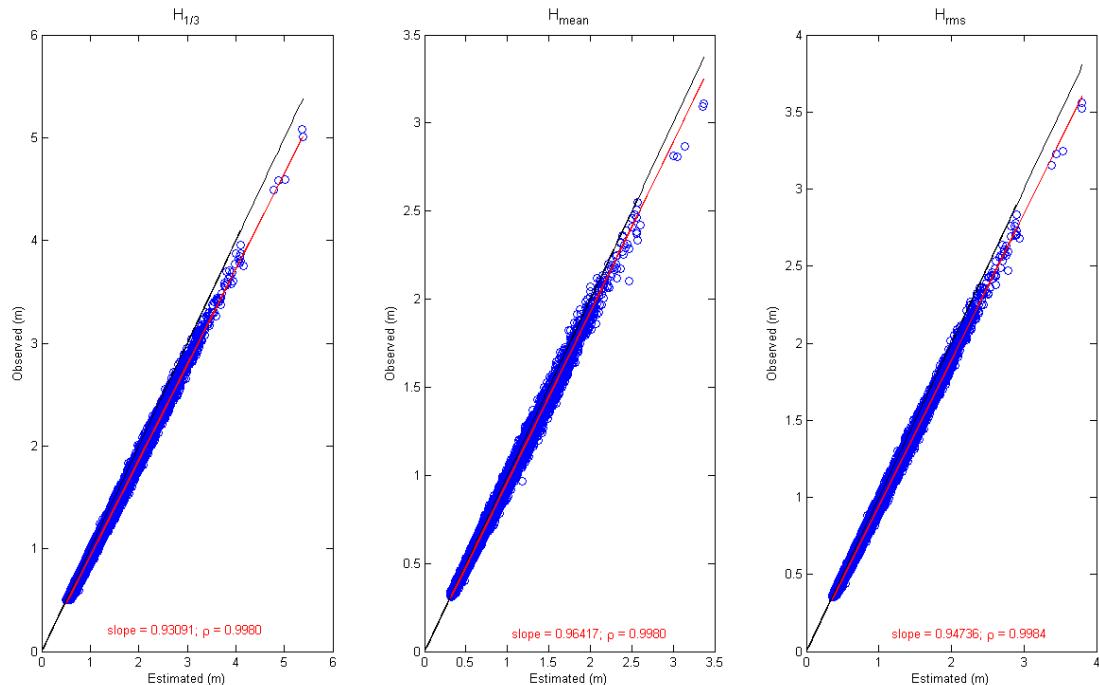


Figure 12.5 Discrepancy between Rayleigh theory and observations of: significant wave height, mean wave height, root-mean-square wave height.

Therefore, it becomes appropriate to compare these observed parameters with the crest-to-trough distribution (see Chapter 7) which does not simply rescale the Rayleigh distribution. The discrepancy for the significant wave height is illustrated in Figure 12.6. Although Tayfun (1990)

suggested using the mean value over all records for the autocorrelation as a simplification, here, for each record, the autocorrelation value has been calculated (from the spectrum) and used. The mean value is $r = 0.62$, very similar to the one obtained by Tayfun (1990): 0.65; and the median is also higher: 0.64, implying a certain asymmetry of distribution of r (see Figure 12.7).

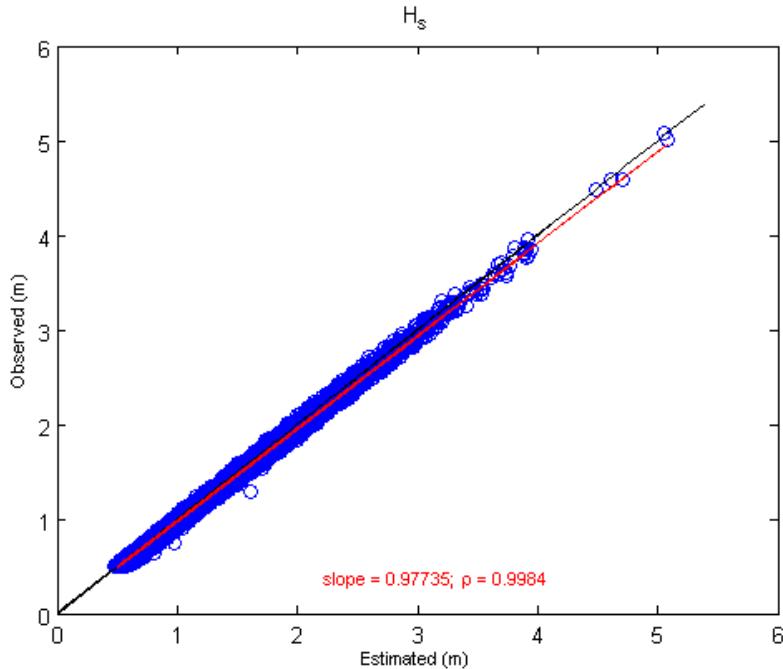


Figure 12.6 Observed and estimated significant wave height with crest-to-trough theory

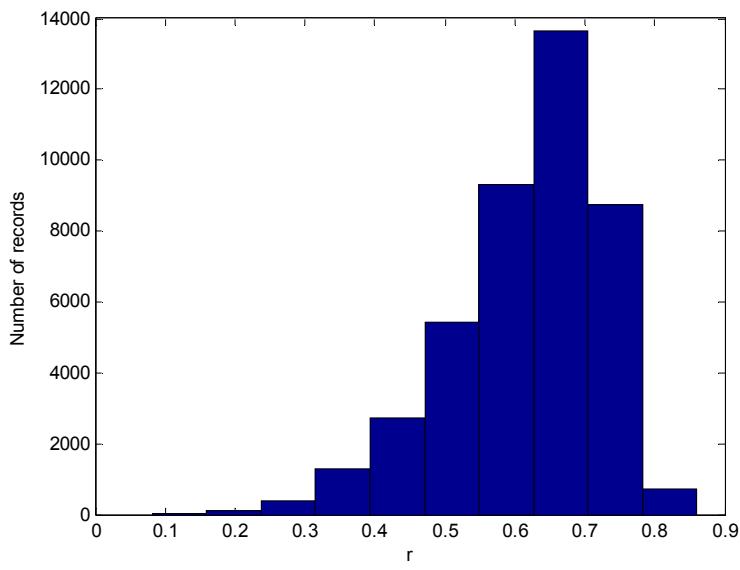


Figure 12.7 Histogram of the autocorrelation factor r of the crest-to-trough theory

In conclusion, the found scaling factor with the significant wave height is in agreement with other found in the literature. However, the disagreement is slightly higher as one looks at higher

waves related to low probabilities, suggesting that the overprediction cannot be simply solved by considering a scaled Rayleigh distribution as in Modified Rayleigh distribution I and II (Chapter 8). Therefore, the crest-to-trough theory is used but the observed significant wave height is still lower than the one predicted by this theory.

Note that the possible reasons for the Rayleigh overprediction are:

a) Spectral width, comprising of two aspects:

- i) In a wide spectrum, the calculated wave envelope does not vary slowly with time and overpredicts wave crests (see Figure 5.2) and therefore wave heights too. In particular the bimodal character of some spectra is one of the most important factors. Rodríguez et. al (2002) pointed out that the increase of the intermodal distance in a bimodal spectrum gives rise to the generally observed overprediction of large wave heights by the Rayleigh distribution.
- ii) Although the wave envelope properly predicted wave crests and troughs, the assumption of estimating the wave height as twice the crest would produce overprediction for the statistics of large wave heights. This assumption, at the same time, assumes: symmetric wave envelopes, the crest having the same value as the trough. In the simplified example below, it is explained why the random asymmetry of waves produces overprediction of large wave heights, whereas the mean is not overpredicted.

| Height | Crest / Trough |
|---|---------------------------------|
| $H_{mean} = H + h = 2\eta_{c/t,mean}$ | $\eta_{c/t,mean} = (H + h) / 2$ |
| $H_{1/2} = H + (H + h) / 2 < 2\eta_{c/t,1/2}$ | $\eta_{c/t,1/2} = H$ |

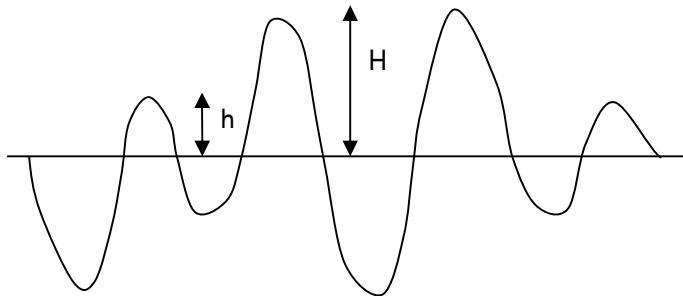


Figure 12.8 Sketch of random asymmetry in waves

- b) The buoy may underestimate extreme waves. In fact, in the WADIC project (1989), they conclude that the WAVERIDER tend to underestimate the wave height.

Errors related with instrumentation probably exist. However, with the present information it is difficult to quantify them and therefore only attention to the first ones will be paid. The crest-to-trough distribution, which is theoretically better than the other analysed above, tries to solve the problem by considering a certain time lag between crest and trough. For i), more research is needed since there is only one fully developed theory relating the wave height and spectral bandwidth: Longuet-Higgins & Cartwright (1956), but it is derived for all local maxima (for the crest)

and not the global maxima that we are interested in. In addition, ε is a very sensitive parameter to the high frequency tail of the spectrum.

12.2.2. Maximum wave height

Apart from the significant wave height, the expected maximum wave height is also an important parameter. Note that the maximum is not a parameter indirectly involved in the distribution of the wave height but a random variable itself. Therefore, in order to properly compare the observed value with the expected value of the linear theory, an averaging process is necessary. Firstly, in a way analogous to the other parameters, the observed maximum wave height is normalized by the standard deviation of each record. This maximum only depends on the number of waves in a record. All the records have the same order of magnitude of waves (250-300 waves) and they have been concatenated to create longer records with higher number of waves. Considering that there are approximately 40.000 records, the total amount of waves (N_T) is about 10 million. For example, for calculating the observed maximum height which corresponds to n waves, firstly, the records are concatenated in such a way that each group has approximately n waves. The maximum of each group is taken. There are approximately N_T / n maximum heights (associated to n waves) and the average of them is the final associated maximum to n waves, obtained from observations.

The fact of having a large amount of data here plays an important role. In general, having more observations give more reliability to the empirical results. Moreover, this particular analysis allows the studying what happens when considering large values of N . According to linear theory, the values of the maxima increase only slowly with increasing N because of this dependency:

$$E_N \{ H_{n\max} \} \approx 2 \left(1 + \frac{0.29}{\ln N} \right) \sqrt{2 \ln N} \quad (12.2)$$

Figure 12.9 illustrates the above results, using a logarithmic scale for N . Once more, as for the significant wave height, the estimations of linear theory overpredict the observations but now (for the maximum wave height) the scaling factor is noticeably lower: 0.863 instead of 0.931. The tendency of a lower scaling factor (higher discrepancy) for higher heights is corroborated (see Section 12.2.1). Forristall (1978) found a slightly higher value for the maximum wave in $N = 1000$ of 116 hours of hurricane waves generated in the Gulf of Mexico: 0.907. On the contrary, Earle (1975) found that the maximum wave height was well predicted by the linear theory but he only analysed records of 200 waves and, instead of considering the expected value of the maxima, he calculated the 50 percentile value, which for 200 waves is approx. 2% lower than the expected one and he used the observed significant (7% lower than the Rayleigh one). So, in fact, Earle (1975) did not exactly use the Rayleigh distribution. However, this 9% (roughly 2% plus 7%) difference does still not explain why the present discrepancy is higher (14%). Sobey et al (1990) also compared the highest wave heights in a standard 20 minute record (approx. 300 waves) from tropical Cyclone Victor using linear theory. His data showed a systematic overprediction of 10%, practically the same as the discrepancy obtained in the analysis of 1000 waves of hurricane-generated waves in the Gulf of Mexico (1978). Despite 10% being higher than the 7 % commonly found value for the significant wave height, it is still lower than the present discrepancy for the maximum wave height: 14%. The maximum wave height is 8.52 m.

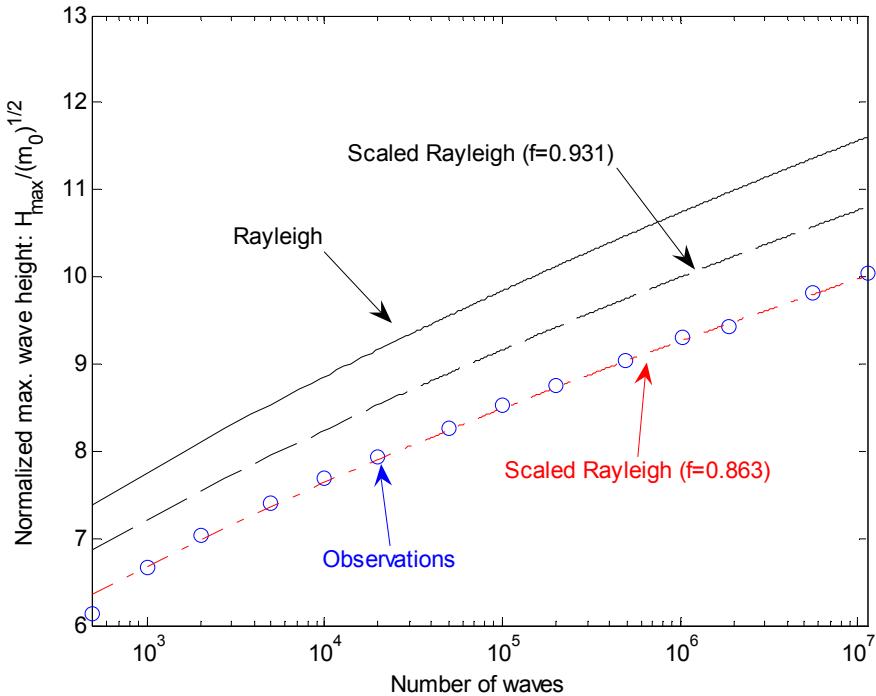


Figure 12.9 Mean normalized maximum wave height for different number of waves

Using the crest-to-trough theory, Massel (1991) calculated the reduction of the maximum wave height (compared to linear theory) for various sample sizes considering the crest-to-trough distribution. It is about 0.93: a reduction of 7% is not high enough to explain the discrepancy found in the observed mean maximum wave height. For the Modified Rayleigh I and II distribution, the discrepancies for the maximum wave height are, respectively, 7% and 4%.

As explained in Chapter 9, one can include the effect of the spectral band width in the calculation of the maxima by considering a reduced effective number of waves (Cartwright, 1958). The spectral band width used is the mean of all the observations: $\varepsilon = 0.69$. The reduction of the expected maximum wave height is practically constant (varying from 2.5 to 1%). Remind in this analysis $N > 500$ and that the reduction was more variable for $N < 100$, see Chapter 9.

Then, one could roughly approximate the estimated overprediction for the maximum wave height, in linear theory, as the sum of the two obtained previously (from crest-to-trough distribution and from considering an effective number of waves) achieving a discrepancy of 8-10.5%. This discrepancy agrees with the results of Sobey et al. (1990) and Forristall (1978) presented above but it does not explain the 14% discrepancy in the Mediterranean data.

Finally, it is important to note that the quality control explained in Chapter 2 has been crucial in order to have reliable results. More precisely, in the analysis of the maximum wave height the effects are noticeable. For example, without rejecting the data with significant wave height lower than half meter, the observations show a large enhancement for large N (see Figure 12.10). The effects for low N are practically null, probably because of the averaging process. The reason is not well-known.

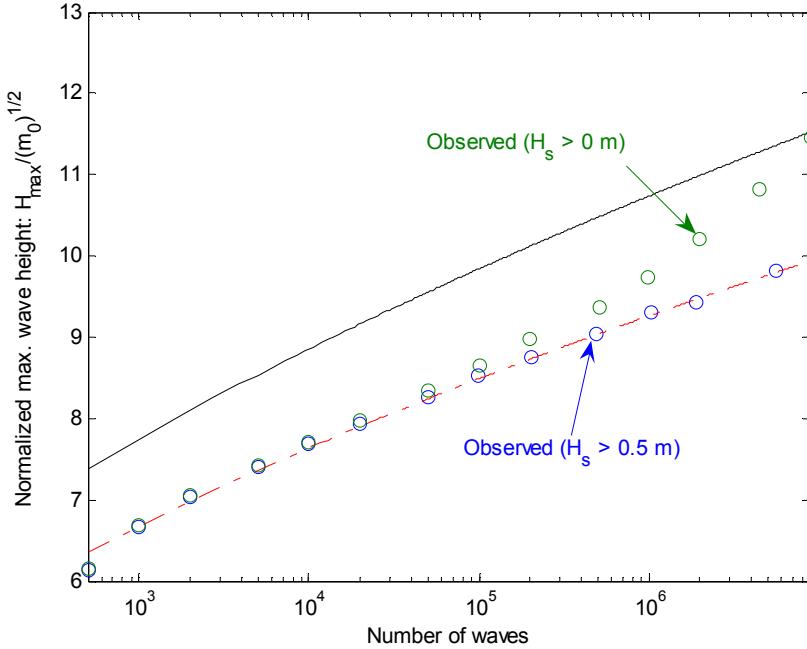


Figure 12.10 Effects of some conditions of the quality control in the maximum wave height (for more details see text above)

12.2.3. Summary of results

In Table 12.1, the observations/expectations ratios for some of the theories considered in the present study (those which assume surface elevation Gaussian distributed) are summarized:

Table 12.1 Relation observations/expectations and % of discrepancy, according to different theories

| Parameter | Rayleigh | Modified Rayleigh I (Longuet-Higgins) | Modified Rayleigh II (Naess) | Crest-to-trough |
|------------------------|-------------|--|---------------------------------|-----------------|
| H_{mean} | 0.96 - 4 % | 1.04 + 4 % | 1.08 + 8 % | 0.96 - 4 % |
| H_{rms} | 0.95 - 5 % | 1.02 + 2 % | 1.06 + 6 % | 0.98 - 2 % |
| H_s | 0.93 - 7 % | 1.00 0 % | 1.04 + 4 % | 0.98 - 2 % |
| H_{max} | 0.86 - 14 % | 0.93 - 7 % | 0.96 - 4 % | 0.93 - 7 % |
| $H_{max}(\varepsilon)$ | - | - 12 % | - 5 % | - 2 % |
| | | | | - 5 % |

The modified Rayleigh I is the one with less discrepancy for all the studied parameters. In second place, there is the crest-to-trough distribution. However, bear in mind that the Modified Rayleigh I theory also considers a scaling factor for the wave crest distribution. In Section 12.3, it will be shown how the crest does not suffer from the same discrepancy as the height. Therefore, the crest-to-trough, becomes the “best” theory although its results are not convincing. In fact, Tayfun (1990) recognized that, from observations, a systematic overprediction of 4% for the mean wave height was found, just as in linear theory; the same is obtained in the present study.

For the maximum wave height there is an important overprediction for all the theories. With $H_{max}(\varepsilon)$ one wants to denote that the estimated values for the maxima are obtained considering

the influence of ε . Remember that the inclusion of ε has to do with the fact that the considered wave amplitude in the linear theory is larger than the real amplitude. Such an influence has been roughly calculated, using the effective number of waves (see Chapter 9), by subtracting 2 % from all discrepancies of the maxima (in absolute value). However, the remaining discrepancy is still considerable.

12.3. Wave crest/trough

12.3.1. Significant wave crest/trough

The significant crest/trough is computed according to linear theory: $\eta_{crest,s} = \eta_{trough,s} = 2\sqrt{m_0}$ and then compared with observations, as well as mean and root-mean-square values. Figure 12.11 and 12.12 illustrate the results. Compared to the wave height, both crest and trough show a better agreement with linear theory: the scaling factor for the significant value is, in both cases, around 0.98 (for the crest slightly higher than the trough), and there is no evident tendency of higher scaling factor for statistics related to larger heights.

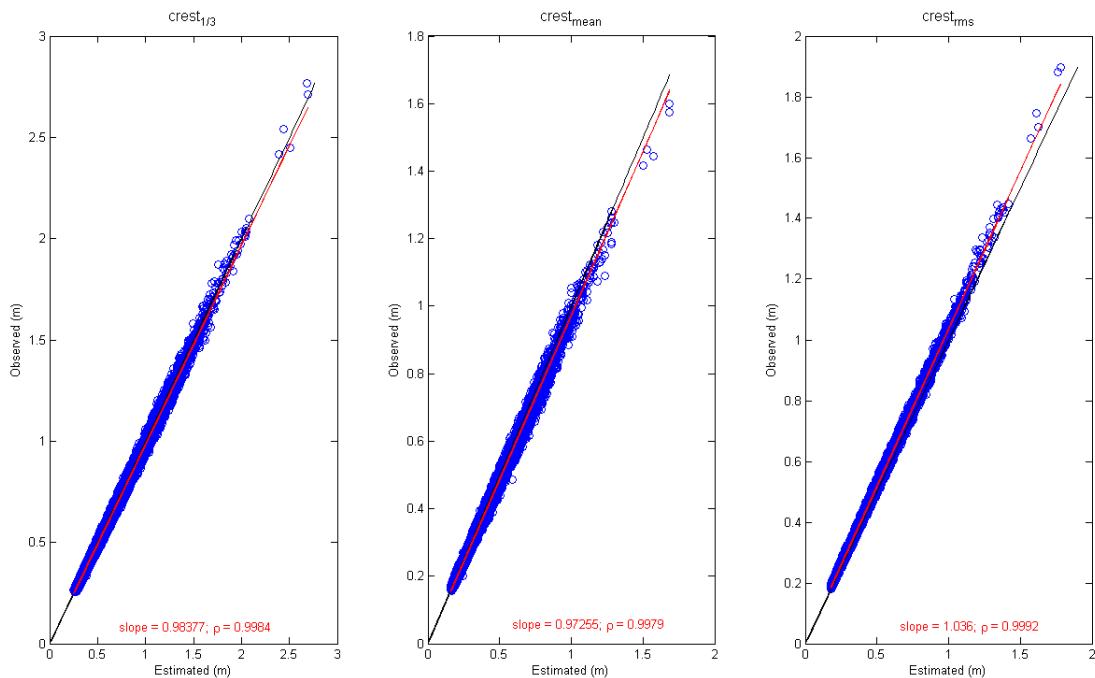


Figure 12.11 Comparison between observed and estimated parameter of wave crest (linear theory)

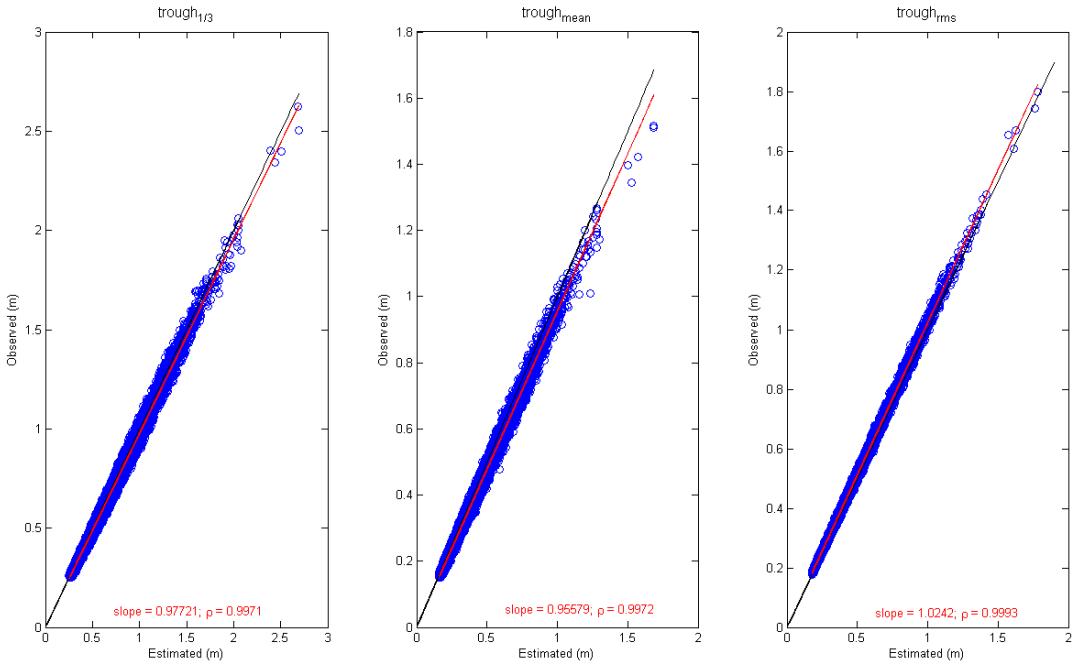


Figure 12.12 Comparison between observed and estimated parameter of wave trough (linear theory)

Surprisingly, the discrepancies found for the crest and trough are little, above all in the case of the crest, suggesting that the Rayleigh distribution is quite acceptable. This agrees with the crest-to-trough theory for the wave height (which assumes Rayleigh distribution for, respectively, crests and trough) whereas the Modified Rayleigh I becomes worse.

12.3.2. Maximum wave crest/trough

Following the same procedure as in the calculation of the maximum wave height, the maximum crest/trough is computed. In Figure 12.13 the results are compared with linear theory expectations. The scaling factor is, respectively, for the crest: 0.991, and for the trough: 0.962.

Therefore, in contrast to the results of the maximum wave height, the maximum crest and trough are reasonably well predicted by the linear theory. In fact, the maximum wave crests corresponds well with the results of Cartwright (1958, see Figure 11.1) Although he considered smaller values of N ($N < 10^4$). In order to further appreciate quantitatively the agreement with linear theory, the 95% confidence limits have been plotted. The crest, except for low waves, is comprised inside the interval whereas the trough is, until $N = 5 \cdot 10^5$, below it. The confidence interval has been calculated as 1.96 the standard sampling error (S) (assumed as being Gaussian distributed): the standard variation divided by the size of the sample. The variance of the maximum crest has been approximated using the asymptotic expression (Cartwright, 1958):

$$S = \frac{\sqrt{Var(\eta_{\max \text{ crest}})}}{\sqrt{n}}, \text{ where } n = \frac{\text{num total waves}}{N} \quad (12.3)$$

$$Var(\eta_{\max \text{ crest}}) = E(\eta_{\max \text{ crest}}^2) - E^2(\eta_{\max \text{ crest}}) \approx \frac{1}{2 \ln(N)} \left(1.6449 - \frac{2.1515}{\ln(N)} \right) \quad (12.4)$$

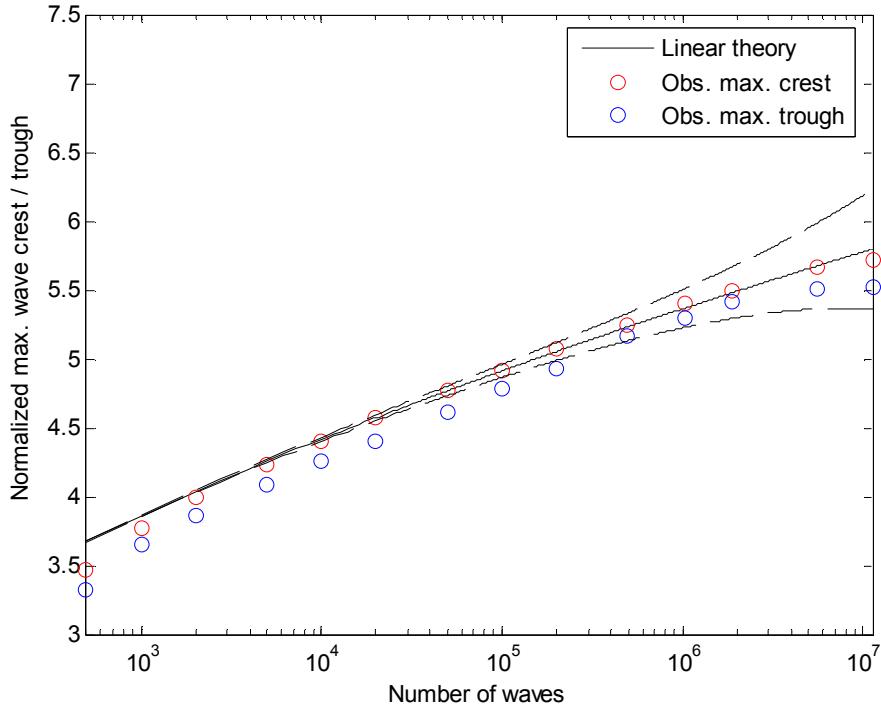


Figure 12.13 Comparison with linear theory with 95% confidence limits.

Although the general agreement with linear theory, the trough is slightly lower than the crest, leading to us to think that nonlinear effects are present and therefore the crest is more peaked and higher than the trough. However, these effects seem to be weak.

Including the effect of the spectral band width, according to Cartwright (1958), the obtained discrepancy is the same as in the wave height: 1 - 2.5% (see Figure 12.14). The estimated maximum wave amplitude with the inclusion of the spectral width lays between the maximum crest level and maximum trough level, suggesting that, in general, the linear theory slightly overpredicts the wave amplitude (both crests and troughs). However, the crest seems to be quite well predicted due to the possible presence of weak nonlinear effects. In the opposite sense, the wave trough appear to be affected by such nonlinearities.

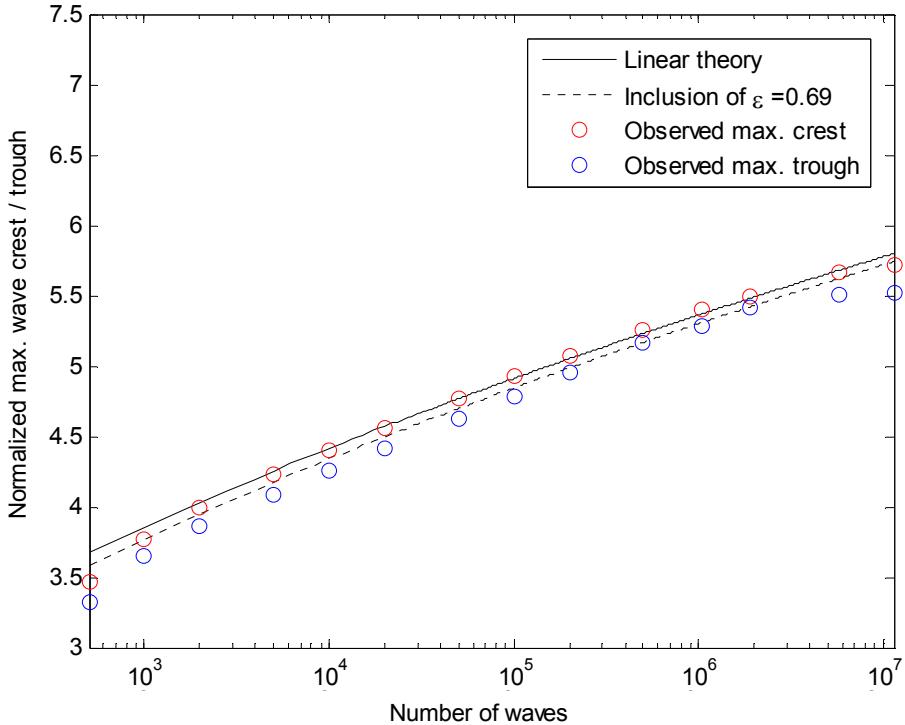


Figure 12.14 Inclusion of the spectral band width in the expected maximum wave crest / trough, comparison with the linear theory

12.4. Nonlinearities

12.4.1. Introduction

From the above showed results in which neither the wave height nor crest/trough are underpredicted by the linear theory, one may suspect that the surface elevation can be almost considered Gaussian distributed. In the present section, one justifies with the observed skewness and kurtosis parameter (of the surface elevation) that nonlinearities can be neglected in this case. Note that skewness is related to the asymmetry in the surface elevation and the kurtosis with the enhancement of wave heights. Finally, the BFI is considered which, theoretically, in the narrow-band case is related to the kurtosis.

12.4.2. Kurtosis and Skewness

In Chapter 6 the nonlinear effects were discussed and it was concluded that they can be explained, in part, with the kurtosis (k) and skewness (s) of the surface elevation. Therefore, although they are not spectral parameters, it is interesting to analyse them and see if they differ from the Gaussian ones. The Gaussian distribution (linear theory) has $k = 3$ and $s = 0$. Sometimes, the kurtosis is “normalized” by subtracting 3 although here the first definition is used.

Figure 12.15 illustrates the agreement of one “mean” record with the Gaussian distribution (in red). With “mean” record, one means a record with kurtosis and skewness similar to the mean values of all the records (see Table 12.2 and 12.3).

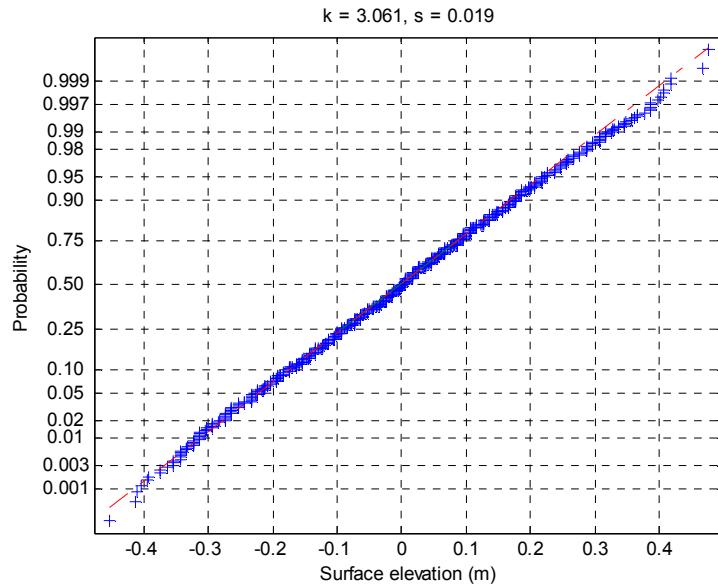


Figure 12.15 Observations of surface elevations, compared with the Gaussian cdf

Both kurtosis and skewness are computed for all records and then the mean and the standard deviation are calculated. In Figure 12.16, the distribution of each parameter is illustrated. For more detailed results see Table 12.2 and 12.3 in which a distinction for each buoy is made.

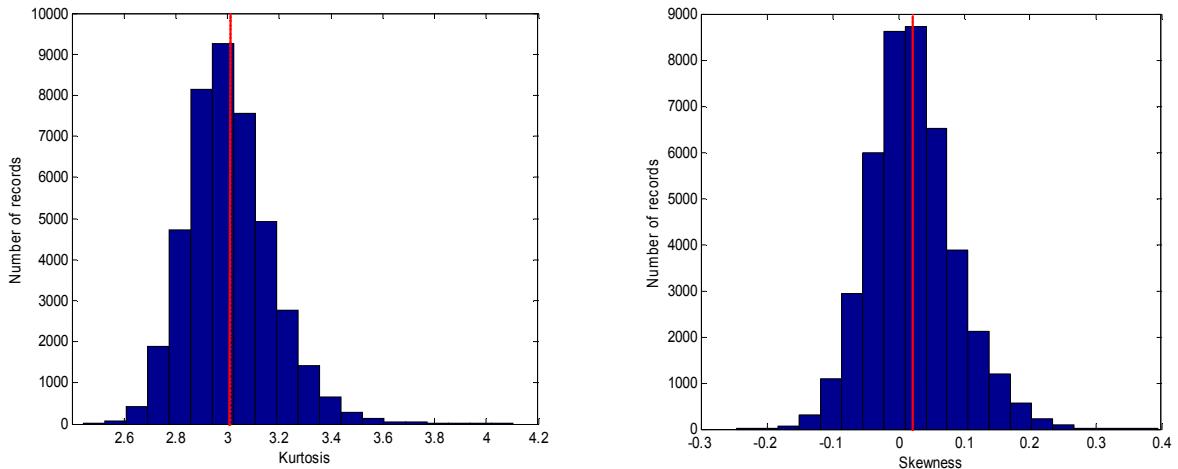


Figure 12.16 Histogram of the kurtosis (left) and skewness (right) parameter of the surface elevation. In red the mean value is marked.

Table 12.2 Distribution of the kurtosis for each buoy

| kurtosis | Roses | Blanes | Llobregat | Tortosa | Total |
|--------------------|--------------|---------------|------------------|----------------|---------------|
| Mean | 2.9853 | 2.9936 | 2.9877 | 3.0361 | 3.0100 |
| Standard deviation | 0.1510 | 0.1575 | 0.1583 | 0.1674 | 0.1625 |
| Maximum | 4.1033 | 4.0175 | 3.9500 | 4.0219 | 4.1033 |
| Minimum | 2.4455 | 2.4667 | 2.5175 | 2.5390 | 2.4455 |

Table 12.3 Distribution of the skewness for each buoy

| skewness | Roses | Blanes | Llobregat | Tortosa | Total |
|--------------------|---------|---------|-----------|---------|---------------|
| Mean | 0.0062 | 0.0069 | -0.0056 | 0.0448 | 0.0219 |
| Standard deviation | 0.0562 | 0.0531 | 0.0528 | 0.0701 | 0.0647 |
| Maximum | 0.3949 | 0.2748 | 0.1738 | 0.3611 | 0.3949 |
| Minimum | -0.2101 | -0.2406 | -0.2467 | -0.2235 | -0.2467 |

From the statistical viewpoint, the kurtosis has to do with the peakedness of the probability density function of the surface elevation: the higher the kurtosis, the more peaked the distribution. Physically, a higher kurtosis is related to a higher probability of encountering higher (large) wave heights. If the probability density function is more peaked, the probability of mid range wave heights is reduced whereas both surface elevation around zero mean level and at high extremes are enhanced. In Figure 12.17, the maximum and mean wave height per record, both normalized by the standard deviation, are plotted against the kurtosis. Having the Rayleigh-Edgeworth distribution in mind, the tendencies are those expected (see Chapter 11), although the correlation is not so high: the maximum wave height increases with the increase of kurtosis whereas for the mean height is the opposite. In the picture of the maximum height a certain dispersion is expected because the number of waves among all records is slightly different. However, this does not appear to be the main cause because, in the case of the mean height, which does not depend on the number of waves, the dispersion is also present. The red point in the second picture, represents the linear theory.

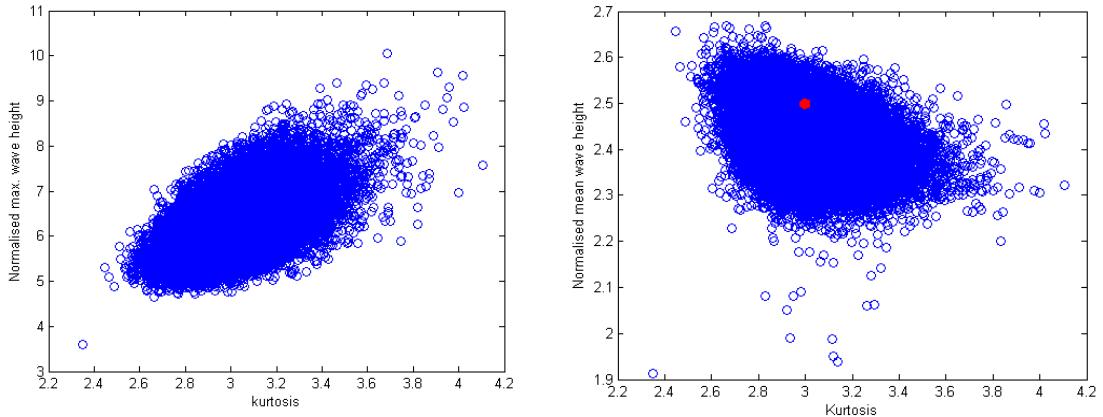


Figure 12.17 Normalized maximum and mean wave height vs. Kurtosis.

In order to clearly see the relation between the maximum wave height and the kurtosis parameter, Figure 12.18 shows the dependency of this variable on N and kurtosis simultaneously. In a analogous way as in the Section 12.2.2, the normalized maximum wave height has been calculated for different values of N . The difference is now that the records have been split in 12 groups (each group with approximately the same number of records), in ascending order of kurtosis. The mean values of kurtosis for each group are: 2.75, 2.84 (yellow), 2.88, 2.92 (magenta), 2.95, 2.98 (green), 3.01, 3.05 (cyan), 3.08, 3.13 (dark blue), 3.20, 3.35 (black). The stratification of colours clearly shows that a relation exists between kurtosis and encountering higher normalized maximum wave heights.

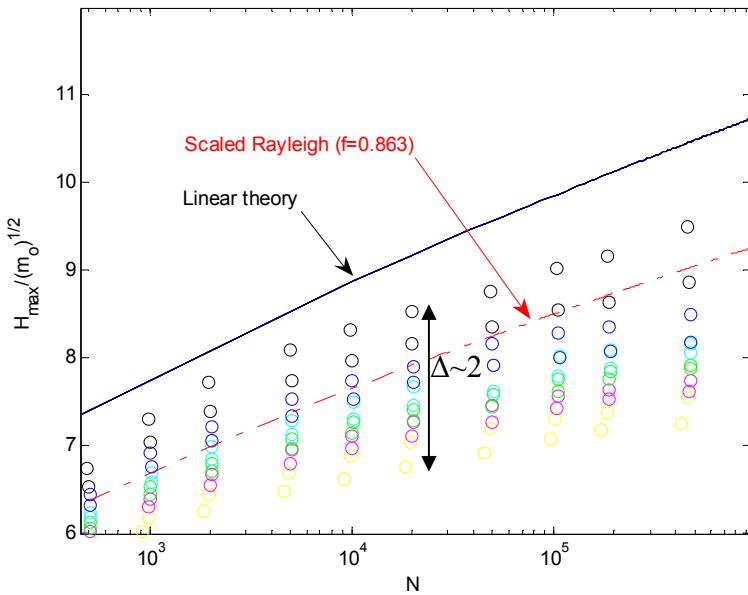


Figure 12.18 Mean normalized maximum wave height for different N . The colors are associated to mean values of kurtosis (see text above)

The mean kurtosis over all records is practically the same as in the case of linear theory but there is a clear dependency of high maximum wave heights for high (larger than 3) values of kurtosis. However, the Rayleigh-Edgeworth distribution, which is the theory which includes the kurtosis parameter (in the definition of BFI), does not seem to be the appropriate probability function to describe this stratification because it does not account for kurtosis lower than the 3 (Gaussian distribution). Point out that a sine wave has a kurtosis of 1.5. Note that, considering the value of $\sqrt{m_0} = 0.3 \text{ m}$ (approximately the mean value of the present data set), the variation of the maximum wave height (see Figure 12.18) can be roughly approximated by $\Delta \approx 2\sqrt{m_0} \approx 0.6 \text{ m}$. In contrast, with the skewness, there is no stratification (see Figure 12.19). The mean values of skewness for each group are: -0.0877, -0.0486 (yellow), -0.0291, -0.0139 (magenta), -0.0008, 0.0116 (green), 0.0241, 0.0374 (cyan), 0.0523, 0.0705 (dark blue), 0.0966, 0.156 (black). The deviation for large wave heights is, in part, expected due to the random variability.

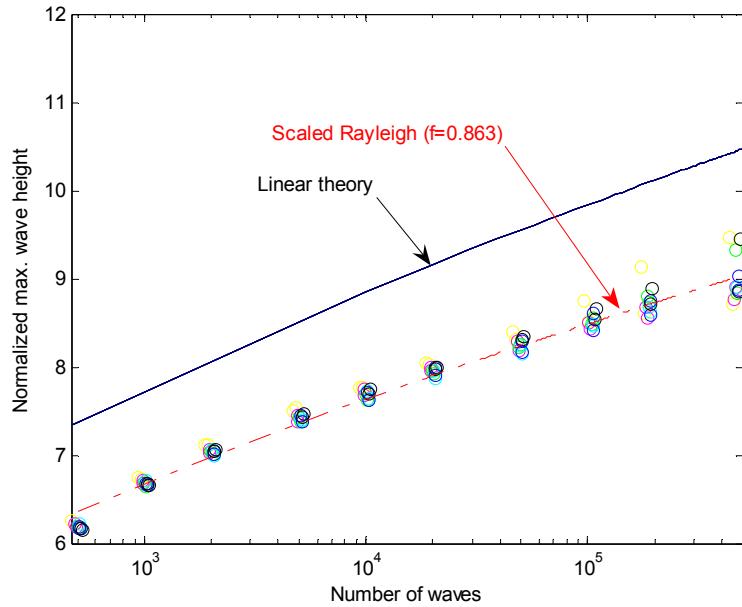


Figure 12.19 Mean normalized maximum wave height for different \mathcal{N} . The colors are associated to mean values of skewness

Figure 12.20 and 12.21 illustrate the time trace of kurtosis and skewness in which the 95 % confidence intervals are plotted in dashed lines. Such intervals have been found by simulating with the Montecarlo method 1000 samples of surface elevation using the Gaussian distribution. Note that the interval of the Tortosa buoy for kurtosis and skewness is wider because the number of data points in the wave record is smaller.

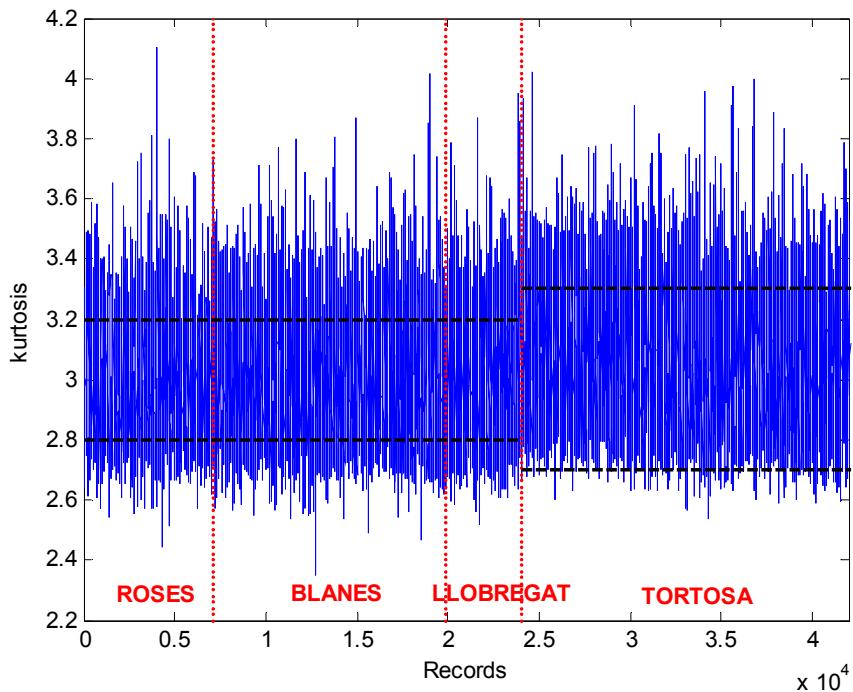


Figure 12.20 Time trace of kurtosis, concatenating all the records of each buoy

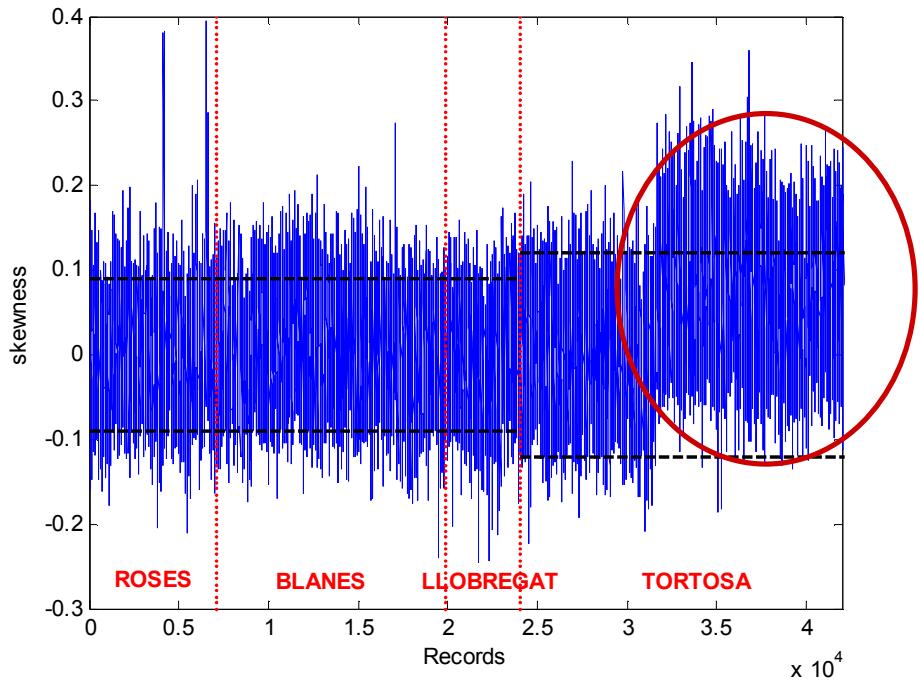


Figure 12.21 Time trace of skewness, concatenating all the records of each buoy

The skewness measures the asymmetry of the probability density function. For example, a positive skewness means that there is an elongated tail to the higher values. This agrees with the skewed profile of surface elevation, with higher and peaked crests and shallower and rounded troughs, meaning that the crests/troughs are larger/shorter (see Figure 6.2).

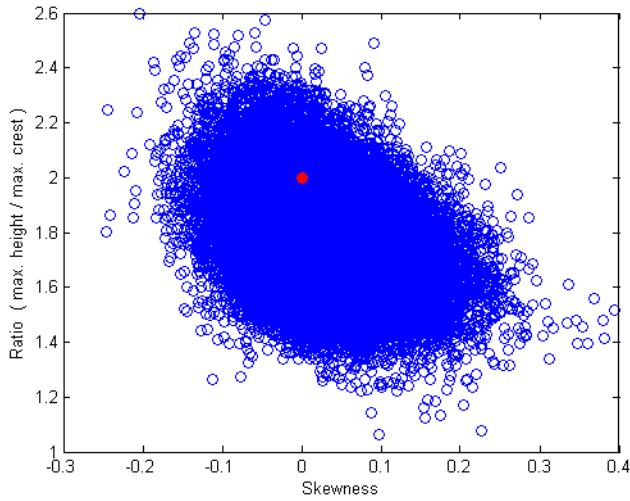


Figure 12.22 Comparison $H_{\max} / \eta_{\max \text{ crest}}$ vs. skewness

In contrast to the kurtosis, the mean value of skewness is relatively higher than the Gaussian one (zero) compared to the kurtosis. However, in this case it is interesting to pay attention to the time trace illustrated in Figure 12.21. It is noticeable that in the last few years, the buoy of Tortosa clearly registered a higher mean compared to the rest of the years and buoys. Although the reason is unknown, it seems related to a registration error because during the same

years, the rest of the buoys do not suffer from this higher mean skewness and, curiously, this “jump” is found at the beginning of 2001, coinciding with the second period of available data in Tortosa (in Chapter 2, it was explained that Tortosa has two periods of available raw data: 1991-1997 and 2001-2006, 2004 is missing). Therefore, at first sight, it seems to be very probable that in the replacement of the buoy, the installed device was not properly moored. Without considering the second part of data of Tortosa, the mean of the skewness drops down from 0.0219 to 0.0036, whereas the mean of the second period is 0.0758. However, the affectation to statistics is little, above all for the wave height. The observed maximum wave crest / trough without considering this period of larger skewness, is compared with the linear theory expectations in Figure 12.23. There is little variation, above all in the mid range of number of waves, in which the crest is slightly lower and the trough higher (the wave height is practically not affected).

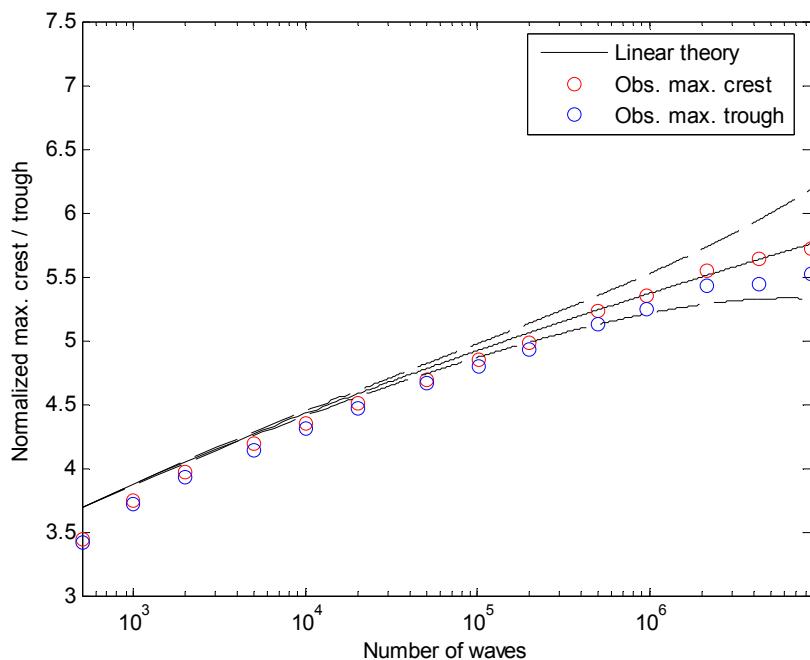


Figure 12.23 Observed mean maximum wave crest / trough compared to the linear theory (the period 2001-2006 of Tortosa buoy is not considered)

In conclusion, the mean kurtosis and skewness are practically the same as the Gaussian ones. However, it is seems that for higher kurtosis, the maximum wave height tends to be higher and for higher skewness, the same happens for the crest.

12.4.3. BFI

The BFI is a spectral parameter which according to Janssen (2003) is closely related to the kurtosis in the narrow band case. It is important to note that, in general, spectral parameters are preferable to statistical ones because in the first case, one only needs the spectrum which can be predicted. This parameter has been calculated for all records. The mean value is 0.32 (see Figure 12.24) but there is no apparent correlation between kurtosis and BFI (see Figure 12.25), and, therefore, between large wave heights and BFI. That reason is probably due to the narrow assumption made by Janssen (2003). In fact, considering the quarter of records with higher maximum wave height, the mean of the BFI remains almost the same. A similar conclusion was

yield by Olagnon & Magnusson (2004). Alber (1978) demonstrated that the BFI vanishes if the wave spectrum is sufficiently broad. That could be the main reason but also the fact of having mixed states. Rotés (2004) analysed more than 4.000 spectrums in the Mediterranean Sea and found that swell reduces considerably the value of the BFI. Nevertheless, although she calculated the BFI using an alternative method (using the peak frequency and half the spectral width at half the peak energy, see Chapter 6), she found little relation between highest waves and highest BFI, either. In addition, the presence of low frequency noise in the spectrum could increase such a reduction.

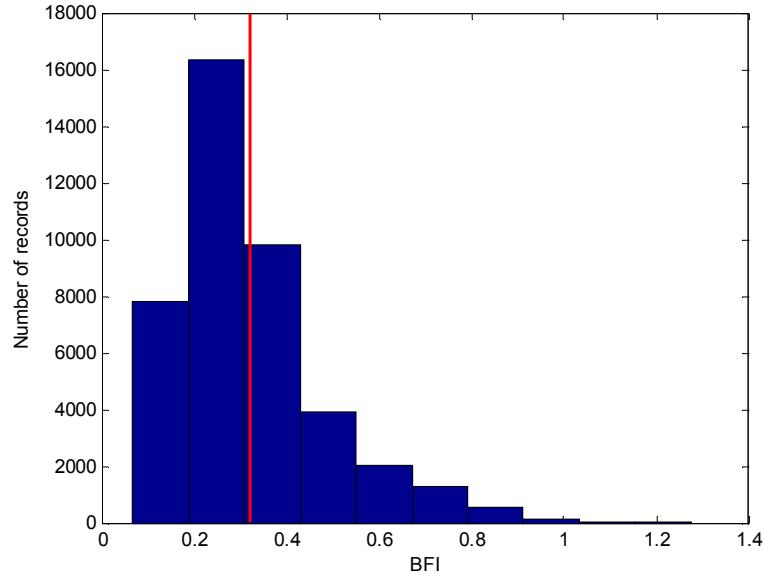


Figure 12.24 Histogram of the BFI values of all records

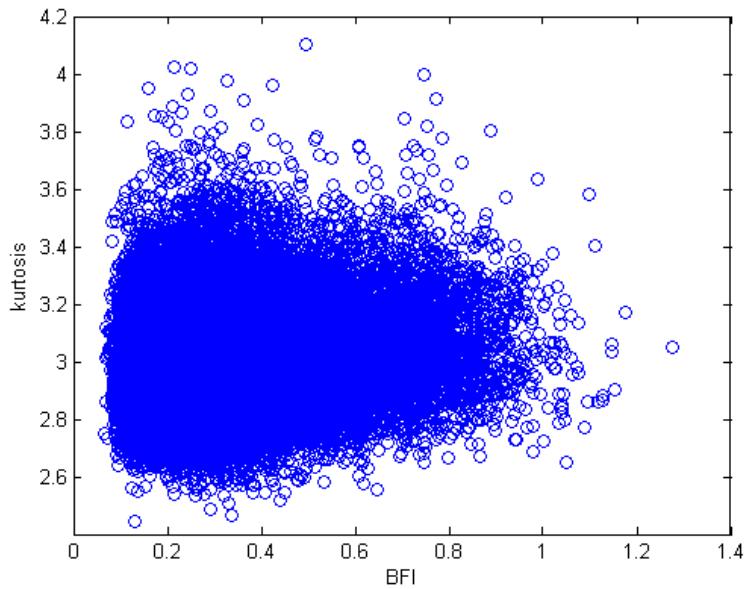


Figure 12.25 Observed kurtosis vs. calculated BFI

12.5. Conclusion

In front of these results, it seems that the overprediction of the linear theory for both the crest and trough is practically negligible and therefore the wave envelope (of the linear theory) itself practically does not overpredict the crests and troughs. On the contrary, the wave height differs considerably. Therefore, the assumption of the wave height being twice the wave amplitude appears to be the most important cause of such an overprediction although other factors are probably also involved. On the other hand, the nonlinearities appear to be weak and, therefore, the assumption of surface elevation being Gaussian distributed appears to be quite acceptable. However, a definitive conclusion cannot be yield only from the analysis of buoy measurements which may tend to avoid high crests.

13. ANALYSIS OF THE NORTH SEA DATA

13.1. Introduction

The aim of the analysis of this additional data from the North Sea (the WADIC project) is not the mere repetition of the analysis of Chapter 12 for another set of data. There are two main reasons:

- They were registered by two laser altimeters installed on an offshore platform. It is well-known that buoys might tend to avoid wave crests whereas the laser altimeters do not suffer from this problem. Therefore, the results concerning the wave crests and troughs will be compared in detail.
- The Mediterranean data consists of a large number of records. However, due to its location and climate, not many high waves are included. In this particular data set which corresponds to two storms in 1985, illustrates more extreme events.

13.2. Results

First of all, before entering into detail in the analysis of parameters such as significant and maximum wave height, illustrative pictures of the observed probability of exceedance of height, crest and trough are shown (see Figure 13.1), compared with the famous Rayleigh distribution. A logarithmic scale is used for the y axis. Qualitatively, at first view, it seems that the crest is reasonably well predicted by the Rayleigh theory whereas the trough and the height are overpredicted. Bear in mind the random character of sampling from the entire population. In order to have an idea of such variability, 69 (the sample number of records as in the North Sea data) samples of 300 wave crests (the mean number of waves per record) have been simulated, according to linear theory, with the Montecarlo method. With such simulation, see Figure 13.2, one intends to demonstrate that the variability found in the observations for the pdf is reasonable. The observed pdf has been calculated according to the simple California method (Vrijling & van Gelder, 2006): i / N .

The green colour means a kurtosis of the surface elevation higher than 3 whereas blue is lower. Later, nonlinearities are analysed more in-depth but here one can advance that, as in the Mediterranean data, there is a certain correlation between high height and high kurtosis. Nevertheless, this does not directly imply that nonlinearities are important.

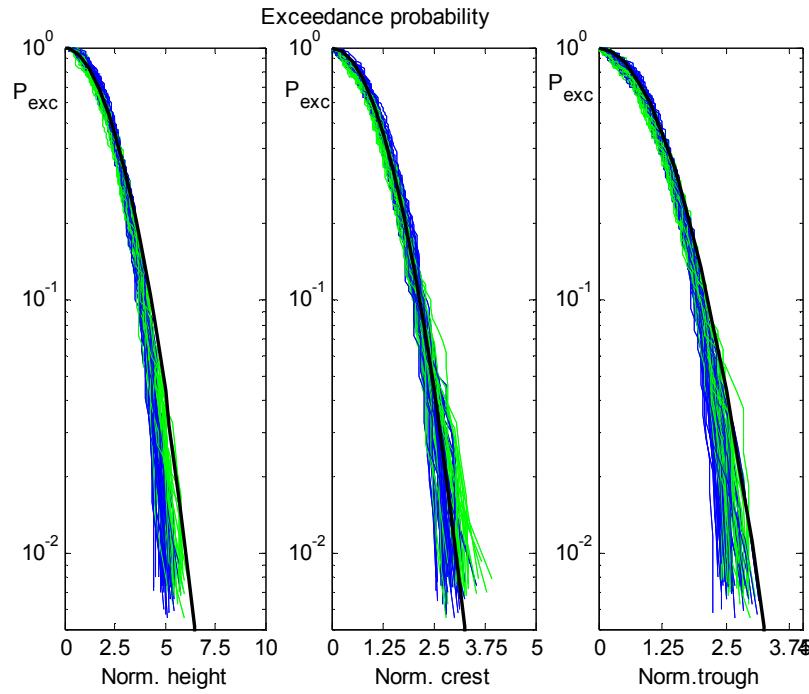


Figure 13.1 Observed exceedance probability compared to the Rayleigh distribution for wave height, crest and trough

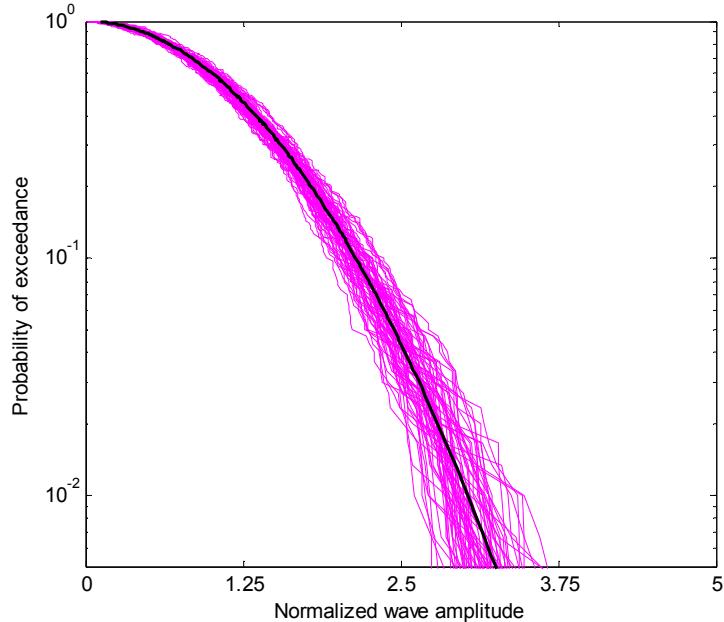


Figure 13.2 Empirical probability of exceedance of 69 simulated samples of 300 wave crests Rayleigh distributed (magenta), using the Montecarlo method, compared to the Rayleigh distribution (black)

The significant, mean and root-mean-square values of, respectively, wave height, crest and trough are illustrated in Figure 13.3, 13.4 and 13.5 in the same manner than in the previous Chapter. The estimated values are obtained from the linear theory and therefore the black line

represents the hypothetic perfect agreement with the linear theory. The linear regression to data is plotted in red.

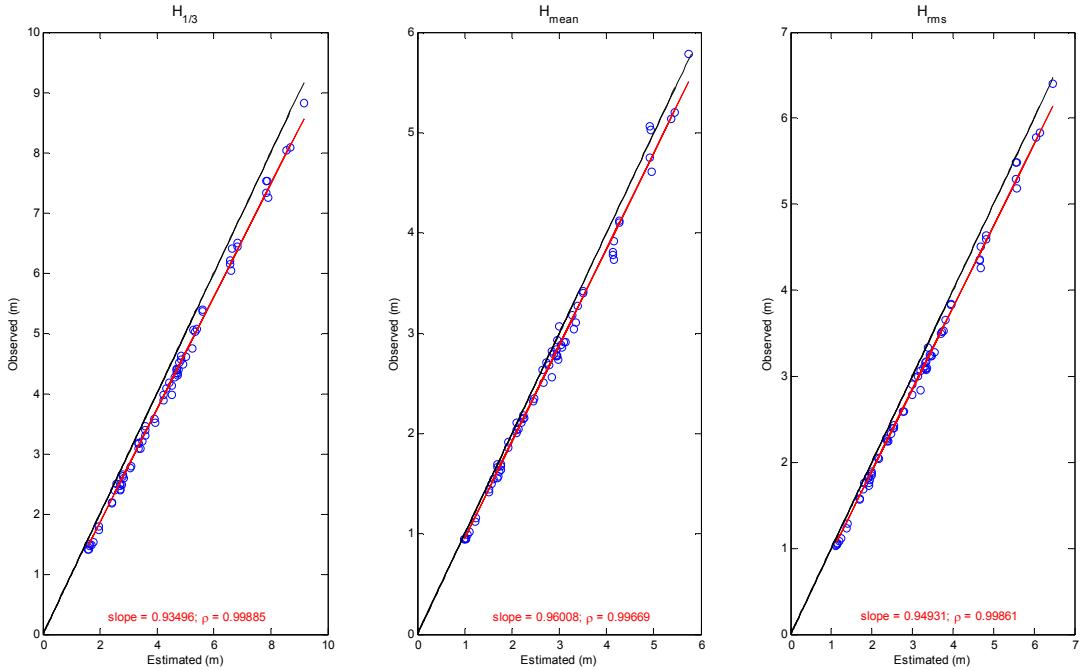


Figure 13.3 Wave height: comparison between observed and calculated parameters with linear theory

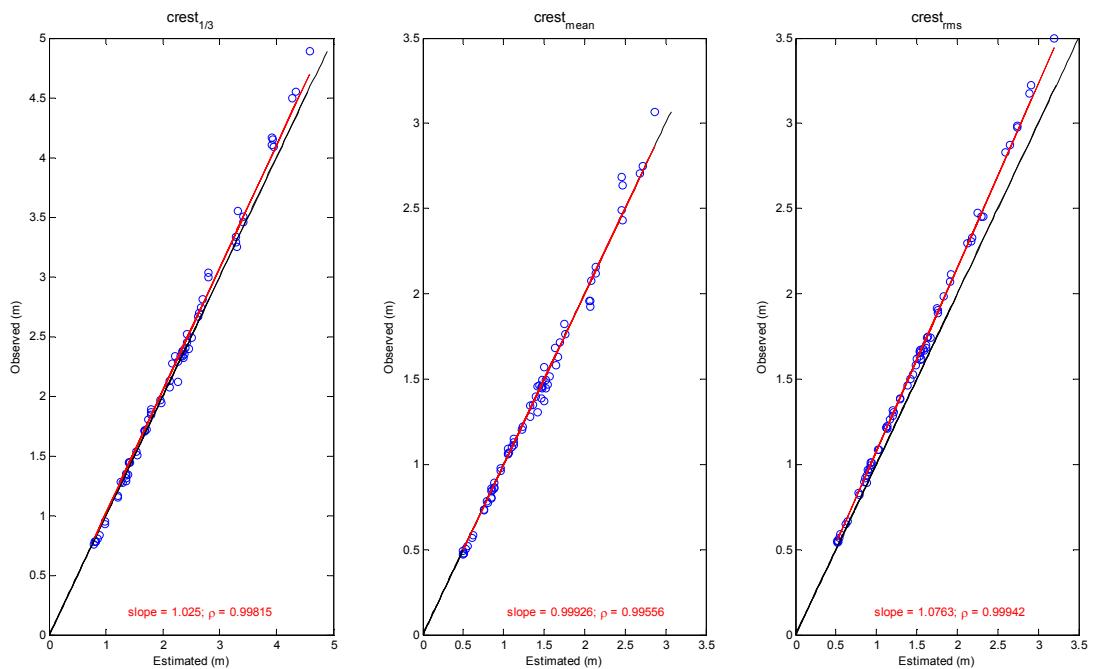


Figure 13.4 Wave crest: comparison between observed and calculated parameters with linear theory

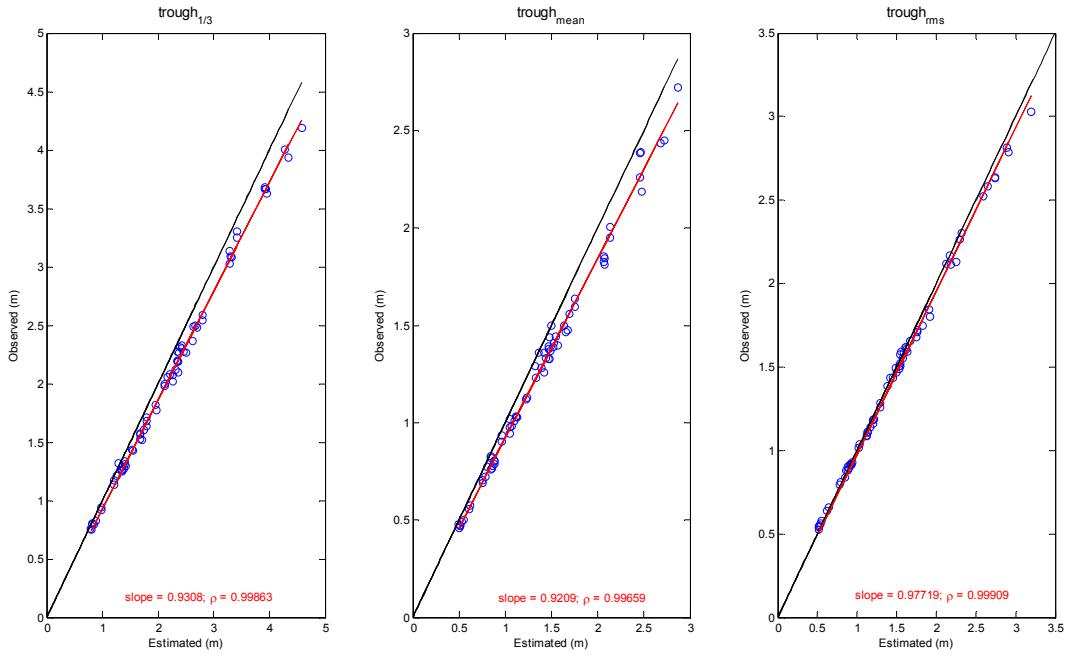


Figure 13.5 Wave trough: comparison between observed and calculated parameters with linear theory

As expected, higher wave heights are detected: the maximum significant wave height is about 9 m whereas in the Mediterranean data is about 5 m. These higher events do not show a different statistical behaviour. Note that the North Sea data rejected due to the condition of deep water had a maximum significant wave height of 11 m.

In order to easily compare the results between the two sources of data sets, Table 2.1 summarises the discrepancies (slopes) of each parameter in comparison to the Rayleigh distribution. Note that in the case of the maximum value, the number of waves is noticeably different for each data set (about 10^7 million for the Mediterranean data and 9500 for the North Sea data).

Table 13.1 Comparison between Mediterranean and North Sea.

| Parameter | Sea | Significant | Mean | Root-mean-square | Maximum |
|-----------|---------------|-------------|-------|------------------|---------|
| Height | Mediterranean | 0.931 | 0.960 | 0.949 | 0.863 |
| | North Sea | 0.935 | 0.958 | 0.945 | 0.832 |
| Crest | Mediterranean | 0.984 | 0.973 | 1.036 | 0.991 |
| | North Sea | 1.025 | 0.999 | 1.076 | 1.057 |
| Trough | Mediterranean | 0.977 | 0.956 | 1.024 | 0.962 |
| | North Sea | 0.931 | 0.921 | 0.977 | 0.829 |

The discrepancies of the wave height are almost the same between the two sets of data for the three values (significant, mean and root-mean-square height) and agree with some of the literature, explained in the previous Chapter. Although it might seem a little daring, one could conclude that these discrepancies can be taken as “universal” values.

On the contrary the discrepancies for the crest and trough are significantly different. The crest is, in general, underpredicted by the linear theory whereas the trough is clearly overpredicted,

suggesting the nonlinear profile of sharp high crests and rounded less deep troughs. The most probable reason of such difference between Mediterranean data and North Sea data is the device used. Therefore, in the case of the buoy measurements off the Catalan coast, the nonlinearities might be present but not properly registered by buoys. Broadly speaking, the reason is the Lagrangian character of the buoy measurement. The buoy tends to avoid high crests and also tends to go to deeper troughs around its location. In fact, Forristall (1999) concluded that surface following buoys effectively cancel out the second-order nonlinearity by making a Lagrangian measurement. That reasoning was also repeated in the paper of Prevosto & Forristall (2004) who affirmed that the buoy gives crest height statistics very close to those obtained from the linear theory. In addition, the WADIC project authors (Allender et al., 1989), in which many different types of wave measurements systems were analysed, confirmed that the registered individual wave profiles from the WAVERIDER were more symmetric than the fixed instruments such as the laser altimeter. Therefore, WAVERIDER seems to miss a large percentage of the very highest waves, either by being dragged under or by traversing around the highest waves. In principle, according to the results of the WADIC project, the laser altimeters are more reliable systems. In fact, in this project, together with other fixed instruments (subsurface current meter/pressure cell triplets), the records from the laser altimeters were used to make up the best estimate data set: being pattern for comparing with all registrations. However, a noticeable drawback is that the laser's times series were sometimes spiky (Allender et al., 1989).

However, bear in mind that, although the differences between buoys and laser altimeters have become clear, the climate conditions also have an important role. Physically, the manifestation of nonlinear characteristics in terms of higher and more pointed crests and shallower and more rounded troughs is related with sporadically breaking of waves and this is more pronounced in stormy conditions with high wind speeds. Actually, Honda & Mitsuyadsu (1975) found that the skewness of surface elevation depends on the wind speed: the higher the wind speed, the higher the skewness. Therefore, although it is evident that the buoy tends to counteract nonlinear effects, it is probable that such nonlinearities are, in reality, less pronounced in the Mediterranean waves than in the North Sea. In addition, note that the data of North Sea is from two major storms with wind speeds of 10 m/s or higher. On the contrary, the Mediterranean data includes more types of climate situations, with a mean wind velocity of approx. 2 m/s.

Therefore, in the Mediterranean data, the real surface profile might be more asymmetric than what records show. In fact, after comparing with the laser altimeters, the higher skewness observed in the last period of Tortosa buoy may not be an anomaly; perhaps, the weak nonlinearities are better measured. Note that wave height does not seem to suffer from these nonlinearities, with or without them the height becomes practically the same, this being in agreement with the theory explained in Chapter 6 (including the skewness) which concluded that the enhancement of crests is the same as the reduction of troughs. Note that Tayfun (1983) concluded that, after including the nonlinear effects (concerning the skewness), it is reasonable to suggest that net effect for the wave height of nonlinearity is negligible, especially in the narrow-band conditions. Massel (1996) also observed that nonlinearities do not appear to affect the distribution of zero-crossing wave heights.

In the analysis of the maxima the discrepancies (see Table 13.1), the above reasoning is more or less consistent: the wave height discrepancies are similar in the two data sets (even higher in the case of the North Sea data) and clearly different for the crest and trough, being respectively

approx. 6% higher and 18% lower than in the linear case. The maximum wave height is 13.55 m (in the Mediterranean data is 8.53 m).

Figure 13.6 illustrates the normalized maximum wave crest and trough compared to the linear theory. They are clearly more separated than in Figure 12.13 shown previously. Note that in this case the total amount of waves is considerably lower: 9500 waves approx and therefore the confidence limits are wider. The crest is more or less getting into the upper band of the 95% confidence interval whereas the trough is out of it.

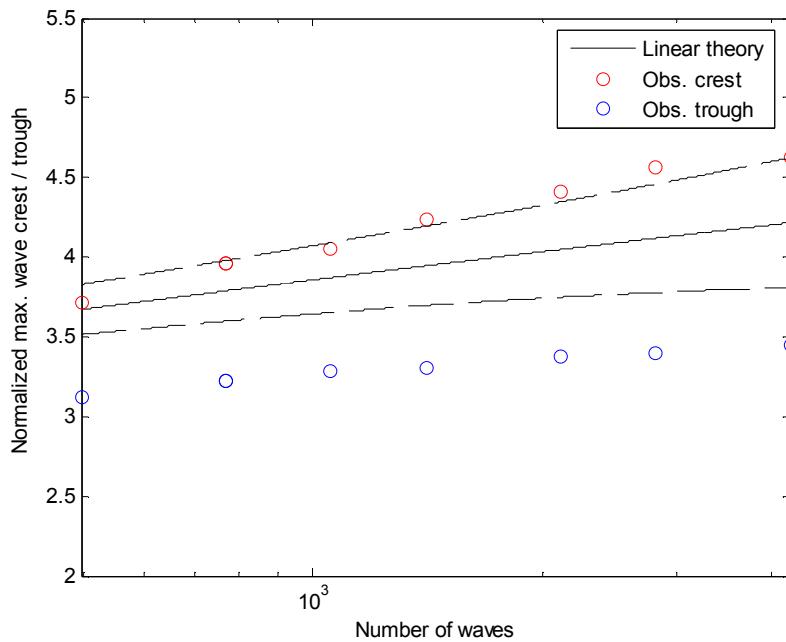


Figure 13.6 Comparison between observed and estimated mean maximum wave crest / trough with the linear theory.

Considering the theory of Tayfun (1994) the crests and troughs should be, respectively, higher and lower than in the linear case but in a symmetric manner (the wave height remaining Rayleigh distributed). However, it has been previously illustrated the wave height is underestimated by the Rayleigh theory due to the spectral bandwidth itself and the assumption of twice the wave amplitude. Therefore, the expected maximum wave amplitude is lower than the one from the linear theory. In Figure 13.7, the estimated wave amplitude is calculated with the inclusion of the mean spectral band width in the number of waves. Now, though there is a certain tendency of troughs being lower than expectations, this estimated amplitude is practically at the same level as the hypothetic mean between crests and troughs (in green asterisks).

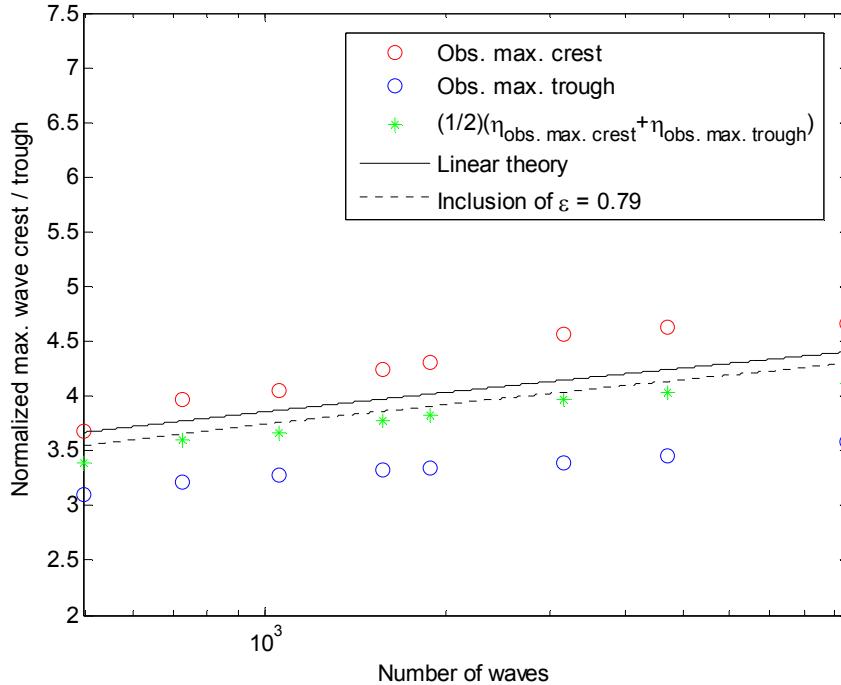


Figure 13.7 Comparison between observed and estimated maximum wave crest / trough with the linear theory, inclusion of spectral band with.

Table 13.2 summarises the ratio observations / predictions for the same theories (which assume a surface elevation Gaussian distributed) studied for the Mediterranean data. The results are quite similar to the ones of the previous data. The Modified Rayleigh II (Naess) overpredicts the observations except for the maxima. For lower wave heights the theory of Longuet-Higgins is better whereas for large wave heights the crest-to-trough becomes the best, although the agreement is not perfect. In this case the effect of the spectral bandwidth for the maxima has been approximate with a 3 % discrepancy.

Table 13.2 Relation observations/expectations and % of discrepancy, according to different theories

| Parameter | Rayleigh | | Modified Rayleigh I (Longuet-Higgins) | | Modified Rayleigh II (Naess) | | Crest-to-trough | |
|------------------------|----------|--------|--|--------|---------------------------------|-------|-----------------|--------|
| H_{mean} | 0.96 | - 4 % | 0.99 | - 1 % | 1.07 | + 7 % | 0.96 | - 4 % |
| H_{rms} | 0.95 | - 5 % | 0.99 | - 1 % | 1.06 | + 6 % | 0.98 | - 2 % |
| H_s | 0.94 | - 6 % | 0.98 | - 2 % | 1.05 | + 5 % | 0.98 | - 2 % |
| H_{max} | 0.83 | - 17 % | 0.86 | - 14 % | 0.93 | - 7 % | 0.89 | - 11 % |
| $H_{max}(\varepsilon)$ | - 14 % | | - 11 % | | - 4 % | | - 8 % | |

After analysing the main parameters for the description of ocean waves, it appears that nonlinear effects are present but only mainly related to the skewness parameter, leading to higher crests and lower troughs, without affecting the wave height as the sum of crests and troughs. Table 13.3 summarises the mean, standard deviation, maximum and minimum of kurtosis and skewness directly obtained from the 69 records.

Table 13.3 Kurtosis and skewness parameter

| Parameter | Kurtosis | Skewness |
|--------------------|---------------|---------------|
| Mean | 2.9942 | 0.1158 |
| Standard deviation | 0.2001 | 0.0875 |
| Maximum | 3.5910 | 0.3499 |
| Minimum | 2.6149 | -0.1001 |

In mean terms, the kurtosis is practically the same as in the Gaussian distribution and similar to the case of the Mediterranean data. This agrees with the fact that the wave height is hardly affected by nonlinear effects. On the contrary, the skewness is noticeably higher than in the Gaussian distribution and also higher than in the Mediterranean observations (approx. 0.02). Hence, this result also agrees with the asymmetric profile derived from the analysis in the previous section. But now, it is necessary to try to quantify such agreement. First of all, as in the Mediterranean data, the 95% confidence intervals of the kurtosis and skewness are calculated (assuming the surface elevation Normal distributed) and plotted in the time traces of such parameters (Figure 13.8). In both kurtosis and skewness, more than 95% of the records are outside the confidence limits. However, for the skewness there is a clearer tendency towards higher values.

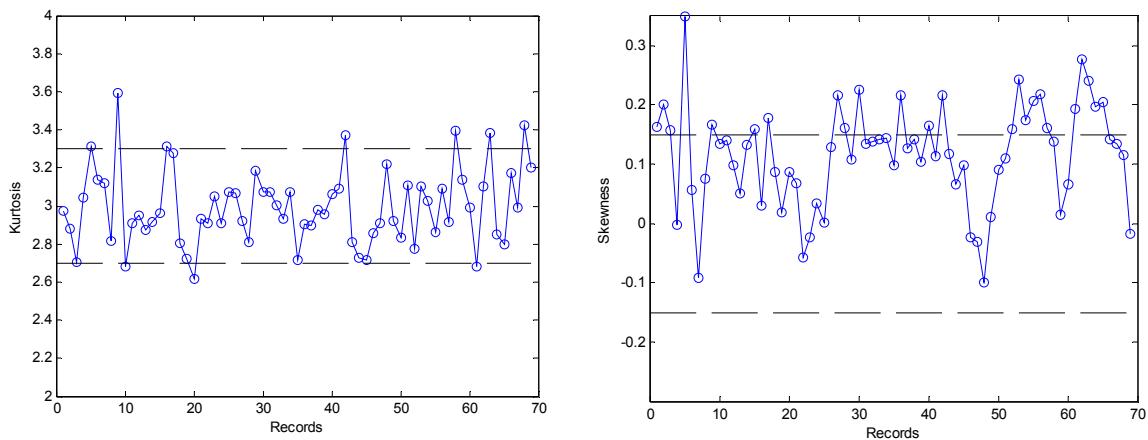


Figure 13.8 Kurtosis and skewness of 69 records with the 95% confidence interval.

In Figure 13.9, it is qualitatively shown that a certain correlation between skewness and normalized maximum wave crest exists. However, in Figure 13.10, confronting the two variables, the correlation does not seem to be so high.

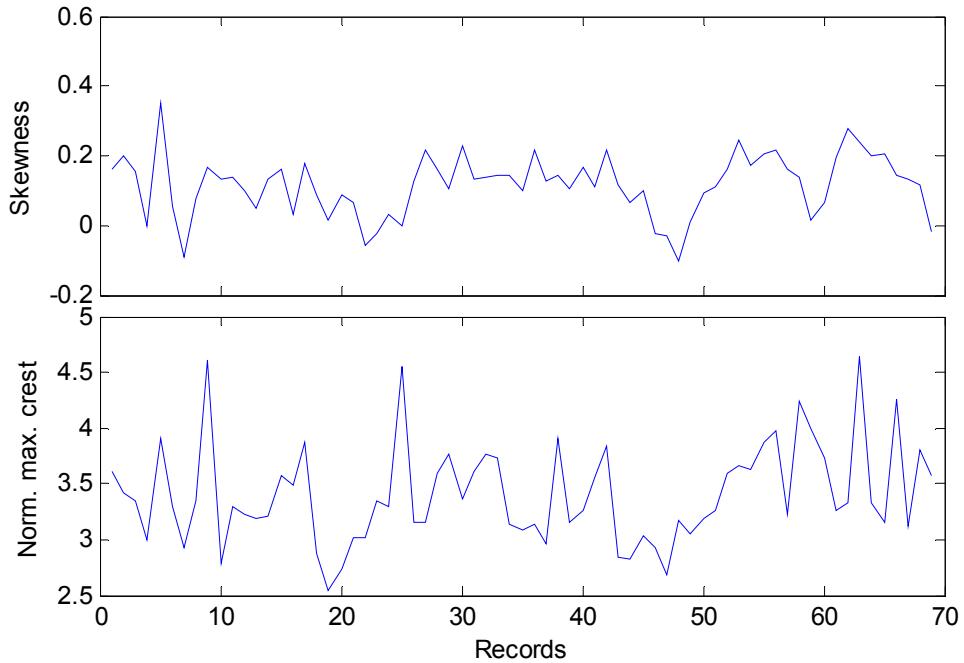


Figure 13.9 Skewness and normalized maximum wave crest for all records

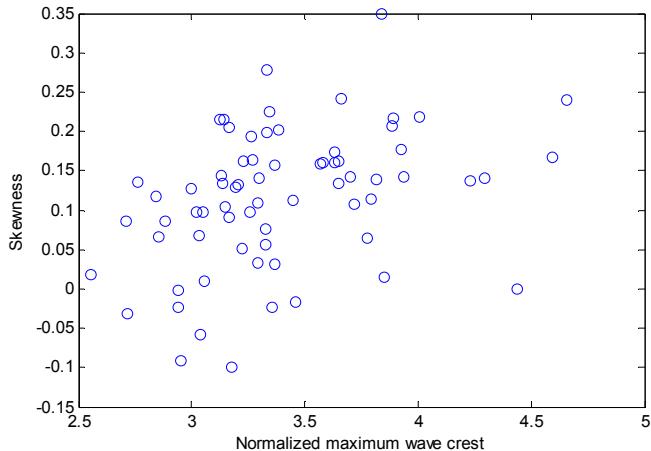


Figure 13.10 Skewness vs. normalized maximum wave crest for all records

In order to try to quantify such affectation, the asymmetric Rayleigh distribution (Tayfun, 1994) is used to estimate the significant and maximum wave crest / trough. Figure 13.11 illustrates the expected values of maximum crest and trough which has been calculated on the basis of the asymmetric distribution of Tayfun (1994) but considering the effect of ε in the number of waves n . The mean value of ε has been considered for such a calculation. Note that, owing to the higher mean value of ε , the reduction of the expected maxima is slightly higher than in the Mediterranean data: about 3% (instead of 1-2.5%). About the skewness, some considerations have to be done. Tayfun (2006) affirmed that the skewness derived from a wave records tends to be an unstable statistic because of its sensitivity to local trends and/or presence of an exceptionally large wave in the record. In addition, he suggested that the skewness would have to be estimated from the surface elevation representing relatively high values because the derived model (Asymmetric Rayleigh

distribution seems to be appropriate to the crest heights of large waves. Therefore, for the calculation of the maxima, the maximum skewness for each set of records (for each n) has been considered.

In general, the expected values underpredict in the observed crests whereas the observed troughs are overpredicted (see Figure 13.11). The difference is greater in the case of the troughs which do not lay within the 95% confidence intervals.

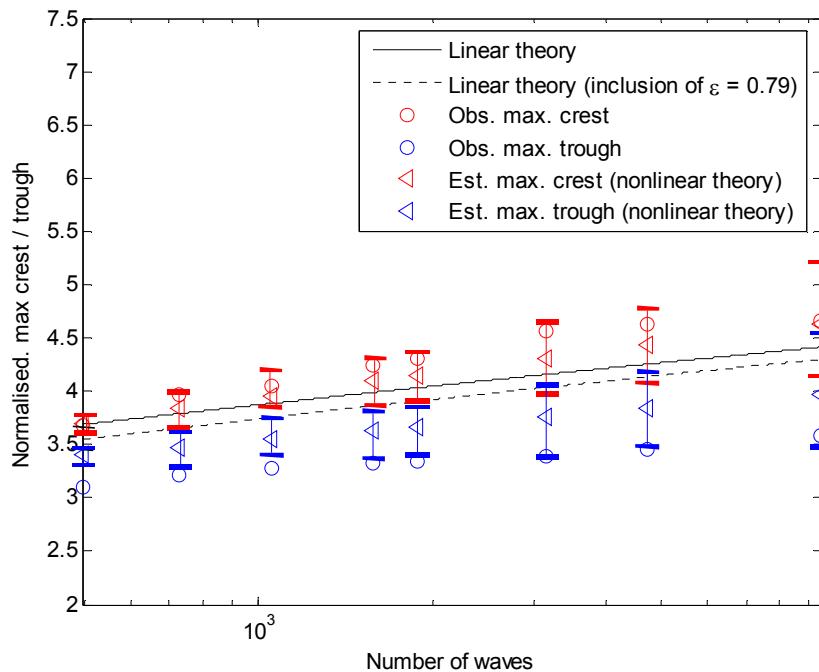


Figure 13.11 Comparison between observed and estimated mean maximum wave crests / troughs, according to Tayfun (1994) and including the effect of ε with the 95% confidence bands

The same theory of Tayfun (1994) has been used to estimate the significant wave crest / trough, another important parameter of wave statistics. In this case, the mean of the skewness of the one third of records with the highest maximum wave height: 0.1407. Remark that this estimation of the skewness is quite arbitrary. It is a way to use a representative skewness of high surface elevations as Tayfun (2006) suggested.

Table 13.4 Comparison between Rayleigh distribution and asymmetric Rayleigh (skewness = 0.1407) estimations for the significant parameter

| Parameter | Rayleigh | Asymmetric R. | |
|-----------|----------|---------------|-------|
| Crest | 1.025 | + 2.5 % | 0.977 |
| Trough | 0.931 | - 6.9 % | 0.979 |

Contrary to the case of the maximum parameter, the asymmetric distribution overpredicts the expected values of both significant wave crest and trough. Such discrepancy is about 2 % for both parameters, suggesting that it is due to the overprediction of the wave envelope, which is Rayleigh distributed. However, it is unclear because in the case of the maxima the expectations are smaller than the observations. Moreover, the result is quite sensitive to the skewness parameter, and it is not clear how it has to be estimated. As an example of the sensitivity of the

results to the skewness, the same calculation of the estimated significant wave crest / trough is made but now using the mean skewness over all the records (0.1158). In Table 13.5 the results are given.

Table 13.5 Asymmetric Rayleigh estimation for the significant parameter (skewness = 0.1158)

| Parameter | Asymmetric R. | |
|-----------|---------------|---------|
| Crest | 0.986 | - 1.4 % |
| Trough | 0.971 | - 2.9% |

The distribution of the BFI parameter is illustrated in Figure 13.12. The mean BFI is slightly higher than in the case of the Mediterranean data: 0.37 (vs. 0.32). However, it is not as high as expected considering that it comes from stormy conditions.

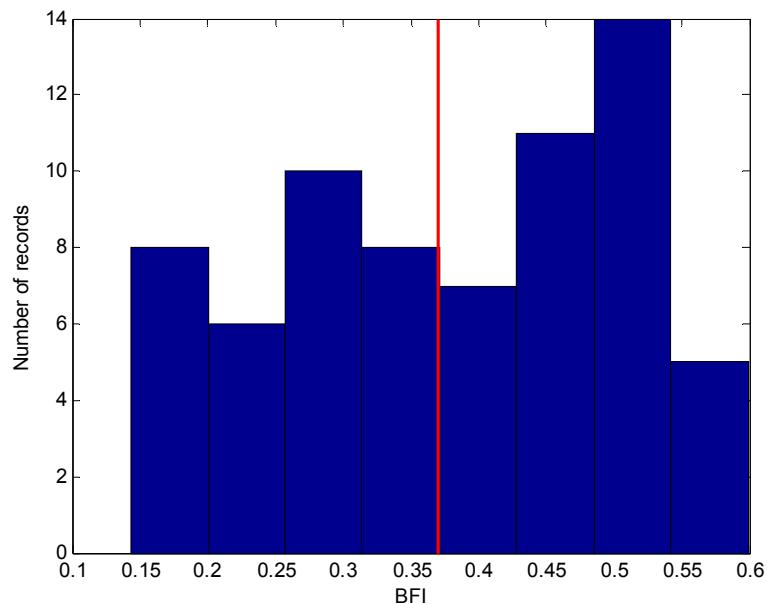


Figure 13.12 Histogram of the BFI parameter

14. CONCLUSIONS

After the analysis of the present study the following can be concluded:

- The correct maintenance of the instruments is essential in order to have reliable measurements of surface elevations. Although, it has recently been improved, the apparent low frequency of checking the proper operability of the buoys has lead to a database of raw time series being rather contaminated. In the present study, the quality control has become very complex and time consuming as it has had to be designed specifically for this data, basically due to the presence of many rough anomalies. In this task, the possible anomalies have been grouped in different types:
 - Rough errors, such as spikes or gaps
 - Buoys limitations: addition of trend, aliasing phenomenon, spectral energy in low frequencies due to knocks against the buoy...

Owing to the above mentioned reasons, nearly 70 % this data set has been refused.

- The type of instrument used for the registration of the surface elevation is very important. The comparison between buoy and laser altimeter data has clearly suggested that, as is predictable, the buoy tends to counteract nonlinear effects (due to its Lagrangian measurement), obtaining crests and troughs nearly well predicted by the Rayleigh distribution. Moreover, many others have come to the same conclusion, especially for the case of the crest. However, one does not pretend to attribute all the discrepancies found between Mediterranean and North Sea only to the device's fault. Note that the North Sea data is from stormy conditions with higher wind speeds, enlarging the nonlinear effects related to higher and more pointed crests and less deep and more rounded troughs.

Therefore, if one intends to analyse in detail the presence of nonlinear effects in the ocean waves is very recommendable to use a fixed instrument as, for example, laser altimeters. The disadvantage is that they require an offshore platform to be mounted on. On the other hand, to properly measure the wave heights, the buoys are acceptable.

- The wave height is clearly overpredicted by the Rayleigh distribution. The overprediction factors obtained in the two sets of data are nearly the same, agreeing with other observations analysed by others. In particular, the significant wave height is overpredicted by 7 %.

Solving this overprediction is a difficult job because many factors interfere:

- The calculated wave envelope in the linear theory overpredicts, in general, crests and troughs, and therefore also the wave heights.
- The assumption of $H = 2\eta_{crest}$ is also producing discrepancy. This assumption implies that the crest and trough have always the same value,

being the wave envelope symmetric, which is not true for real random waves.

In general the overprediction is worse for wider spectrum. The combination of the above mentioned effects causes overprediction of the wave height to not be constant for the different parameters (H_{mean} , H_s ...). Furthermore, there is a certain tendency in which the discrepancy is higher for larger wave heights and therefore the scaled Rayleigh distributions (e.g. Modified I and II) are not quite convincing (discrepancies, in general, higher than 3%). The overprediction due to the calculation of the wave envelope has been estimated to be 1-3% for the maxima. However, the main important reason of the overprediction seems to be the assumption of $H = 2\eta_{crest}$. The crest-to-trough theory tries to solve the problem by accounting for a time lag between crest and trough but it still considers symmetric envelopes. Although the complex expressions derived, the results, although quite good, are not convincing (between 2 and 7% discrepancy). More research is needed in order to quantify and predict such overprediction.

In conclusion, the Rayleigh distribution is “good” within 10% accuracy (except for the maxima which is worse) which is not acceptable for civil engineering purposes. In order to have more accurate results one can use the Modified Rayleigh I or the crest-to-trough theory, the last one often better for larger wave heights (in fact, remember that the simplified expressions are derived for wave heights higher than the mean wave height).

Nonlinear effects do not generally seem to affect the wave height but a clear correlation has been found between larger heights and kurtosis. This parameter has been found to be little related with the BFI, calculated as explained in the present study. More research is needed to try to estimate better such a spectral parameter. In fact Janssen, Mori and Onorato, in a recent conference (2th-6th June 2008) improved their own old results for the calculation of BFI, by considering the influence of the directional spectral width.

- The crest and trough are less affected by the overprediction problem; at most they are affected by 1-3%. The nonlinear effects, related to the skewness, seem to be present. The Asymmetric Rayleigh distribution predicts reasonably well such nonlinearities but the dependence on the skewness, a statistical parameter obtained from the time series instead from the spectrum, is not desirable (because the spectrum, contrary to the time series, can be predicted). Moreover, the results are quite sensitive to skewness, which generally speaking is not a stable statistic. The effect of nonlinearities is strong in the maxima, especially in the trough (more than 10%) but in the significant parameter is lower than 5%.

APPENDIX A: Quality control

In the following tables, the detailed information about rejected records is given. The order in the calculation is the same as in the shown in the first table in which the meaning of each type of error is clarified

| Error | Meaning |
|--------|--|
| e1 | Record length different from $D/\Delta t$ |
| e 2 | Vertical acceleration $> g/2$ |
| e 3 | 3 consecutives second derivatives = 0 |
| e_50 | $H_s < 0.5m$ |
| e4 | $\eta_{max\ crest} > 2.83 \cdot \eta_{crest,s}$ in one isolated data point |
| e_al | $f_{Nyq}/f_m < 2.2$ |
| e_bump | $E(0) > 0.004 m^2 / Hz$ |
| e_sh | $h < L_0/2$ |
| visual | Rejected after visual check |

| ROSES | | | | | | | | | | | |
|--------------|---------|----|--------|-------|-------|----|------|--------|------|--------|--------------|
| Year | Initial | e1 | e2 | e3 | e_50 | e4 | e_al | e_bump | e_sh | visual | Final |
| 2001 | 452 | 0 | 48 | 87 | 124 | 0 | 0 | 27 | 0 | 0 | 166 |
| 2002 | 8.725 | 0 | 2.258 | 2.820 | 1.441 | 1 | 0 | 117 | 4 | 0 | 2.084 |
| 2003 | 7.678 | 0 | 2.444 | 2.289 | 955 | 0 | 0 | 117 | 5 | 0 | 1.868 |
| 2004 | 5.847 | 0 | 565 | 2.164 | 1.299 | 3 | 0 | 67 | 1 | 0 | 1.748 |
| 2005 | 5.592 | 3 | 3.967 | 715 | 396 | 77 | 0 | 76 | 0 | 0 | 358 |
| 2006 | 7.962 | 12 | 4.585 | 1.592 | 662 | 0 | 0 | 190 | 4 | 0 | 917 |
| TOTAL | 36.256 | 15 | 13.867 | 9.667 | 4.877 | 81 | 0 | 594 | 14 | 0 | 7.141 |

| BLANES | | | | | | | | | | | |
|--------------|---------|----|-------|--------|-------|----|------|--------|------|--------|---------------|
| Year | Initial | e1 | e2 | e3 | e_50 | e4 | e_al | e_bump | e_sh | visual | Final |
| 2002 | 5.647 | 1 | 197 | 2.181 | 941 | 0 | 0 | 54 | 0 | 0 | 2.273 |
| 2003 | 8.227 | 1 | 2.732 | 1.404 | 1.351 | 7 | 0 | 70 | 0 | 0 | 2.662 |
| 2004 | 6.373 | 1 | 631 | 1.855 | 1.354 | 1 | 0 | 109 | 9 | 4 | 2.409 |
| 2005 | 8.119 | 0 | 497 | 2.678 | 1.763 | 0 | 0 | 139 | 0 | 2 | 3.040 |
| 2006 | 8.457 | 4 | 2.240 | 3.018 | 863 | 0 | 0 | 15 | 0 | 0 | 2.317 |
| TOTAL | 36.823 | 7 | 6.297 | 11.136 | 6.272 | 8 | 0 | 387 | 9 | 6 | 12.701 |

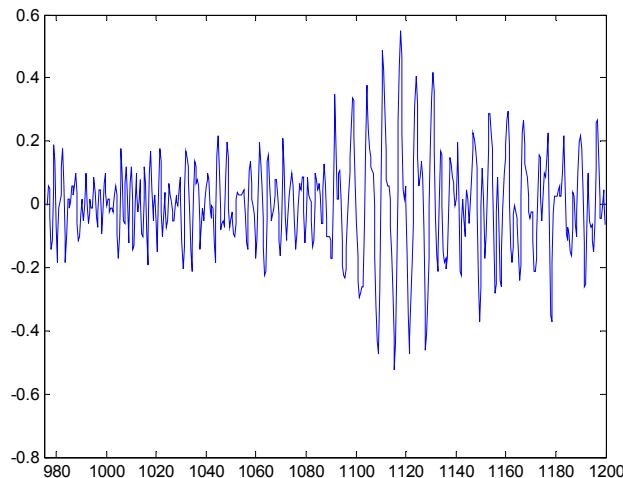
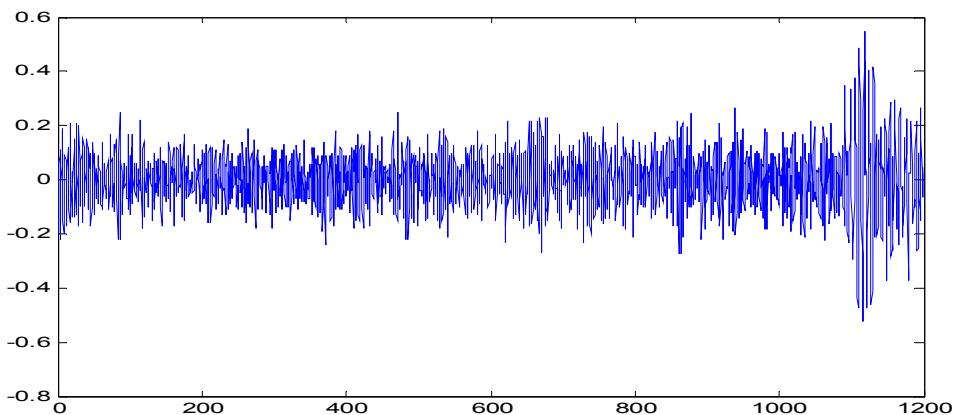
| LLOBREGAT | | | | | | | | | | | |
|-----------|---------|----|-------|-------|-------|----|------|--------|------|--------|--------------|
| Year | Initial | e1 | e2 | e3 | e_50 | e4 | e_al | e_bump | e_sh | visual | Final |
| 2001 | 1.939 | 0 | 711 | 408 | 296 | 0 | 0 | 11 | 13 | 0 | 500 |
| 2002 | 7.051 | 0 | 2.256 | 2.092 | 769 | 0 | 0 | 20 | 0 | 0 | 1.914 |
| 2003 | 8.332 | 2 | 3.512 | 2.212 | 765 | 0 | 0 | 32 | 13 | 0 | 1.796 |
| 2004 | 274 | 0 | 239 | 25 | 5 | 0 | 0 | 0 | 0 | 0 | 5 |
| | 17.322 | 2 | 6.479 | 4.712 | 1.830 | 0 | 0 | 63 | 26 | 0 | 4.215 |

| TORTOSA | | | | | | | | | | | |
|---------|---------|----|--------|-----|--------|----|-------|--------|------|--------|---------------|
| Year | Initial | e1 | e2 | e3 | e_50 | e4 | e_al | e_bump | e_sh | visual | Final |
| 1991 | 1.003 | 0 | 71 | 5 | 471 | 0 | 76 | 0 | 0 | 0 | 380 |
| 1992 | 1.935 | 0 | 75 | 16 | 593 | 0 | 139 | 0 | 0 | 0 | 1.112 |
| 1993 | 2.990 | 4 | 62 | 14 | 1.016 | 1 | 218 | 0 | 0 | 0 | 1.675 |
| 1994 | 3.219 | 3 | 69 | 22 | 997 | 1 | 185 | 2 | 0 | 0 | 1.940 |
| 1995 | 1.132 | 1 | 17 | 9 | 290 | 0 | 31 | 0 | 0 | 0 | 784 |
| 1996 | 2.098 | 2 | 63 | 24 | 704 | 0 | 40 | 0 | 0 | 0 | 1.265 |
| 1997 | 829 | 1 | 18 | 9 | 331 | 0 | 34 | 0 | 0 | 0 | 436 |
| 1998 | - | - | - | - | - | - | - | - | - | - | - |
| 1999 | - | - | - | - | - | - | - | - | - | - | - |
| 2000 | - | - | - | - | - | - | - | - | - | - | - |
| 2001 | 6.706 | 7 | 2.396 | 20 | 1.690 | 25 | 480 | 7 | 0 | 1 | 2.080 |
| 2002 | 3.101 | 0 | 897 | 15 | 933 | 7 | 250 | 1 | 0 | 0 | 998 |
| 2003 | - | - | - | - | - | - | - | - | - | - | - |
| 2004 | 6.924 | 1 | 2.696 | 108 | 1.431 | 11 | 537 | 18 | 0 | 8 | 2.114 |
| 2005 | 7.009 | 2 | 2.417 | 59 | 1.964 | 15 | 418 | 6 | 0 | 1 | 2.127 |
| 2006 | 7.762 | 8 | 1.653 | 6 | 2.221 | 13 | 452 | 0 | 0 | 0 | 3.409 |
| | 44.708 | 29 | 10.434 | 307 | 12.641 | 73 | 2.860 | 34 | 0 | 10 | 18.320 |

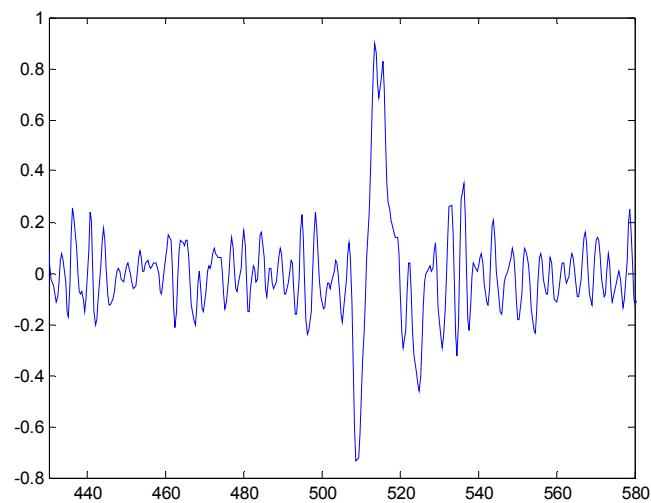
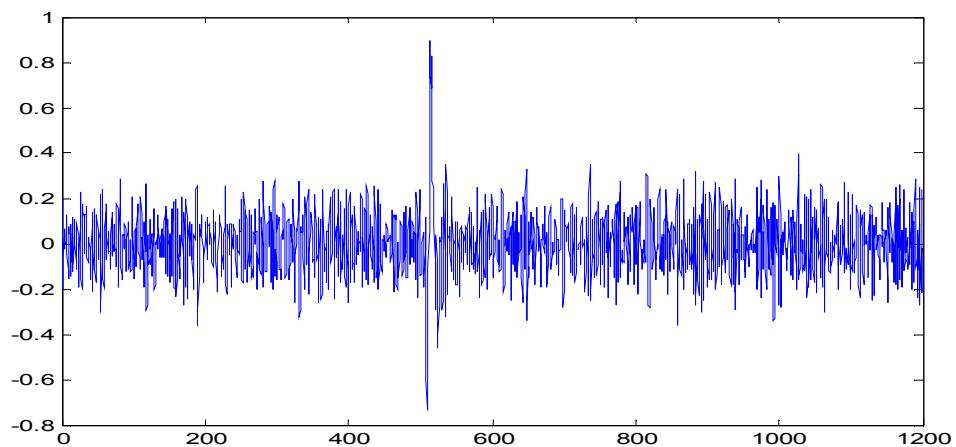
APPENDIX B: Freak waves

Freak waves have been defined as those with $H \geq 2.83H_s$. Although the particular analysis of freak waves occurrence is not covered in the present study, a complementary analysis is carried out by looking for all the freak waves present in the analysed data, attending to the above mentioned criterion. Three records have been detected in the Mediterranean data whereas none in those of North Sea. They are illustrated below (the first pictures are of the entire record whereas the second ones are an enlargement). The x-axis represents time (in seconds) and the y-axis surface elevation (in meters).

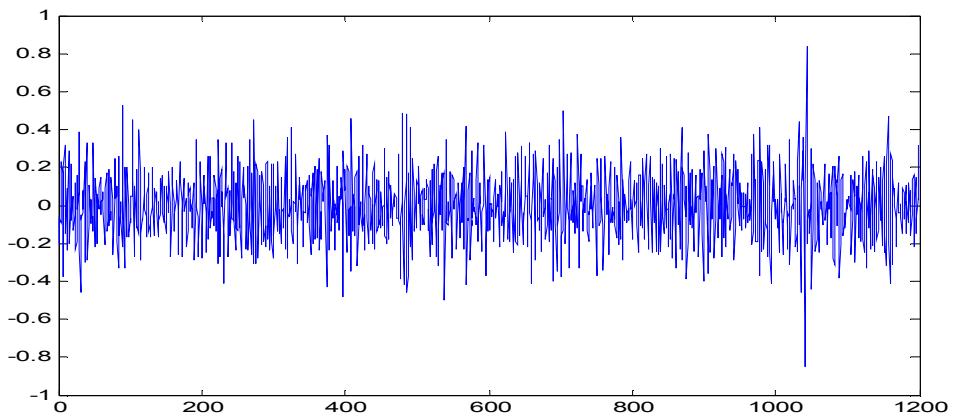
Llobregat: 01/06/02 03:20h

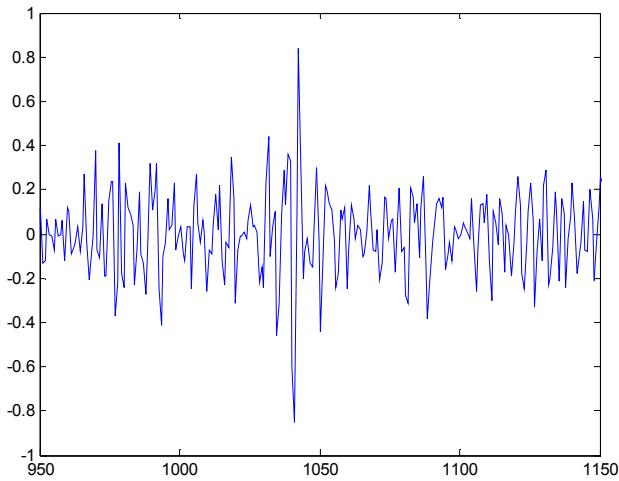


Roses: 13/07/03 18:20h



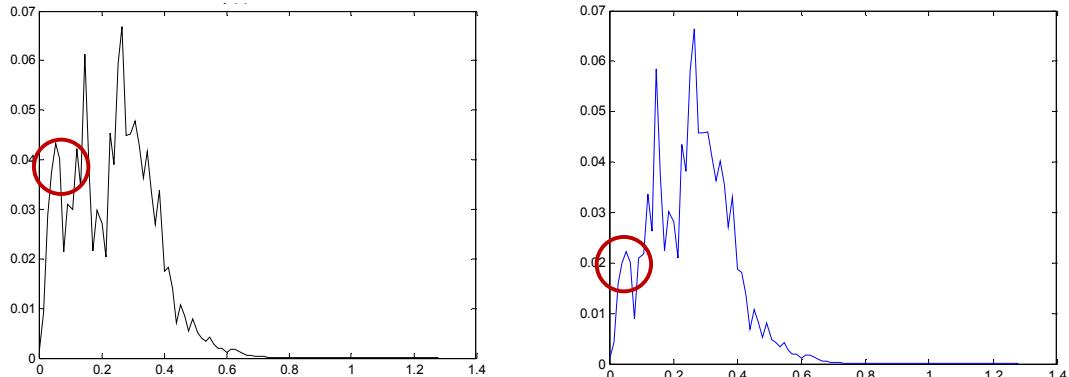
Tortosa: 30/12/05 08:00h





At first sight, the last two are more similar to the academic shape of a freak wave: a single wave which is extremely high compared to the ones around it. In the first case it is not clear. Instead of a freak wave it seems to be a non stationary wave record in which a constant standard deviation cannot be applied for all the record.

It is important to remark that these three freak waves were removed by the quality control because of their low significant wave height. Also, freak waves appear to be related to spectral energy at low frequencies. In the following pictures (x-axis frequency in Hz and y-axis spectral energy in m^2/Hz), there is the original spectrum (left) of the freak wave's record of Roses, and the same, after subtracting the freak wave (right).



If considering the amount of waves of the filtered data (about 12.6 million waves), the 3 freak waves represents 1 freak wave per 4.2 million waves. In linear theory, considering 4.2 million waves, the expected value for the normalized maximum wave height (normalized dividing by the significant wave height) is approx. 2.8, practically the same value as the threshold in the definition of freak wave. Therefore it seems that for this data, the freak waves can be predicted by the linear theory although it may be a coincidence.

APPENDIX C: MATLAB code

The Matlab code, that was developed for the present study, basically consists of two programs which process the total amount of data in an automatic way. Each buoy-year is analysed separately and then saved in *.mat* files. The first program reads all the raw data of the selected year-buoy and proceeds with a part of the quality analysis more related to rough errors. Afterwards, in the second program, the rest of the quality analysis is done and also the statistical and spectral analysis. Finally, in order to study particular facets or interactions between parameters, and plot some illustrative figures, smaller complementary programs have been designed and used, extracting the needed information of each file. An example of such complementary programs is the one which concatenate the records in order to obtain the observations of maximum wave heights for different number of waves.

For the Norwegian data the first part has been modified and adapted to the different file types.

1 First analysis

```

clear all
year=input('Year: ','s');
buoy=input('Buoy(Tortosa/Llobregat/Roses/Blanes): ','s');
%-----Define parameters
g=9.81;
%-----Define the parameters of the number of errors
e1=0; %The lenght is not correct
e2=0; %The vertical acceleration is larger than a half g
e3=0; %There are "gaps"
%-----Define characteristics parameters of each buoy/year
if strcmp(buoy,'Tortosa')==1
    dt=1/1.28;
    if
strcmp(year,'2001')==1|strcmp(year,'2002')==1|strcmp(year,'2004')==1|strcmp(year,'2005')==1|strcmp(year,'2006')==1
        len=1535;
    else
        len=1536;
    end
    ncolumn=4;
    colH=2;
    ext='.RAW';
else
    dt=1/2.56;
    if strcmp(buoy,'Llobregat')==1
        if
strcmp(year,'2001')==1|strcmp(year,'2002')==1|strcmp(year,'2003')==1|strcmp(year,'2004')==1
            ext='.1RW';
            len=3072;
            ncolumn=1;
    end
end

```

```

        colH=1;
    else
        ext='.RAW';
        len=1535;
        ncolumn=4;
        colH=1;
    end

end
if strcmp(buoy,'Roses')==1
    ext='.2RW';
    len=3072;
    ncolumn=1;
    colH=1;
end
if strcmp(buoy,'Blanes')==1
    ext='.3RW';
    len=3072;
    ncolumn=1;
    colH=1;
end
end

%The interval time for all the buoys is the same. The duration may be
different
D=dt*len;
t=dt/2:dt:D-dt/2;
t=t';
data=[];
cont_data2=[];
cont_data3=[];
cont=[];
conter1=[];
conter2=[];
conter3=[];
er3='No';
er3bis='No';
l=1;
%-----Quality control
%-----Yearly analysis
filenames=dir(['D:/Els         meus         Documents/5è      de
camins/tesina/dvd/rawdata/',year,'/',buoy,'/*',ext,'']);
for i=1:length(filenames)
    fid=fopen(['D:/Els         meus         Documents/5è      de
camins/tesina/dvd/rawdata/',year,'/',buoy,'/',filenames(i).name,'']);
    if strcmp(buoy,'Tortosa')==1
        r=fscanf(fid,'%f,%f,%f,%f',[ncolumn,len]);
    else
        r=fscanf(fid,'%f',[ncolumn,len]);
    end
    fclose(fid);
    r=r';
    rcolumnH=r(:,colH)/100;%units m
    rtt=diff(rcolumnH,2)/(dt^2);%second derivative
    rtt_ind=find(abs(rtt)<10^(-4));
    if length(rcolumnH)==len;%-----Length control

```

```

if max(abs(rtt))<(1/2)*g%-----Spikes control
    while l<length(rtt_ind)-1%-----"Gaps" control
        a=rtt_ind(l);
        b=rtt_ind(l+1)-1;
        c=rtt_ind(l+2)-2;
        if
            a==b;% (rtt_ind(l)+2)==(rtt_ind(l+1)+1)==rtt_ind(l+2);
                if b==c;
                    ll=l;
                    l=2*length(rtt);
                    l2=ll+3;
                    %er3='Yes';
                end
            end
            l=l+1;
        end
    %
    l2=ll+3;
    if l>length(rtt)
        while l2<length(rtt_ind)-1;
            if strcmp(er3,'No')==1;
                a=rtt_ind(l2);
                b=rtt_ind(l2+1)-1;
                c=rtt_ind(l2+2)-2;
                if a==b;
                    if b==c;
                        er3='Yes';
                        l=1;
                    end
                end
                l2=l2+1;
            else
                l2=length(rtt_ind);
            end
        end
    end
    if strcmp(er3,'No')==1
        data=[data,rcolumnH];
        siz=size(data);
        j=siz(1,2);
        cont=[cont;i,j];
    else
        cont_data3=[cont_data3,rcolumnH];
        disp(['Error',filenames(i).name,'']);
        conter3=[conter3;i];
        e3=e3+1;
    end
else
    cont_data2=[cont_data2,rcolumnH];
    disp(['The vertical acceleration is larger than
1/2g in the record ',filenames(i).name,'']);
    e2=e2+1;
    conter2=[conter2;i];
end
else
    disp(['The length of the time record is not the
appropriate ',filenames(i).name,'']);
    e1=e1+1;

```

```

        conter1=[conter1;i];
    end
    %i=i+1;
    er3='No';
    l=1;
end
%-----Remove trend
y=detrend(data);
s=size(y);
n=s(1,2);%number of records
%-----Checking
check=10;%number of randomly chosen records
u=ceil(rand(check,1)*n);
figure(1)
for i=1:check;
    subplot(5,2,i),plot(t,y(:,u(i)));
    title(['',buoy,' ',filenames(cont(u(i),1)).name,' ',year,'']);
end
%Error 2
if e2>0
    uc2=ceil(rand(check,1)*length(conter2));
    figure(2)
    for i=1:check;
        subplot(5,2,i),plot(t,cont_data2(:,uc2(i)));
        title(['',buoy,' ',filenames(conter2(uc2(i),1)).name,' ',year,'']);
    end
end
%Error 3
if e3>0
    uc3=ceil(rand(check,1)*length(conter3));
    figure(3)
    for i=1:check;
        subplot(5,2,i),plot(t,cont_data3(:,uc3(i)));
        title(['',buoy,' ',filenames(conter3(uc3(i),1)).name,' ',year,'']);
    end
end
%-----Save the results
%save(['C:/THESISnew/RESULTS/data',year,buoy,'']);
save([D:/Els meus Documents/5è camins/tesina/NOU/RESULTATSqualitycontrol/',year,buoy,'']); de

```

2 Second analysis

```

clear all
year=input('Year: ','s');
buoy=input('Buoy(Tortosa/Llobregat/Roses/Blanes): ','s');
tap=input('Tapering?(Y/N) ','s');
load(['D:/Els meus Documents/5è camins/tesina/NOU/RESULTATSqualitycontrol/',year,buoy,'.mat']); de
%Density definition
if strcmp(buoy,'Tortosa')==1
    h=60;
elseif strcmp(buoy,'Roses')==1

```

```

h=46;
elseif strcmp(buoy,'Llobregat')==1
    h=45;
elseif strcmp(buoy,'Blanes')==1
    h=74;
end
resp='No'

%=====Statistical analysis=====
dev=std(y); %Standard deviation
H=[];
T=[];
wave=[];
not_wave=[];
chicrest=[];
chitrough=[];
cont_prova=[];
num_1third=[];
Hmax=[];
H1third=[];
H1third_crest=[];
chilthird=[];
chilthird_trough=[];
chilthird_H=[];
chilthird_trough_H=[];
chimax=[];
chimin=[];
num_not_wave=[];
num_wave=[];
Hmean=[];
Tmean=[];
T1third=[];
Hrms=[];
cont_new=[];
cont_notrecord=[];
chimean=[];
chimean_trough=[];
chirms=[];
chirms_trough=[];
relHchicrest=[];
relHchitrough=[];
relHchicrest2=[];
relHchitrough2=[];
e10=0;
e_fix2=0;

figure(1)%Rayleigh distribution (with Hnorm)
subplot(1,2,1)
plot((0:0.25:9),exp(log(-8*log(Ray((0:0.25:9)/4))))/2),'r')
hold on
subplot(1,2,2)
semilogy((0:0.25:9),Ray((0:0.25:9)/4),'r')
hold on
ii=0;
iii=0;
for i=1:n;
    ind_zero=find(diff(sign(y(:,i)))== -2 | diff(sign(y(:,i)))== -1);

```

```

for j=1:length(ind_zero)-1
    chi=y(ind_zero(j)+1:ind_zero(j+1),i
        inc_t_zero1=y(ind_zero(j),i)*dt/(y(ind_zero(j),i)-
    y(ind_zero(j)+1,i));
        inc_t_zero2=y(ind_zero(j+1),i)*dt/(y(ind_zero(j+1),i)-
    y(ind_zero(j+1)+1,i));
        tchi_ini_aprox=t(ind_zero(j))+inc_t_zero1;
        tchi_fin_aprox=t(ind_zero(j+1))+inc_t_zero2;

        if (max(chi)-min(chi))>0.05%&tchi_fin_aprox-
tchi_ini_aprox>0%reject very small "waves
            if abs(max(chi))>0.025;
                if tchi_fin_aprox-tchi_ini_aprox>2*dt
                    chicrest=[chicrest,max(chi)];
                    chitrough=[chitrough,min(chi)];
                    H=[H,max(chi)-min(chi)];
                    T=[T,tchi_fin_aprox-tchi_ini_aprox];
                    wave=[wave,j];
                else
                    not_wave=[not_wave,j];
                end
            else
                not_wave=[not_wave,j];
            end
        else
            not_wave=[not_wave,j];
        end
    end
num_1third=[num_1third,floor(length(H)/3)];
H=H';
[H_sort,indexH_sort]=sort(H);

[chicrest_sort,indexchicrest_sort]=sort(chicrest);

[chitrough_sort,indexchitrough_sort]=sort(chitrough);
H1third=[H1third,mean(H_sort((length(H)-
num_1third(length(num_1third))):length(H)))];
chilthird=[chilthird,mean(chicrest_sort((length(H)-
num_1third(length(num_1third))):length(H)))];

Hmax=[Hmax,max(H)];
[m indmax]=max(abs(y(:,i)));
indmax2=indmax+1;
indmax3=indmax-1;
if indmax==len;
    indmax2=indmax-2;% 
end
if indmax==1;
    indmax3=indmax+2;
end
if H1third>0.5;
    if strcmp(resp,'No')
        if
abs(y(indmax,i))<2.83*(chilthird(length(chilthird)))%&(max(abs(y(:,length(Hmax)))))<3*chilthird;%%Threshold
            ii=ii+1;
            prova=sort(H'./chicrest);
    end
end

```

```

relHchicrest=[relHchicrest,prova(floor(length(H)/2))];
    relHchitrough=[relHchitrough,mean(H'./abs(chitrough))];
    relHchicrest2=[relHchicrest2,mean(H')/mean(chicrest)];
    relHchitrough2=[relHchitrough2,
    mean(H')/mean(abs(chitrough))];
    chilthird=[chilthird,mean(chicrest_sort((length(H)-
    num_1third(length(num_1third))):length(H)))];
    chilthird_trough=[chilthird_trough,abs(mean(chitrough_s
    ort((1:num_1third(length(num_1third))))))];
    chilthird_H=[chilthird_H,mean(chicrest(indexH_sort((len
    gth(H)-num_1third(length(num_1third))):length(H))))];
    chilthird_trough_H=[chilthird_trough_H,abs(mean(chitrou
    gh(indexH_sort((length(H)-
    num_1third(length(num_1third))):length(H)))))];
    H1third_crest=[H1third_crest,mean(H(indexxchicrest_sort(
    length(H)-
    num_1third(length(num_1third))):length(H)))];
    chimax=[chimax,max(chicrest)];
    chimin=[chimin,abs(min(chitrough))];
    num_not_wave=[num_not_wave,length(not_wave)];
    num_wave=[num_wave,length(wave)];
    Hnorm_sort=H_sort/dev(i);
    Hnorm_sort=sort(Hnorm);
    Hmean=[Hmean,mean(H)];
    chimean=[chimean,mean(chicrest)];
    chimean_trough=[chimean_trough,abs(mean(chitrough))];
    Hrms=[Hrms,sqrt(mean(H.^2))];
    chirms=[chirms,sqrt(mean(chicrest.^2))];
    chirms_trough=[chirms_trough,sqrt(mean(chitrough.^2))];
    Tmean=[Tmean,mean(T)];
    T1third=[T1third,mean(T(indexH_sort((length(H)-
    num_1third(length(num_1third))):length(H))))];

    ind=[1:length(H)];
    P=ind./length(H);
    subplot(1,2,1)
    plot(Hnorm_sort,exp(log(-8*log(1-P))/2))
    hold on
    subplot(1,2,2)
    semilogy(Hnorm_sort,1-P)
    hold on
    cont_new=[cont_new;cont(i,:),ii];
    elseif
    abs(y(indmax2,i))>2.83*(chilthird(length(chilthird))&abs(y
    (indmax3,i))>2.83*(chilthird(length(chilthird)))
        ii=ii+1;
        prova=sort(H'./chicrest);

    relHchicrest=[relHchicrest,prova(floor(length(H)/2))]
    ;
    relHchitrough=[relHchitrough,mean(H'./abs(chitrough))]
    ;
    relHchicrest2=[relHchicrest2,mean(H')/mean(chicrest)]
    ;
    relHchitrough2=[relHchitrough2,
    mean(H')/mean(abs(chitrough))];

```

```

chilthird=[chilthird,mean(chicrest_sort((length(H)-
num_1third(length(num_1third))):length(H)))];
chilthird_trough=[chilthird_trough,abs(mean(chitrough_s
ort((1:num_1third(length(num_1third))))))];
chilthird_H=[chilthird_H,mean(chicrest(indexH_sort((len
gth(H)-num_1third(length(num_1third))):length(H)))]);
chilthird_trough_H=[chilthird_trough_H,abs(mean(chitrou
gh(indexH_sort((length(H)-
num_1third(length(num_1third))):length(H)))))];
H1third_crest=[H1third_crest,mean(H(indexchicrest_sort(
(length(H)-
num_1third(length(num_1third))):length(H))))];
chimax=[chimax,max(chicrest)];
chimin=[chimin,abs(min(chitrough))];
num_not_wave=[num_not_wave,length(not_wave)];
num_wave=[num_wave,length(wave)];
Hnorm_sort=H_sort/dev(i);
Hmean=[Hmean,mean(H)];
chimean=[chimean,mean(chicrest)];
chimean_trough=[chimean_trough,abs(mean(chitrough))];
Hrms=[Hrms,sqrt(mean(H.^2))];
chirms=[chirms,sqrt(mean(chicrest.^2))];
chirms_trough=[chirms_trough,sqrt(mean(chitrough.^2))];
Tmean=[Tmean,mean(T)];
T1third=[T1third,mean(T(indexH_sort((length(H)-
num_1third(length(num_1third))):length(H))))];
ind=[1:length(H)];
P=ind./length(H);
subplot(1,2,1)
plot(Hnorm_sort,exp(log(-8*log(1-P))/2))
hold on
subplot(1,2,2)
semilogy(Hnorm_sort,1-P)
hold on
cont_new=[cont_new;cont(i,:),ii];
cont_prova=[cont_prova;i];
else
iii=iii+1;
num_1third(length(num_1third))=[];
H1third(length(H1third))=[];
jaja=chilthird;
jeje=max(chicrest);
jojo=max(abs(chitrough));
chilthird(length(chilthird))=[];
Hmax(length(Hmax))=[];
dev(length(Hmax))=[];
y(:,length(Hmax))=[];
cont_notrecord=[cont_notrecord,i];
end
else
e_fix2=e_fix2+1;
num_1third(length(num_1third))=[];
H1third(length(H1third))=[];
chilthird(length(chilthird))=[];
Hmax(length(Hmax))=[];
% cont_notrecord=[cont_notrecord;cont(i,:)];
cont_notrecord=[cont_notrecord,i];

```

```

        end
    else
%
    e10=e10+1;

        num_1third(length(num_1third))=[];
        H1third(length(H1third))=[];
        ch1third(length(ch1third))=[];
        Hmax(length(Hmax))=[];
        cont_notrecord=[cont_notrecord,i];
    end

    T=[];
    H2=H;
    H=[];
    not_wave=[];
    wave=[];
    chicrest=[];
    chitrough=[];
end
perc_not_wave=(num_not_wave./num_wave)*100;
perc_not_wave_mean=mean(perc_not_wave);
saveas(1,['D:/Els          meus          Documents/5è           de
CAMINS/tesina/NOU/RESULTATSanalysis1NOU/Figures/Hprob/',year,buoy,''],'
fig')
saveas(1,['D:/Els          meus          Documents/5è           de
CAMINS/tesina/NOU/RESULTATSanalysis1NOU/Figures/Hprob/',year,buoy,''],'
emf')
hold off

if length(cont_notrecord)>0.5
    y(:,cont_notrecord)=[];
    dev(:,cont_notrecord)=[];
    s=size(y);
    n=s(1,2);

end
kurt=kurtosis(y);
ske=skewness(y);

%Gaussian distribution
[pr prr]=sort(Hmax);

i=ceil(rand(1,9)*s(1,2));

for l=1:length(i)
    surf=linspace(min(y(:,i(l))),max(y(:,i(l))));
    p=(1/(sqrt(2*pi)*dev(i(l))))*exp(-
(surf.^2)/(2*(dev(i(l))^2)));
    for j=2:length(p)
        int_p(j)=trapz(surf(1:j),p(1:j));
    end

    figure(2)%cdf (normal plot)
    subplot(3,3,1)
    normplot(y(:,i(l)))

```

```

        title(['kurt:           ', num2str(kurt(i(1))), ',      skew:
', num2str(ske(i(1))), ''])
        saveas(2,['D:/Els      meus      Documents/5è      de
CAMINS/tesina/NOU/RESULTATSanalysis1NOU/Figures/cdfgaus/',year,buoy,'']
,'fig')
        saveas(2,['D:/Els      meus      Documents/5è      de
CAMINS/tesina/NOU/RESULTATSanalysis1NOU/Figures/cdfgaus/',year,buoy,'']
,'emf')

figure(3)%pdf (histogram)
subplot(3,3,1)
[bincounts binpositions]=hist(y(:,i(1)),20);
delta=abs(binpositions(1)-binpositions(2));
area_hist=delta*len;
y_norm=normpdf(sort(y(:,i(1))),0,dev(i(1)));
hist(y(:,i(1)),20)
hold on
plot(sort(y(:,i(1))),area_hist*y_norm,'r','LineWidth',2);
title(['kurt:           ', num2str(kurt(i(1))), ',      skew:
', num2str(ske(i(1))), ''])
saveas(3,['D:/Els      meus      Documents/5è      de
CAMINS/tesina/NOU/RESULTATSanalysis1NOU/Figures/pdfgaus/',year,buoy,'']
,'fig')
saveas(3,['D:/Els      meus      Documents/5è      de
CAMINS/tesina/NOU/RESULTATSanalysis1NOU/Figures/pdfgaus/',year,buoy,'']
,'emf')

=====Spectral analysis=====
Y=[];
E=[];
G=1;

%-----Tapering (optional)
if strcmp(tap,'Y') ==1
%    pertap=input('% tapering? ');
pertap=10;
r=pertap/100;
for i=1:s(1,2)
    ytap(:,i)=y(:,i).*tukeywin(len,r);
end
ybef=y;
y=ytap;
G=1-5/(8*(1/r));
end

%-----Without splitting the record
if strcmp(buoy,'Tortosa') ==1
len2=1536;
else
len2=3072;
end
fftw('planner','hybrid');%optimise the fft method

Y=fft(y,len2);
df=1/D;
D2=len2*dt;
df2=1/D2;

```

```

freq2=[0:len2-1]*df2;
%define the density spectrum
E=(1/G)*(1/df)*(1/2)*(abs(Y)/(len/2)).^2;
E([1,2],:)=zeros(2,n);

%-----Splitting
p=16;%24;%input('Number of segments: ');
dsf2=p*df2;
num2=(len2)/p;
freqq2=[0:num2-1]*dsf2;

for i=1:num2
    for j=1:n
        EE(i,j)=(1/p)*sum(E(1+(i-1)*p:i*p,j));
    end
end
dsf=p*df;

%Definition of the "real" spectrum (0.03Hz-Nyq freq)
S=EE(1:num2/2+1,:);%I do not consider f<0.03Hz
Sini=E(1:len2/2+1,:);
f=freqq2(1:num2/2+1);
fini=freq2(1:len2/2+1);
fNyq=1/(2*dt);
EE=[];
Y=[];

```

%Aliasing

```

m=[];
for i=1:5
    for j=1:s(1,2)
        int=(f'.^(i-1)).*S(:,j);
        m(i,j)=trapz(f,int);
    end
end
m0=m(1,:); m1=m(2,:); m2=m(3,:);
Tm=m0./m1;
fm=1./Tm;
ind_aliasing=find(fNyq./fm<2.2);
e_aliasing=length(ind_aliasing);

H1third(ind_aliasing)=[];
ch1lthird(ind_aliasing)=[];
relHchicrest(ind_aliasing)=[];
Hmax(ind_aliasing)=[];
num_1third(ind_aliasing)=[];
relHchitrough(ind_aliasing)=[];
relHchicrest2(ind_aliasing)=[];
relHchitrough2(ind_aliasing)=[];
ch1lthird_trough(ind_aliasing)=[];
ch1lthird_H(ind_aliasing)=[];
ch1lthird_trough_H(ind_aliasing)=[];
H1third_crest(ind_aliasing)=[];
chimin(ind_aliasing)=[];
chimax(ind_aliasing)=[];
num_not_wave(ind_aliasing)=[];

```

```

num_wave(ind_aliasing)=[];
m0(ind_aliasing)=[];
m1(ind_aliasing)=[];
m2(ind_aliasing)=[];

Hmean(ind_aliasing)=[];
chimean(ind_aliasing)=[];
chimean_trough(ind_aliasing)=[];
Hrms(ind_aliasing)=[];
chirms(ind_aliasing)=[];
chirms_trough(ind_aliasing)=[];
Tmean(ind_aliasing)=[];
T1third(ind_aliasing)=[];
perc_not_wave(ind_aliasing)=[];
kurt(ind_aliasing)=[];
ske(ind_aliasing)=[];
dev(ind_aliasing)=[];
y(:,ind_aliasing)=[];
ybef(:,ind_aliasing)=[];
S(:,ind_aliasing)=[];
Sini(:,ind_aliasing)=[];
cont_new(ind_aliasing,:)=[];

s=size(y);
n=s(1,2);

%Bump
E_zero=S(1,:);
ind_bump=find(E_zero>0.004);
e_bump=length(ind_bump);

prova=mean(4*sqrt(m0(ind_bump)))

H1third(ind_bump)=[];
chi1third(ind_bump)=[];
relHchicrest(ind_bump)=[];
Hmax(ind_bump)=[];
num_1third(ind_bump)=[];
relHchitrough(ind_bump)=[];
relHchicrest2(ind_bump)=[];
relHchitrough2(ind_bump)=[];
chi1third_trough(ind_bump)=[];
chi1third_H(ind_bump)=[];
chi1third_trough_H(ind_bump)=[];
H1third_crest(ind_bump)=[];
chimin(ind_bump)=[];
chimax(ind_bump)=[];
num_not_wave(ind_bump)=[];
num_wave(ind_bump) [];

m0(ind_bump)=[];
m1(ind_bump)=[];
m2(ind_bump) [];

Hmean(ind_bump)=[];
chimean(ind_bump)=[];

```

```

chimean_trough(ind_bump)=[] ;
Hrms(ind_bump)=[] ;
chirms(ind_bump)=[] ;
chirms_trough(ind_bump)=[] ;
Tmean(ind_bump)=[] ;
T1third(ind_bump)=[] ;
perc_not_wave(ind_bump)=[] ;
kurt(ind_bump)=[] ;
ske(ind_bump)=[] ;
dev(ind_bump)=[] ;
y(:,ind_bump)=[] ;
ybef(:,ind_bump)=[] ;
S(:,ind_bump)=[] ;
Sini(:,ind_bump)=[] ;
cont_new(ind_bump,:)=[] ;

s=size(y) ;
n=s(1,2) ;

%Deep water
T0=sqrt(m0./m2) ;
L0=g*T0.^2/(2*pi);%Hypothesis: deep water

for i=1:length(L0) ;
    if h>=L0(i)/2 %Checking the hypothesis of deep water
        water(i)=1;%Deep water
    elseif h<L0/20
        water(i)=2;%Shallow water
    else
        water(i)=3;%Intermediate water
        dist(i)=L0(i)/2-h;
    end
end

ind_shallow=find(water>1) ;
e_shallow=length(ind_shallow) ;

H1third(ind_shallow)=[] ;
chilthird(ind_shallow)=[] ;
relHchicrest(ind_shallow)=[] ;
Hmax(ind_shallow)=[] ;
num_1third(ind_shallow)=[] ;
relHchitrough(ind_shallow)=[] ;
relHchicrest2(ind_shallow)=[] ;
relHchitrough2(ind_shallow)=[] ;
chilthird_trough(ind_shallow)=[] ;
chilthird_H(ind_shallow)=[] ;
chilthird_trough_H(ind_shallow)=[] ;
H1third_crest(ind_shallow)=[] ;
chimin(ind_shallow)=[] ;
chimax(ind_shallow)=[] ;
num_not_wave(ind_shallow)=[] ;
num_wave(ind_shallow)=[] ;
Hmean(ind_shallow)=[] ;
chimean(ind_shallow)=[] ;

```

```

chimean_trough(ind_shallow)=[];
Hrms(ind_shallow)=[];
chirms(ind_shallow)=[];
chirms_trough(ind_shallow)=[];
Tmean(ind_shallow)=[];
T1third(ind_shallow)=[];
perc_not_wave(ind_shallow)=[];
kurt(ind_shallow)=[];
ske(ind_shallow)=[];
dev(ind_shallow)=[];
y(:,ind_shallow)=[];
ybef(:,ind_shallow)=[];
S(:,ind_shallow)=[];
Sini(:,ind_shallow)=[];
cont_new(ind_shallow,:)=[];

s=size(y);
n=s(1,2);

%Plot randomly spectra
check=10;%number of randomly chosen spectra
u=ceil(rand(check,1)*s(1,2));
figure(4)
for i=1:check;
    subplot(5,2,i),
    plot(fini(:,Sini(:,u(i))),'b')%%the area of interest is f=<fnyq
    hold on
    plot(f(:,S(:,u(i)),'k')
    title([''',buoy,'          ',filenames(cont_new(u(i),1)).name,
',year,'']);
    end
    saveas(4,['D:/Els           meus           Documents/5è           de
CAMINS/tesina/NOU/RESULTATSanalysis1NOU/Figures/spectra/',year,buoy,']
,'fig')
    saveas(4,['D:/Els           meus           Documents/5è           de
CAMINS/tesina/NOU/RESULTATSanalysis1NOU/Figures/spectra/',year,buoy,']
,'emf')
    hold off

%Spectral parameters
m=[];
for i=1:5
    for j=1:s(1,2)
        int=(f'.^(i-1)).*S(:,j);
        m(i,j)=trapz(f,int);
    end
end
m0=m(1,:); m1=m(2,:); m2=m(3,:); m3=m(4,:); m4=m(5,:);

Hm0=4*sqrt(m0);
chim0=Hm0/2;

HmR=((2*pi*m0).^(1/2));
HrmsR=((8*m0).^(1/2));

chimR=HmR/2;

```

```

chirmsR=Hrms/2;%I have to check if it is directly half the Hrms

%Other parameters
Tm=m0./m1;
T0=sqrt(m0./m2);%Mean period between downcrossings
Tc=sqrt(m2./m4);%Mean period between crests
eps_2=sqrt(1-((m2.^2)./(m0.*m4)));%Spectral width EPS4
nu=sqrt((m0.*m2./(m1.^2))-1);%Spectral width EPS2
for j=1:s(1,2)
    Qp(j)=(2./(m0(j).^2)).*trapz(f,f'.*(S(:,j).^2));
    kappa(j)=sqrt((trapz(f,S(:,j).* (cos(2.*pi.*f.*T0(j))))')).^2+(trapz(f,S(:,j).* (sin(2.*pi.*f.*T0(j))))')).^2)./m0(j);
end

%Peak frequency
[Smax,indfp]=max(S);
fp=f(indfp);
fp=fp';

Ss=(2*pi*Hm0)./(g*(T0.^2));%Significant Steepness

L0=g*T0.^2/(2*pi);
k0=2*pi./(L0);

fftw('wisdom', []);

save(['D:/Els meus Documents/5è de
CAMINS/tesina/NOU/RESULTATSanalysis1NOU/',year,buoy,'']);

```

3 Example of complementary program: Maximum wave height

```

clear all
BFI_tot=[];
BFI_tot2=[];
kurt_tot=[];
ske_tot=[];
num_wave_tot=[];
Hnmax_tot=[];
Hnmean_tot=[];
crestnmax_tot=[];
troughnmax_tot=[];
dev_tot=[];
crest1third_all=[];
H1third_all=[];
m0_all=[];

buoy='Tortosa';
any=[1991,1992,1993,1994,1995,1996,1997,2001,2002,2004,2005,2006];
% any=[2001,2002,2004,2005,2006];
for i=1:length(any);
    year=num2str(any(i));
    load(['D:/Els meus Documents/5è de
CAMINS/tesina/NOU/RESULTATSanalysis1NOU/',year,buoy,'.mat'],'m0','Hmean'

```

```

    ', 'ske', 'BFI', 'kurt', 'num_wave', 's', 'Hmax', 'chimax', 'chimin', 'dev', 'Hm0
    ', 'chim0', 'H1third', 'chilthird', 'chilthird_trough', 'eps_2');
    eps_tot=[eps_tot,eps_2];
    m0_all=[m0_all,m0];
    H1third_all=[H1third_all,H1third];
    crest1third_all=[crest1third_all,chilthird];
    Hnmax_tot=[Hnmax_tot,Hmax./dev];
    Hnmean_tot=[Hnmean_tot,Hmean./dev];
    crestnmax_tot=[crestnmax_tot,chimax./dev];
    troughnmax_tot=[troughnmax_tot,chimin./dev];
    BFI_tot=[BFI_tot,BFI];
    kurt_tot=[kurt_tot,kurt];
    ske_tot=[ske_tot,ske];
    s_tot=[s_tot,s(1,2)];
    dev_tot=[dev_tot,dev];
    num_wave_tot=[num_wave_tot,num_wave];

end
lenT=length(crestnmax_tot);

buoy='Llobregat';
any=[2001,2002,2003,2004];
for i=1:length(any)

    year=num2str(any(i));
    load(['D:/Els          meus          Documents/5è          de
CAMINS/tesina/NOU/RESULTATSanalysis1NOU/',year,buoy,'.mat'],'m0','Hmean
','ske','BFI','kurt','num_wave','s','Hmax','chimax','chimin','dev','Hm0
','chim0','H1third','chilthird','chilthird_trough','eps_2');
    eps_tot=[eps_tot,eps_2];
    m0_all=[m0_all,m0];
    H1third_all=[H1third_all,H1third];
    crest1third_all=[crest1third_all,chilthird];
    Hnmax_tot=[Hnmax_tot,Hmax./dev];
    Hnmean_tot=[Hnmean_tot,Hmean./dev];
    crestnmax_tot=[crestnmax_tot,chimax./dev];
    troughnmax_tot=[troughnmax_tot,chimin./dev];
    BFI_tot=[BFI_tot,BFI];
    kurt_tot=[kurt_tot,kurt];
    ske_tot=[ske_tot,ske];
    s_tot=[s_tot,s(1,2)];
    dev_tot=[dev_tot,dev];
    num_wave_tot=[num_wave_tot,num_wave];
end
lenL=length(crestnmax_tot)-(lenT);

buoy='Roses';
any=[2001,2002,2003,2004,2005,2006];
for i=1:length(any)

    year=num2str(any(i));
    load(['D:/Els          meus          Documents/5è          de
CAMINS/tesina/NOU/RESULTATSanalysis1NOU/',year,buoy,'.mat'],'m0','ske',
'Hmean','BFI','kurt','num_wave','s','Hmax','chimax','chimin','dev','Hm0
','chim0','H1third','chilthird','chilthird_trough','eps_2');
    eps_tot=[eps_tot,eps_2];
    m0_all=[m0_all,m0];

```

```

H1third_all=[H1third_all,H1third];
crest1third_all=[crest1third_all,chilthird];
Hnmax_tot=[Hnmax_tot,Hmax./dev];
Hnmean_tot=[Hnmean_tot,Hmean./dev];
crestnmax_tot=[crestnmax_tot,chimax./dev];
troughnmax_tot=[troughnmax_tot,chimin./dev];
BFI_tot=[BFI_tot,BFI];
kurt_tot=[kurt_tot,kurt];
ske_tot=[ske_tot,ske];
s_tot=[s_tot,s(1,2)];
dev_tot=[dev_tot,dev];
num_wave_tot=[num_wave_tot,num_wave];
end
lenR=length(crestnmax_tot)-(lenL+lenT);

buoy='Blanes';
any=[2002,2003,2004,2005,2006];
for i=1:length(any)

    year=num2str(any(i));
    load(['D:/Els meus Documents/5è de
CAMINS/tesina/NOU/RESULTATSanalysis1NOU/',year,buoy,'.mat'],'m0','Hmean
','ske','BFI','kurt','num_wave','s','Hmax','chimax','chimin','dev','Hm0
','chim0','H1third','chilthird','chilthird_trough','eps_2');
    eps_tot=[eps_tot,eps_2];
    m0_all=[m0_all,m0];
    H1third_all=[H1third_all,H1third];
    crest1third_all=[crest1third_all,chilthird];
    Hnmax_tot=[Hnmax_tot,Hmax./dev];
    Hnmean_tot=[Hnmean_tot,Hmean./dev];
    crestnmax_tot=[crestnmax_tot,chimax./dev];
    troughnmax_tot=[troughnmax_tot,chimin./dev];
    BFI_tot=[BFI_tot,BFI];
    kurt_tot=[kurt_tot,kurt];
    ske_tot=[ske_tot,ske];
    s_tot=[s_tot,s(1,2)];
    dev_tot=[dev_tot,dev];
    num_wave_tot=[num_wave_tot,num_wave];
end
lenB=length(crestnmax_tot)-(lenR+lenT+lenL);

tot=sum(s_tot);
wave_tot=sum(num_wave_tot);

n_vector1=[5*1e2,1e3,2*1e3,5*1e3,1e4,2*1e4,5*1e4,1e5,2*1e5,5*1e5,1e
6,2*1e6,5*1e6,1e7];%,2*1e7];

for c=1:length(n_vector1)
    num_clusters(c)=round(sum(num_wave_tot)/n_vector1(c));
    num_records_cluster(c)=floor(tot/num_clusters(c));
        for j=1:num_clusters(c)
            if j<num_clusters(c)
                Hnmax_cluster(j)=max(Hnmax_tot((j-
1)*num_records_cluster(c)+1:j*num_records_cluster(c)));
                chinmax_cluster(j)=max(crestnmax_tot((j-
1)*num_records_cluster(c)+1:j*num_records_cluster(c)));
            end
        end
    end
end

```

```

chinmin_cluster(j)=max(roughnmax_tot((j-
1)*num_records_cluster(c)+1:j*num_records_cluster(c)));
N(j)=sum(num_wave_tot((j-
1)*num_records_cluster(c)+1:j*num_records_cluster(c)));
else
    Hnmax_cluster(j)=max(Hnmax_tot((j-
1)*num_records_cluster(c)+1:tot));
    chinmax_cluster(j)=max(crestnmax_tot((j-
1)*num_records_cluster(c)+1:tot));
    chinmin_cluster(j)=max(roughnmax_tot((j-
1)*num_records_cluster(c)+1:tot));
    N(j)=sum(num_wave_tot((j-
1)*num_records_cluster(c)+1:tot));
end
end
Hnmaxmean(c)=(mean(Hnmax_cluster))';
chinmaxmean(c)=(mean(chinmax_cluster))';
chinminmean(c)=(mean(chinmin_cluster))';
Nmean(c)=(mean(N))';
N=[];
Hnmax_cluster=[];
chinmax_cluster=[];
chinmin_cluster=[];
end

N=linspace(min(Nmean),max(Nmean),10000);

Ehnmax=2*(1+0.29./log(N)).*sqrt(2*log(N));
Ecrestnmax=(1+0.29./log(N)).*sqrt(2*log(N));

%Calculation of confidence intervals
% DH=sqrt(2);
% SH=DH./sqrt(N);

Dcrest=sqrt((1./(2*log(N))).*(1.6449-2.1515./(log(N))));;
Screst=Dcrest./sqrt(wave_tot./N);

figure(1)
semilogx(N,Ecrestnmax,'b');
axis([min(N) max(N) 3 7.5])
axis manual
hold on
semilogx(N,Ecrestnmax+1.96*Screst,'b--');
semilogx(N,Ecrestnmax-1.96*Screst,'b--');
semilogx(Nmean,chinmaxmean,'or');
semilogx(Nmean,chinminmean,'ob');

```

REFERENCES

- Alber, I. (1978). The effects of randomness of the stability of two-dimensional surface wave trains. *Proc. Roy. Soc. London*, Vol. 363, A, 525-546
- Alber, I. & Saffman, P. (1978). Stability of random nonlinear deepwater waves with finite bandwidth spectra. Tech. Rep. 31326-6035-RU-00, TRW Defense and Space System Group.
- Allender, J., Audunson, T., Barstow, S.F., Bjerken, S., Krogstad, H.E., Steinbakke, P., Vartdal, L., Borgman, L. E. & Graham,C. (1989). The WADIC project: a comprehensive field evaluation of directional wave instrumentation. *Ocean Eng.*, Vol. 16, 5/6, 505-536.
- Al-Humoud, J. & Tayfun, M. A., Askar, H. (2001). Distribution of nonlinear wave crests. *Ocean Eng.*, Vol. 29, 1929-1943.
- Brigham, E. O. (1988). *The fast fourier transform and its applications*. Prince-Hall International Editions.
- Cartwright, D. E. & Longuet-Higgins, M. S. (1956). The statistical distribution of the maxima of a random function. *Proc. Roy. Soc. London*, Vol. 237, A, 212-232.
- Cartwright, D. E. (1958). On estimating the mean energy of sea waves from the highest waves in a record. *Proc. Roy. Soc. London*, Vol. 247, A, 22-48.
- Chen, E., Borgman, L. E. & Yfantis, E. (1979). Height and period distribution of hurricane waves. *Civil Engineering in the Oceans, San Francisco*.
- Earle, M. D. (1975). Extreme wave conditions during hurricane Camille. *J. Geophys. Res.*, Vol. 80, 3, 377-379.
- Forristall, G. Z. (1978). On the statistical distribution of wave heights in a storm. *J. Geophys. Res.*, Vol. 83, C5, 2353-2358.
- Forristall, G. Z. (1984). The distribution of measured and simulated wave heights as a function of spectral shape . *J. Geophys. Res.*, Vol. 89, C6, 10547-10552.
- Forristall, G. Z. (1999). Wave crest distributions: observations and second-order theory . *J. Phys. Oceanogr.*, Vol. 30, 1931-1943
- Goda, Y. (1970). Numerical experiments on wave statistics with spectral simulation. *Rep. Harbour Res. Inst.*, Vol. 9, 3
- Haring, R. E., Osborne, A. R., Spencer, L. P. (1976). Extreme wave parameters based on continental shelf storm wave records. *Proc. 15th Coastal Eng. Conf.*, 151-170
- Holthuijsen, L. H. (2007). *Waves in oceanic and coastal waters*. Cambridge University Press.
- Honda, T. & Mitsuyasu, H. (1975). The statistical distributions for the elevation, velocity and acceleration of the surface of wind waves. *J. Oceanogr. Society of Japan*, Vol. 1, 93-104.
- Janssen, P. A. (2003). Nonlinear four-wave interactions and freak waves. *J. Phys. Oceanogr.*, Vol. 33, 4, 863-884.
- Jha, A. K. & Winterstein, S. R. (2000). Nonlinear random ocean waves: prediction and comparison with data. *Proc. TCE/OMA2000 Joint Conference, Energy for the New Millennium, New Orleans*.

- Lansen, J. A., Holthuisen, L. H., & van der Westhuysen, A. J. (2006). Estimating freak wave occurrence in the Agulhas current with Swan. *Proc. 30th International Coastal Eng. Conf.*, 677-688.
- Lansen, J. A. (2006). Estimating freak wave occurrence in the Agulhas current with Swan. *Master thesis, TU Delft*.
- Longuet-Higgins, M. S. (1952). On the statistical distribution of the heights of sea waves. *J. Mar. Res.*, Vol. XI, 3, 245-267.
- Longuet-Higgins, M. S. (1957). The statistical analysis of a random, moving surface. *Proceedings of the Royal Society of London. Series A*, Vol. 249, 966, 321-387.
- Longuet-Higgins, M. S. (1963). The effect of non-linearities on statistical distributions in the theory of sea waves. *J. Fluid Mech.*, Vol. 17, 459-480.
- Longuet-Higgins, M. S. (1975). On the joint distribution of the periods and amplitudes of sea waves. *J. Geophys. Res.*, Vol. 80, 18, 2688-2694.
- Longuet-Higgins, M. S. (1980). On the distribution of the heights of sea waves: some Effects of nonlinearity and finite band width. *J. Geophys. Res.*, Vol. 85, C3, 1519-1523.
- Longuet-Higgins, M. S. (1984). Statistical properties of wave groups in a random sea state. *Proc. Roy. Soc.London. A*, Vol. 312, 1521, 219-250.
- Longuet-Higgins, M. S. (1985). Accelerations in steep gravity waves. *J. Phys. Oceanogr.*, Vol. 15, 11, 1570-1579.
- Lyons, R. (1998). *Windowing functions improve FFT results*. Test and Measurement World, Parts I and II, June and September
- Massel, S. R. (1996). *Ocean surface waves: their physics and prediction*. Advanced series on ocean engineering – Volume 11. World Scientific.
- Mori, N., & Janssen, P. A. (2006). On kurtosis and occurrence probability of freak Waves. *J. Phys. Oceanogr.*, Vol. 36, 7, 1471-1483.
- Mori, N., & Yasuda, T. (2001). On A weakly non-gaussian model of wave height distribution for random wave train. *J. Phys. Oceanogr.*, Vol. 36, 7, 1471-1483.
- Naess, A. (1985). On the distribution of crest to trough wave heights. *Ocean Eng.*, Vol. 12, 3, 221-234.
- Nolte, K. G., ASCE, M. & Hsu, F. H. (1979). Statistics of larger waves in a sea state. *Ocean Eng.*, Vol. 105, WW4, 389-404.
- Olagnon, M. & Magnusson, A. K. (2004). Spectral parameters to characterize the risk of rogue waves occurrence in a sea state. *Actes de colloques – IFREMER*, Vol. 39.
- Prevosto, M. & Forristall, G. Z. (2004). Statistics of wave crests from models vs. measurements. *J. Offshore Mechanics and Artic Eng.*, Vol. 126, 43-53.
- Rice, S. O. (1945). Mathematical analysis of random noise. *Bell System Technol.*, Vol. 24: 46-156.
- Rodríguez, G., Guedes Soares, C., Pacheco, M. & Pérez-Martell, E. (2002). Wave height distribution in mixed sea states. *J. Offshore Mechanics and Artic Eng.*, Vol. 124, 34-39.
- Rotés Casanova, A. (2004). *Clima espectral del oleaje en la costa catalana*. Dissertation ETSECCPB-UPC, Barcelona.
- Shiavi, R. (1991). *Introduction to Applied Statistical Signal Analysis*. Aksen Associates Incorporated Publishers.
- Sobey, R. J., Chandler, B. D. & Harper, B. A., (1990). Extreme waves and wave counts in a hurricane. *Proceedings of 22nd Coastal Eng. Conference*, 1, 358-370.

- Tayfun, M. A. (1981a). Breaking-limited wave heights. *J. Waterway, Port, Coastal and Ocean Eng.*, ASCE, Vol. 107, 59-69.
- Tayfun, M. A. (1981b). Distribution of crest-to-trough wave heights. *J. Waterway, Port, Coastal and Ocean Eng.*, ASCE, Vol. 107, 149-158.
- Tayfun, M. A. (1983). Nonlinear effects on the distribution of crest-to-trough wave heights. *Ocean Eng.*, Vol. 10, 2, 97-106.
- Tayfun, M. A. (1984). Nonlinear effects of the distribution of amplitudes of sea waves. *Ocean Eng.*, Vol. 11, 3, 245-264.
- Tayfun, M. A. (1990). Distribution of large wave heights. *J. Waterway, Port, Coastal and Ocean Eng.*, ASCE, Vol. 116, 6, 686-707.
- Tayfun, M. A. (1994). Distributions of envelope and phase in weakly nonlinear random waves. *J. Eng. Mechan.*, Vol. 120, 5, 1009-1025.
- Tayfun, M. A. (2004). Statistics of wave crests in storms. *J. Waterway, Port, Coastal and Ocean Eng.*, ASCE, Vol. 119, 2, 172-192.
- Tayfun, M. A. (2006). Statistics of nonlinear wave crests and groups. *Ocean Eng.*, Vol. 33, 1589-1622.
- Tayfun, M. A. (2007). Wave-height distributions and nonlinear effects. *Ocean Eng.*, Vol. 34, 1631-1649.
- Thompson, E. F. (1974). Results from CERC wave measurement program, paper presented at International Symposium on Ocean Wave Measurement and Analysis, ASCE, New Orleans, Sept. 9-11
- Tucker, M. & Pitt, E. G. (2001). *Waves in ocean engineering*. Elsevier ocean engineering book series – Volume 5.
- Vinje, T. (1989). The statistical distribution of wave heights in a random seaway. *Applied Ocean Res.*, Vol. 11, 3, 143-152.
- Vrijling, J. K. & van Gelder, P. H. A. J. M. (2006). Probabilistic Design in Hydraulic Engineering. Lecture notes, TU Delft.

OTHER CONSULTED BIBLIOGRAPHY

- Barstow, F. S., Krogstad, H. E., Lønseth, L., Mathisen, J. P. Mørk, G. & Schjølberg, P. (2002). Intercomparison of seastate and zerocrossing parameters from the WACSIS field experiment and interpretation using video evidence. *Proc. 21st International Offshore Mec. Artic Eng. Conf.*, Oslo.
- Beji, S. (1995). Note on a nonlinearity parameter of surface waves. *Coastal Eng.*, Vol. 25, 81-85.
- Bolaños, R., Rotés, A. & Sánchez-Arcilla, A. (2004). Spectral wave climate at northern Spain's Mediterranean coast. Laboratoty of Maritime Engineering, ETSECCPB-UPC, Barcelona.
- Botelho Machado, U. (2003). Probability density funtions for non-linear random waves and responses. *Ocean Eng.*, Vol. 30, 1027-1050.
- Canavos, G. C. (1988). *Probabilidad y Estadística: Aplicaciones y Métodos*. McGrawHill.
- Fedele, F. (2006). Extreme events in nonlinear random seas. *J. Offshore Mechanics and Artic Eng.*, Vol. 128, 11-16.
- Frigo, M. & Johnson, S. (1998). FFTW: an adaptative software architecture for the FFT. In *ICASSP 98*, Vol. 3, 1381-1384
- Gómez, J., Espino, M., Puigdefàbregas, J., Jerez, F. & Cateura, J. (2007). Boies d'onatge. Dades obtingudes l'any 2007. Annual report of XIOM.
- Larsen, L. H. (1981). The influence of bandwith on the distribution of heights of sea states *J. Geophys. Res.*, Vol. 86, C5, 4299-4301.
- Leadbetter, M. R. (1966). On crossings of levels and curves by a wide class of stochastic processes. *The Annals of Mathematical Statistics*, Vol. 37, 1, 260-267.
- Leadbetter, M. R. & Spaniolo, G. V. (2002). On statistics at level crossings by a stationary process. *Statistica Neerlandica*, Vol. 56, 2 152-164.
- Leadbetter, M. R. & Spaniolo, G. V. (2002). On statistics at level crossings by a stationary process. *Statistica Neerlandica*, Vol. 56, 2 152-164.
- Lds-group (2003). Understanding FFT Windows.
- Mori, N. (2004). Ocurrence probability of a freak wave in nonlinear wave field. *Ocean Eng.*, Vol. 31, 165-175.
- Ochi, M. K., (1998). *Ocean waves: The stochastic approach*. Cambridge Ocean Technology Series 6.
- Percival, D. B. & Walden, A. T. (1993). *Spectral analysis for physical applications: Multiplayer and conventional univariate techniques*. Cambridge University Press.
- Losada, M. A. & Corniero, M. A. (1982). Una nueva aproximación a la función de distribución de la altura de ola máxima. *Revista de obras publicas*, enero 1982, 871-878.
- Tayfun, M. A. (1993). Sampling-rate errors statistics of wave heights and periods. *J. Waterway, Port, Coastal and Ocean Eng.*, ASCE, Vol. 130, 4, 155-161.
- Day, R. A. (1993). *How to write and publish a scientific paper*. 3r edition. Cambridge University Press.
- Rydén, J. (2006). A note on asymptotic approximations of distributions for maxima of wave crests. *Stochastic environmental research and risk assessment*, Vol. 20, 4, 238-242.
- Rychlik, I. (2000). On some reliability applications of Rice's formula for the intensity of level crossings. *Extremes*, Vol. 3, 4, 331-348

- Shiavi, R. (1991). *Introduction to applied statistical signal analysis*. Aksen Associates Incorporated Publishers, IRWIN.
- Socquet-Juglard, H. (2005). Spectral evolution and probability distributions of surface ocean gravity waves and extreme waves. Ph. D. thesis, University of Bergen, Norway.
- Srokosz, M. A. (1998). A new statistical distribution for the surface elevation of weakly nonlinear water waves. *Notes and correspondence*, Vol. 28, 149-155.