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.

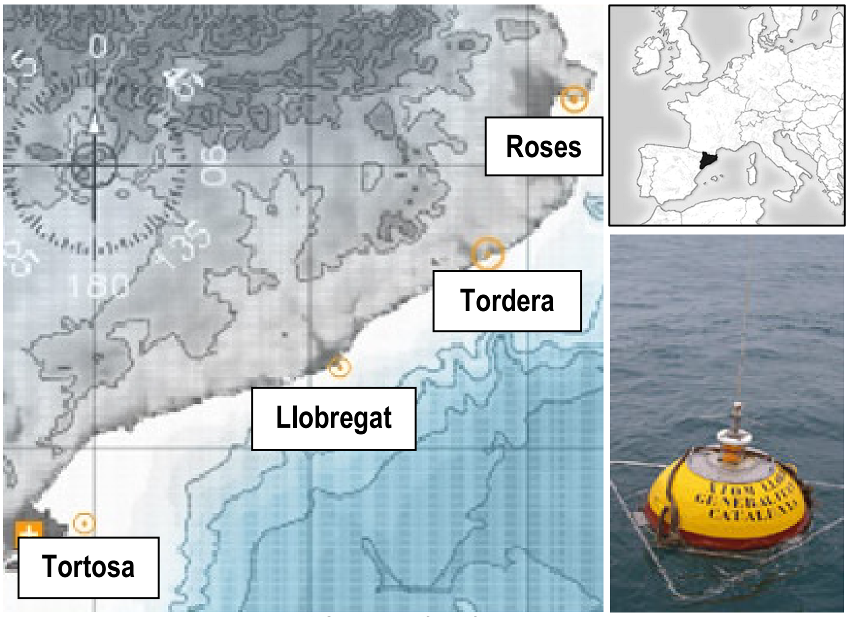


Figure 2.1 Situation of the four buoys

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).

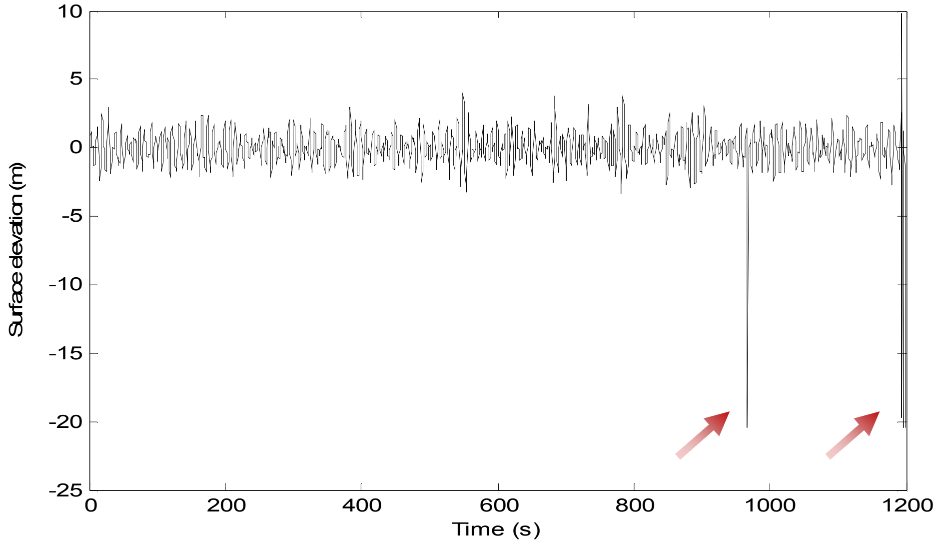


Figure 2.2 Record with spikes (Tortosa 29/03/2004 09:00h)

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).

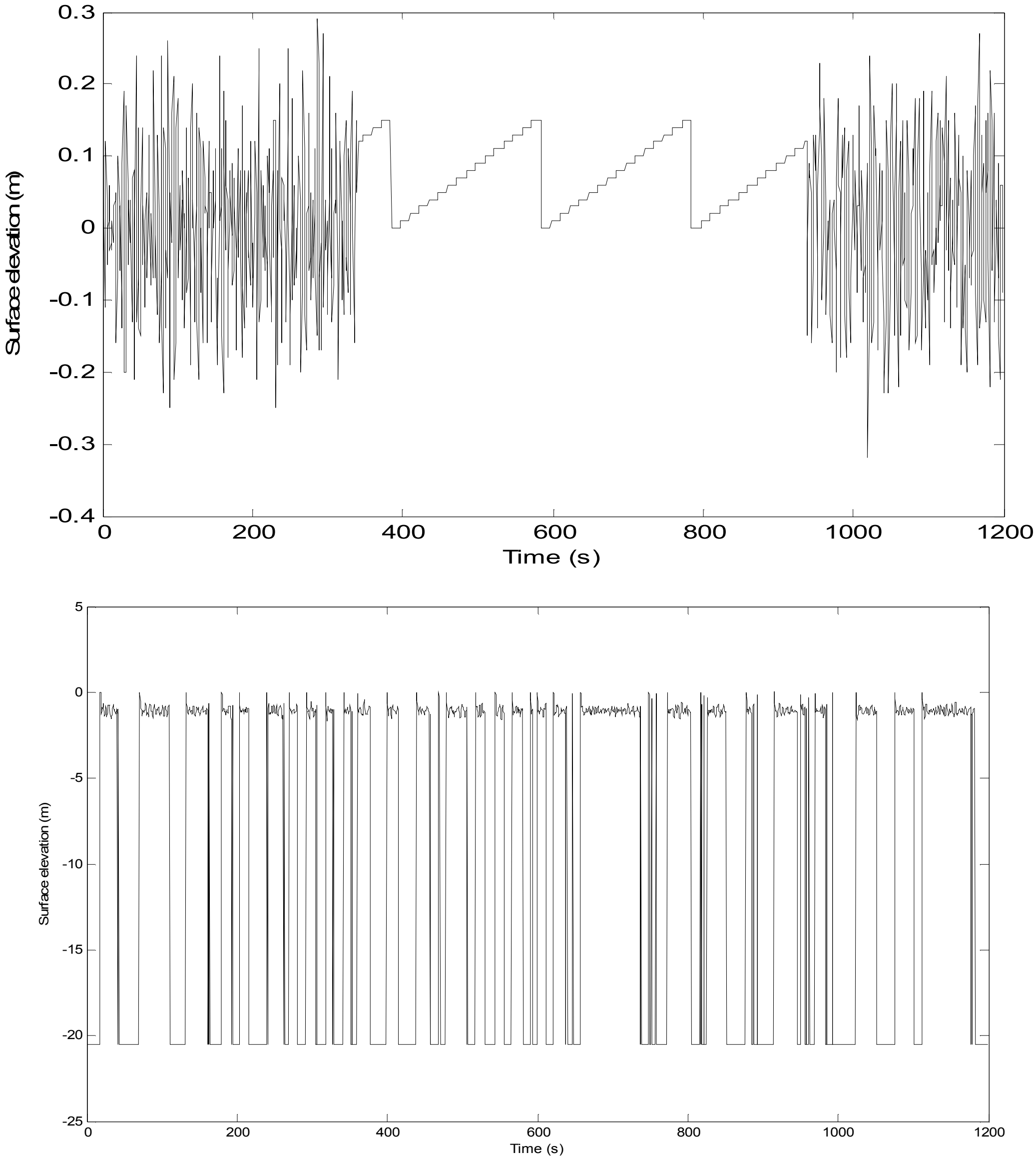


Figure 2.3 Upper panel: Record with a large gap in the middle (Tortosa 14/07/2004 02:00h); lower panel: Record with several gaps of constant value (Roses 30/04/2004 15:20h)

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 120º
- 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).

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

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:

$$H \geq 2.83H_s \tag{2.2}$$

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):

$$Length = D/\Delta t \tag{2.3}$$

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:

$$\frac{f_N}{f_m} < 2.2 \tag{2.4}$$

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.

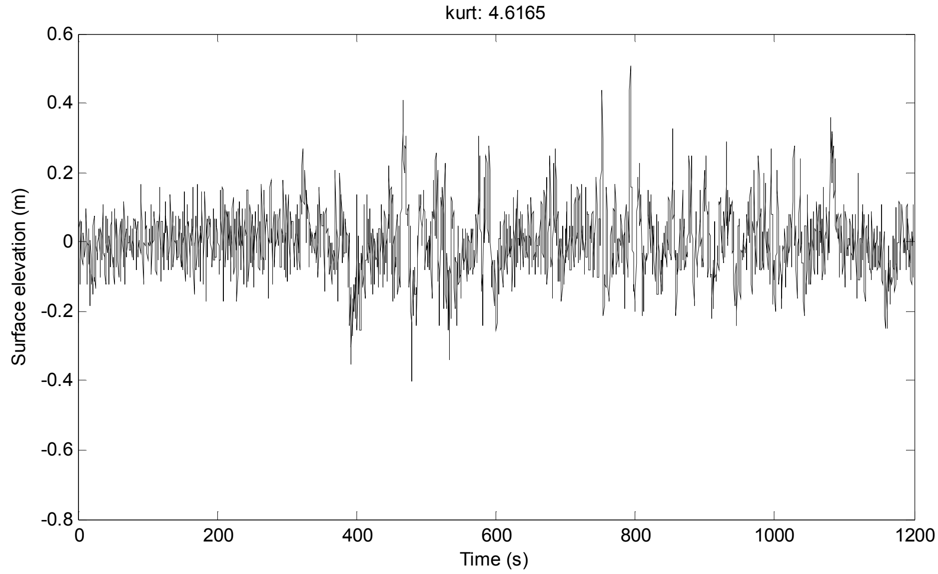
Bump

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 m^2/Hz has been found acceptable, very similar to the one used by Rotés (2004).

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

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.



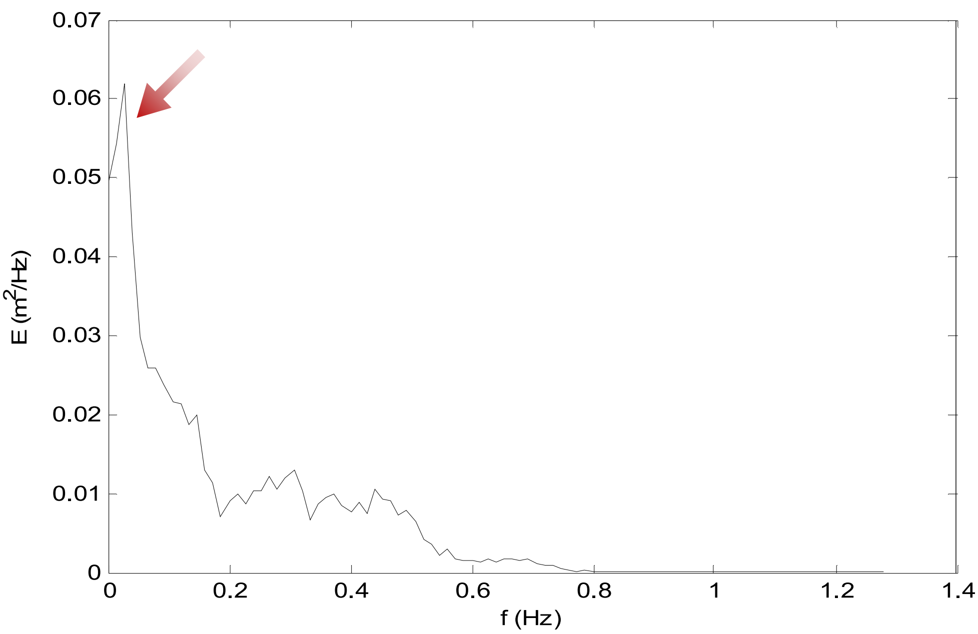


Figure 2.4 Surface elevation and associated spectrum of a distorted record, probably due to a bump against the buoy

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 adittion, 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.

$$kurt > 4, \quad |ske| > 0.3 \tag{2.6}$$

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:

$$h < L_0/2 \tag{2.7}$$

where h is the depth and L_0 the wave length:

$$L_0 = \frac{gT_0}{2\pi} \tag{2.8}$$

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

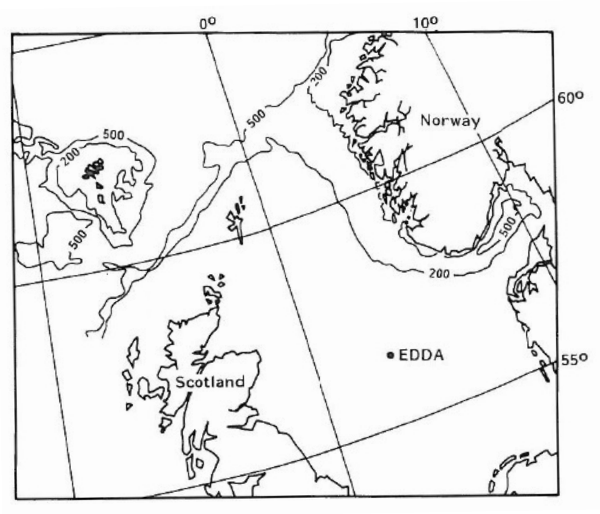


Figure 2.5 The location of the WADIC experiment, near the EDDA platform (Allender et. al., 1989)

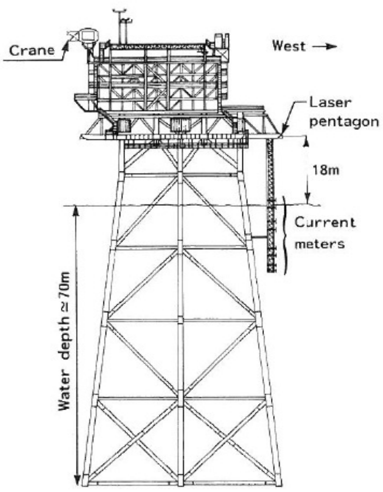


Figure 2.6 Instrument locations on the Edda platform (Allender et al., 1989)

[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.