CHAPTER 61

DESIGN WAVES AND THEIR PROBABILITY DENSITY FUNCTIONS

by

J Rossouw

Abstract

Design wave heights were estimated from measured data using a virtually continuous data set consisting of 8 years of waverider data. Evidence is given which shows that waveriders tend to malfunction during storms. Special care was taken to select independant and identically distributed samples from the data before fitting a number of probability distributions to the selected wave heights. Bootstrap techniques were used to select the models that give the best fit to the data as well as to determine the confidence bands for the predicted design wave heights. It is shown that once the model for the long term distribution of wave height is chosen, relatively narrow confidence bands can be obtained for the most probable value of up to the 100 year return period wave if maximum use is made of the available data. Uncertainty about the selection of the model and the representativeness of the measurements however reduces the usefulness of these confidence bands. A plea is also made in the paper to stop using the concept of a wave with a certain recurrence interval but rather to specify a wave with a given risk of being exceeded within the design life of the structure.

1. The Data Set

Data obtained from the four sites shown in Figure 1 were used to compile a data set which gave a 90 per cent coverage over an 8 year period. These four sites from Cape Town in the West to Port Elizabeth in the East showed remarkable similarity in simultaneously recorded wave heights over distances as much as 700 km apart. An example of simultaneously recorded wave heights at stations 2 and 3 is shown in Figure 2. Similar examples were also shown in Rossouw et al (1982). The reason for the similarity in wave height over such large areas is due to the large size of the weather systems responsible for the higher wave events. A typical example of such a weather system is shown in Figure 3. Large spacial

---

Senior Lecturer, Dept of Civil Engineering, University of Stellenbosch, Republic of South Africa

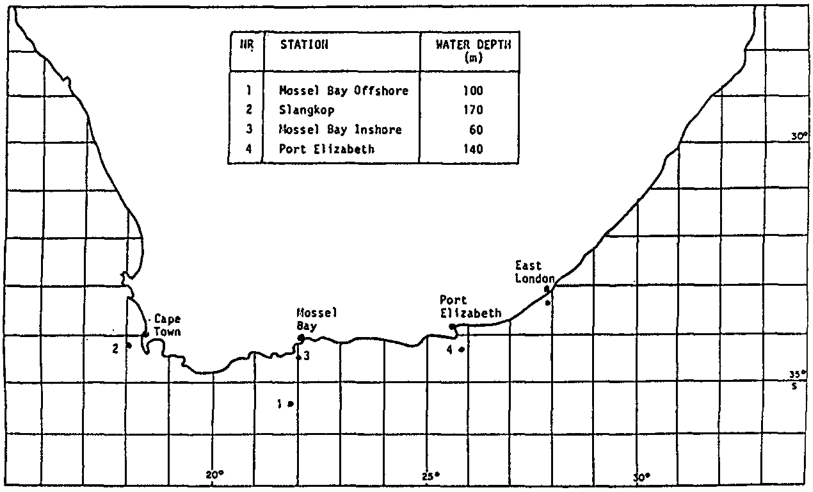


FIGURE 1: RECORDING SITES

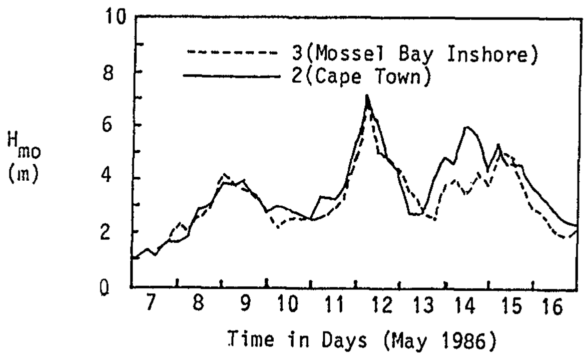
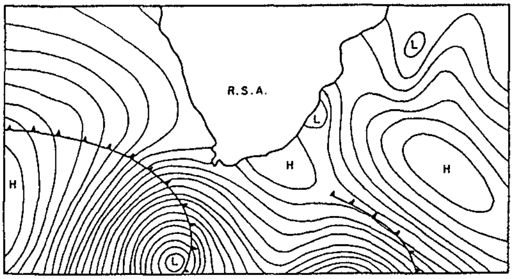


FIGURE 2: COMPARISON IN H_{mo} BETWEEN CAPE TOWN AND MOSSEL BAY



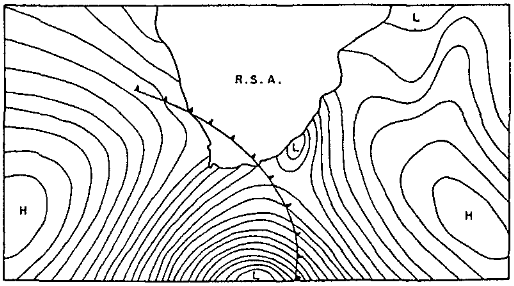


FIGURE 3: PASSAGE OF A COLD FRONT CAUSING HIGH WAVES ALONG THE RSA SOUTHERN COAST (ROSSOUW (1982))

variation in wave heights is not to be expected within such a weather system. These weather systems regularly move from West to East past the South African south coast and generates high waves at the four recording sites shown in Figure 1.

The data set was compiled by using the measurements at site 1 as basis and by filling gaps in this data set with data from stations 2, 3 and 4 in that order of preference. In this process it was noted that the waverider has a tendency to malfunction during the peak of the storms. Examples of this is shown in Figure 4. After filling the gaps a careful study was made of the weather maps during the periods where no records were available from any of the four stations, to ensure that no major storms were ommitted from the data set.

2. Sampling from the data set

The basic data set consisted of 10 537 values of Hmo recorded at 6 hourly intervals. To make the maximum use of this data set it is important to obtain the maximum number of samples from the data set which will be independant and identically distributed.

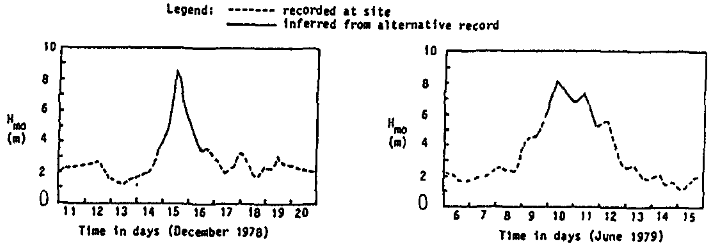


FIGURE 4: TWO EXAMPLES WERE WAVERIDER MALFUNCTIONED DURING PEAK OF STORM

2.1 Independence of the data

To test for independence between the recorded Hmo values, the serie-correlation coefficient were calculated using a lag of 6h, 12h, 18h, etc. The results are summarized in the Table 1 below:

TABLE 1: SERIE-CORRELATION COEFFICIENT AS FUNCTION OF LAG

LAG (HRS)	6	12	18	24	30	36	42	48	54	60	90	120
CORR COEFF	0,85	,71	,58	,46	,35	,27	,21	,16	,13	,093	,026	,008

The 6 hourly H_{mo} values are highly correlated with a serie-correlation coefficient of 0,85. The serie-correlation coefficient gradually reduces with increasing lag and only becomes smaller than 0,1 with a lag of 60 hours. To ensure that independant values of H_{mo} are selected it will therefore be necessary to ensure that not more than one sample is taken in each 60 hour period. When studying the recorded H_{mo} values, difficulty was experienced in selecting an independent event for each 60 hour period. This is illustrated in Figure 5 where the wave heights recorded during the passage of a series of cold fronts at roughly 5 day intervals are shown.

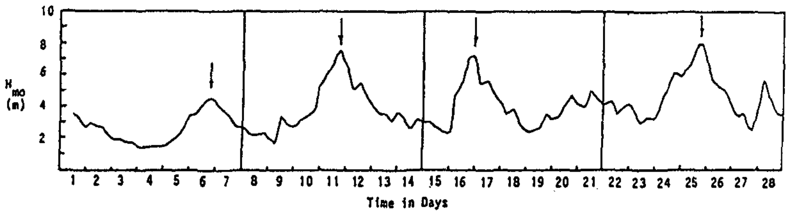


FIGURE 5: SELECTION OF MAXIMUM H_{mo} PER WEEK

It was therefore decided to rather base the selection of independent H_{mo} values on an event basis. The maximum H_{mo} recorded in each week of the recording period was therefore selected as a basis for the fitting of probability distributions and care was taken to ensure that the same event did not feature in successive weeks.

2.2 Identical distribution of the data

In an effort to obtain identically distributed samples, the months which shows similarity in recorded height were grouped together. In the Table 2 below the mean and standard deviation of the maximum weekly H_{mo} values are given.

TABLE 2: MEAN AND STANDARD DEVIATION OF MAX WEEKLY H_{mo} PER MONTH

	MAY	JUNE	JULY	AUG	SEPT	OCT	NOV	DEC	JAN	FEBR	MARCH	APRIL	YEAR
NUMBER OF WEEKS	31	28	29	30	31	28	29	31	32	29	30	31	359
MEAN	4,726	4,864	4,897	4,920	5,087	4,629	4,003	4,019	3,919	3,745	4,090	4,158	4,418
STANDARD DEVIATION	1,630	1,237	0,898	0,995	1,210	1,191	0,825	1,171	0,787	1,031	1,132	0,913	1,198

The mean values for the winter months May to September are fairly constant around 4,9 m whereas the mean for the summer months November to April are also nearly constant around 4,0 m. The standard deviations show no clear pattern with high values in the months where extreme storms occured (i e May) and low values in the months that were free of such storms (i e January). For the purposes of selecting identically distributed data, the stormy months May to September were grouped together. The 149 weekly maximum values of H_{mo} obtained during these months were therefore used to represent the best estimate of independent and identically distributed samples.

2.3 Influence of sampling method on the predicted design wave heights

The influence of the sampling method on the predicted design wave heights were studied by calculating the most probable value of the 10 and 100 year wave according to the Extreme 1 distribution using the method of moments to fit the data. The sampling method used varied from using all 6 hour records for all months (10 537 records) to only using the maximum H_{mo} recorded in each year (8 records). The results are summarized below:

TABLE 3: EXTREME 1 DISTRIBUTION - COMPARISON OF SAMPLING METHODS

3.1 DATA FROM ALL MONTHS

Sampling method	N	H_{10}	H_{100}
All data	10 357	9,80	11,63
Max H_{mo} per week	359	9,73	11,88
Max H_{mo} per month	96	9,52	11,68
Max H_{mo} per year	8	9,53	11,88

3.2 DATA FROM STORMY MONTHS

Sampling method	N	H_{10}	H_{100}
All data	4 463	9,92	11,91
Max H_{mo} per week	149	9,50	11,72
Max H_{mo} per month	40	9,29	11,46
Max H_{mo} per year	8	9,53	11,88

As can be seen from the above table, the predicted design waves are very insensitive to the method of sampling with the most probable 10 year H_{mo} varying from 9,3 m tot 9,9 m and the 100 year H_{mo} from 11,3 m to 11,9 m. Neither the dependence of the data when using all the data, nor the non-identical distribution of the data when the calmer summer months are included, seem to seriously influence the result.

According to Wallis (1988) correlation between data should not alter the expected values but will influence the uncertainty of the estimates as measured by the confidence limits or r m s errors. In the example above however it should be considered fortuitous that such similarity in results were obtained when varying the number of samples from 8 to 10 537.

The fact that inclusion of the data from the calmer summer months in the data set did not seriously influence the result is not surprising due to the small difference in wave height between the summer and winter and the fact that the winter storms will dominate in the total data set.

3. Model selection

The procedure that is most often followed for selecting an appropriate model for the long term distribution of wave height is to fit the data to a number of these models and to select the model which fits the data best according to some goodness of fit criteria. This procedure have led many engineers and researchers to develop a preference for a particular model. In an effort to obtain a better fit of the model to the data, the number of parameters used in the model are often also increased.

The method of model selection described above has been severely critisized by a number of researchers such as Wallis (1988) and Linhart and Zucchini (1986). Authors such as Wallis (1988) have found that if they generate data according to a particular distribution and then use the above procedure of model selection, they frequently select a different distribution than the one used for the generation of the data. He illustrates for instance that by using 4 of the most popular models for the long term distribution of flood intensity and generate data according to one of these distributions, the correct distribution will be chosen less than 50 per cent of the time if 100 data points are used.

An approach to improve model selection is suggested by Linhart and Zucchini (1986). They propose the use of bootstrap sampling [Efron (1982)] whereby the models are not only fitted to the original data set but also to data obtained from resampling the original data set. Here it is important to have independant and identically distributed samples. The process of model selection proposed by them are as follows:

(i) Select a number of likely models (Weibul, Extreme I, Log-normal, etc).

(ii) Select a goodness of fit criteria - say Kolmogoroff discrepancy.

(iii) Select a random sample of size n (with replacement) from the original observations to obtain a bootstrap sample.

(iv) Calculate the parameters of the models selected in (i) above using the method of maximum likelihood.

(v) Calculate the maximum discrepancy for each distribution.

(vi) Repeat steps (iii) to (v) a large number of times (say 100) and keep track of the discrepancies.

(vii) The model which gives the lowest average discrepancy over the 100 repetitions is chosen as the most appropriate.

This bootstrap sampling technique was applied to the 149 weekly maximum H_{mo} values recorded during the stormy months. Six probability distributions were considered i e the Gamma, Normal, Log-normal, Exponential, Weibul, and Extreme I distributions. The Kolmogoroff discrepancy was used as goodness of fit criteria and the method of maximum likelihood for parameter estimation. The results are summarized in the Table below showing the percentage of the time that the various models were selected, the average discrepency for each model and the most probable value of the 100 year wave (H_{100}) for each model

TABLE 4: BOOTSTRAP SELECTION OF MODELS

Model	% Selected	Ave discrepancy	H_{100} (m)
Log-normal	54	0,066	10,7
Gamma	28	0,071	9,9
Extreme I	12	0,074	12,3
Normal	6	0,096	8,9

The result of this bootstrap method clearly illustrates the dillemma we face in model selection. The goodness of fit criteria chosen indicate that the first three models listed above all fit the bulk of the data reasonably well. The upper tail of these models do however differ significantly which result in large differences when these distribution are extrapolated to obtain design values. Emphasis on the upper tail of the distribution can be incorporated in the goodness of fit criteria but a sound statistical criteria whereby the degree of emphasis could be decided, is not available.

4. Confidence limits

Two types of uncertainties pertaining to the prediction of the extreme values of significant wave height lend themselves to statistical analysis i e:

(i) If the long term distribution of wave height is perfectly known such as would be the case if we had an infinitely long and perfect wave record, there is still uncertainty about the largest wave that will occur within the next N years.

(ii) If we are sure about the model that describes the long term distribution of wave heights, but we have to estimate the parameters of this distribution from a limited wave record, we are uncertain about the values of these parameters.

Other uncertainties which cannot so readily be addressed in statistical terms include:

(iii) Decisions about the most appropriate model to describe the long term distribution of wave height.

(iv) Whether the data recorded over a number of years can be considered representative of the long term distribution of the wave height.

Case (i)

This case can best be illustrated by means of an example. Let us assume that we have an infinitely long wave record and the maximum H_{mo} recorded during each year of the record follows an Extreme I distribution as follows:

$$H_{mo}(p) = 8{,}09 - 0{,}85 \ln(-\ln p) \quad (1)$$

where $p = 1 - \frac{1}{T}$ with T the return period in years.

The most probable 100 year wave, i e the wave that will occur on average once in a 100 years, is then given for

$p = 1 - \frac{1}{100} = 0{,}99$ by

$H_{mo}(0{,}99) = H_{mo_{100\ years}} = 12{,}0$ m

If the infinitely long record is broken into 100 year intervals, the highest H_{mo} in each 100 years will obviously not be the same but follow a distribution as shown in Figure 6.

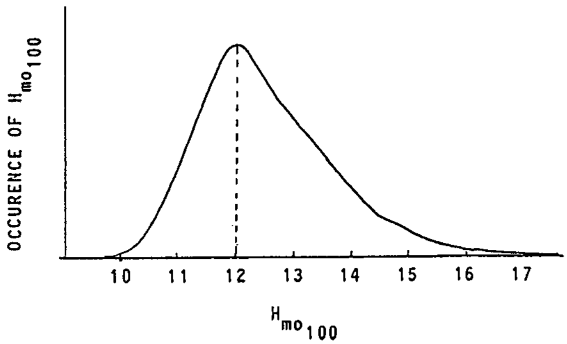


FIGURE 6: DISTRIBUTION OF $H_{mo_{100\ years}}$

The general expression relating the wave height with a given risk (r) of being exceeded during T years to probability (P) is given by

$$p_T^r = (1-r)^{1/T} \quad \text{.................................} \quad (2)$$

If a designer therefore is designing a structure with a design life of say 30 years and he is only willing to take a 10 per cent risk of that wave height being exceeded during the 30 years, he must in our example substitute

$p_T^r = (1 - 0,1)^{1/30} = 0,9965$ in equation 1 to obtain:

$$(H_{mo})_{30}^{0.1} = 8,09 - 0,85 \ln(- \ln ,9965) = 12,9 \text{ m}$$

The reason for labouring this well known concept even to a learned audience such as the attendees of an ICCE, is that the author has found great confusion amongst engineers and designers with respect to concepts such as a wave with a given recurrence interval and its confidence levels. The risk stated above has nothing to do with the confidence bands that are most often quoted. The confidence bands refer to the accuracy with which the wave with a given return period can be estimated if an infinitely long record is not available. With a wave record of only a few years duration there will be considerable uncertainty about the values of 12,0 m and 12,9 m obtained for $H_{100 \text{ years}}$ and $(H_{mo})_{30}^{0.1}$ in the example above. This uncertainty is reflected by the confidence bands.

At this point I would also like to make a plea that the concept of a wave with a given return period be scrapped in all specifications for design. I cannot see any sense in specifying that a structure should be designed for a "100 year wave" if the life of the structure will only be 30 years. Even more ridiculous is the specification that an operation that will last for one month should be designed for a "10 year wave". It makes a lot more sense to specify a wave height with a given risk of being exceeded within the 30 year life of the structure or within the one month of the operation. Let us therefore always specify H_h^r when h is the design life of the structure or the duration of an operation and r is the risk of exceeding H during the period h. In this way the designer and his client will know exactly where they stand and not be lulled into a false sense of security in the case where they have for example designed a structure with a 30 year life for the socalled "100 year wave", without realizing that they face a 26% risk of seeing the design wave within the life of the structure.

Case (ii)

The most commonly quoted confidence bands are those where a model for the long term distribution of wave height is assumed and the accuracy with which the parameters of the distribution can be estimated based on a limited record length, is assessed. The accuracy of the estimate of a wave height with a given probability of exceedance is then mainly a function of the record length. The confidence bands narrows with increasing record length and in the case of an infinitely long record length reduces to zero in which case the problem reduces to Case (i) above.

Confidence bands for a number of the more popular distributions have been established either theoretically or by Monte Carlo simulation. Zucchini and Adamson (1984) have also shown that bootstrap sampling techniques can be successfully employed to estimate confidence bands once an appropriate model for the long term distribution of the appropriate variable (waves in our case) have been selected. Efron (1987) further expanded on this method.

For the purpose of establishing confidence bands it is important to use independant and identically distributed samples. For the data used in this paper such samples could include the maximum H_{mo} recorded per year (8 samples), the maximum H_{mo} recorded in each stormy month (40 samples) or the maximum H_{mo} recorded in each week during the stormy months (149 samples). Assuming that these samples belong to an Extreme I distribution, the confidence with which the most probable value of the 100 year H_{mo}, or $(H_{mo})_{100}^{0.63}$, could be estimated was calculated using the bootstrap technique. The results are summarized in Table 5.

It is interesting to note that if the parent distribution is assumed known, and if maximum use is made of the 8 years of available data, the confidence bands for the most probable value of the 100 year H_{mo} becomes relatively narrow i e the value at 95% confidence is only 12% higher than the most probable value.

Cases (iii) and (iv)

In the general case where both the distribution and its parameters must be estimated from a limited data base, strict statistical treatment becomes impossible. The large number of distributions that have been proposed for the long term distribution of wave height and the different goodness of fit criteria that can be used in selecting the best fit model, makes it impossible to assess the certainty whereby a given model can be selected. This has been clearly illustrated earlier. If we add to these the uncertainties pertaining to the representativeness of the samples, the problem becomes even more complicated. Doubts about the representativeness of the sample stems from such factors as the loss of records during storms, long term variations in wave climate, extreme events occurring that belong to a different parent distribution and did not form part of the sample, etc. Most of these uncertainties can be of the same order of magnitude or larger than those associated with the accuracies with which the parameters of a distribution can be estimated for a few years of data.

TABLE 5: 95% CONFIDENCE BANDS AS FUNCTION OF SAMPLING METHOD

Sampling Method	Number of samples	Probability associated with (H_{mo})	$\frac{(H_{mo})_{100}^{0.63}\,[95\%]}{(\hat{H}_{mo})_{100}^{0.63}}$
Max per year	8	1 - 1/100 = 0,99	1,35
Max per month in stormy months	40	1 - 1/500 = 0,998	1,20
Max per week in stormy months	149	1 - 1/2166 = 0,9995	1,12

5. Summary and Conclusions

1. There is a tendency for waveriders to malfunction near the peak of severe storms. Care should be taken with every data set to ensure that major storms are not truncated or even totally ommitted.

2. The general practice of obtaining design wave heights by fitting distributions to data recorded at 3 hr to 6 hr intervals without regard for the independence of the records still seem to be the best approach to use, especially with short data sets. Correlation between the data will however influence the uncertainty of the estimates as measured by confidence limits or rms errors.

3. The selection of an appropriate model for the long term distribution of wave height still remains the most uncertain part in the process of design wave height determination. Bootstrap techniques can be of some help in this process.

4. It should be remembered that the confidence bands associated with extreme events that can be treated statistically only forms part of the overall uncertainty. In all cases the true confidence bands will be wider.

5. It is recommended that the concept of a wave with a given return period be replaced by a wave with a given risk (r) of being exceeded within the design life (h) of a structure (H_h^r). It should be realized that this wave height is only a most probable value and has an associated confidence interval which can be rather wide and is impossible to accurately assess in typical cases of h and r with the short history of wave recording.

6. Acknowledgements

The support of the Division of Earth, Marine and Atmospheric Sciences and Technology of the South African Council for Scientific and Industrial Research in supplying the data and helping with some of the analyses is gratefully acknowledged.

References

Efron B (1982): The jackknife, the bootstrap and other resampling plans. CBMS-NSF regional conference series in applied mathematics (Philadelphia).

Efron B (1987): Better bootstrap confidence intervals. Journal of the American Statistical Association. Vol 82, No 397. p 171 - 200.

Linhart, H and Zucchini, W (1986): Model selection, John Wiley, N.Y.

Rossouw J, Coetzee L W and Visser C J (1982): A South African Wave Climate Study, Proc 18th Coastal Engineering Conference, Cape Town, ASCE, Vol 1.

Wallis J R (1988): Catastrophes, computing, and containment: Living with our restless habitat. IBM Research Report. RC 13406 (60023).