CHAPTER 9

Comparisons of extreme wave height estimates

As shown in chapters 1-8, several methods for estimation of extreme wave height h_{max} are available. Irrespective of the length of the original data series, the final estimate of h^*_{max} should be treated as a random value. Each of the considered methods is based on specific assumptions, and therefore the estimates obtained with the help of these methods should by definition be somewhat different. It is thus very relevant to compare the basic features of the estimates. Let us make the comparison and summarise the main differences in all the methods.

Table 9.1 demonstrates corresponding $\bar{h}_{\max}$ estimates. These were obtained using a wave height data series, which was the same for all the methods. The wave heights in the Baltic Sea were simulated with a hydrodynamic model driven by the observed wind.

The time interval between wave height readings is six hours. Table 9.1 shows only the summary data for mean wave heights. We also analysed wave heights of other probabilities of exceedance, and our conclusions remained largely unchanged.

Table 9.1

Extreme values $\bar{h}_{\max}$ of mean wave heights at return periods of 50 and 100 years obtained with the use of different methods. The Baltic Sea

Method	$h^{(50)}$, (m)	$h^{(100)}$, (m)
IDM, $\hat{n}=1$, $h_{0.5}$=0.66 (m), s=1.8	6.7	7.3
IDM, $\hat{n}=4$, $h_{0.5}$=0.66 (m), s=1.8	5.7	6.2
IDM, (2.15), $\hat{n}=10$ (=0.2), $h_{0.5}$=0.66 (m), s=1.8	4.8	5.5
AMS (2.10), $\hat{n}=1$, a=1.73, b=3.96	6.4	6.8
AMS (2.10), $\hat{n}=4$, a=1.97, b=3.14	5.3	5.6
AMS (2.10), $\hat{n}=10$ (ρ=0.2 in (2.15)), a=2.14, b=2.65	4.6	5.0
AMS, sample estimated a=2.50, b=3.25 95% confidence interval (m):	5.0 4.5–5.4	5.2 4.6–5.6
POT, Z=2.5 (m), λ=6.0, (213 storms) 95% interval for return period T (year)	4.7 47–53	4.9 95–106
POT, Z=3.0 (m), λ=2.4, (85 storms) 95% interval for return period T (year)	4.4 44–58	4.6 88–116
POT, Z=3.4 (m), λ=1.0, (35 storms) 95% interval for return period T (year)	4.3 37–63	4.4 73–125
POT, Z=3.6 (m), λ=0.4, (15 storms) 95% interval for return period T (year)	4.2 36–84	4.3 71–167
MENU	6.2	6.9
BOLIVAR, 1st maximum	5.0	5.2
BOLIVAR, 2nd maximum	4.0	4.3
BOLIVAR, 3rd maximum	3.8	4.0

Note: $\hat{n}$=1 corresponds to wave height data recorded at every observation time, i.e. with 6 hour intervals; $\hat{n}$=4 corresponds to data extracted once in four observation times, i.e. once every 24 hours; and $\hat{n}$=10 refers to data taken once every ten observation times, which, in accordance with relation (2.15), is equivalent to using non-correlated (or independent) observations.

The true value of h_{max} is, *a priori*, unknown. It must be located in some range with bounds (h_1,h_2), the width of which depends on the initial assumptions of the methods in use. Therefore, a single value (i.e one point) estimate of h_{max} does not say much about the advantages and shortcomings of the methods. Wave heights also exhibit inter-annual, seasonal and synoptic variability as described by equation (8.5). Therefore, for all of the above methods, the estimated confidence interval of parametric or non-parametric quantiles h_p^* does not correspond to true variability of h_{max}. If the time series length is increased infinitely, all the methods predict a zero confidence limit range, while in real conditions, mostly due to existence of natural variability of different kinds and scales, there is a lower finite limit of this range. These considerations are relevant for the analysis of data in Table 9.1. To compare the results we need to have an estimate of the true value of h_{max}. Because the AMS method has the strongest theoretical foundation and reflects the existence of the inter-annual variability, let us assume that the estimate obtained with the help of the AMS method (2.17) is the most truthful. Thus in Table 9.1 we provide various range estimates obtained with the help of the AMS method and compare them with single value estimates that are obtained using the other methods.

The AMS method

This method is based on processing of the last elements of the data series. The theoretical foundation of this method is the most elaborated, and from the outset the method was designed for prediction of extreme values. Parameters *a* and *b* in relation (2.5) are evaluated either using original data (more specifically, annual maxima) or relations (2.10) corrected with respect to the correlation range (because the number *n* of readings in the sample enters formula (2.10)).

Processing of observations for the Baltic Sea yields a hundred year wave height h_{max} of 5.2 m. If internal correlation is not taken into account, relation (2.10) yields a hundred year wave height $\bar{h}_{\max}$ of 6.8 m. If four consecutive observations are considered correlated (i.e. correlation range is equal to a day), h_{max} =5.6 m. For the correlation range of two and a half days ($\hat{n}$ =10) $\bar{h}_{\max}$ =5.0 m. This means that taking into account the correlation between neighbouring observations leads to smaller estimates of $\bar{h}_{\max}$. In normal practice the AMS estimates are made with parameter values determined using the annual maxima from the sample.

The IDM method

The initial distribution contains the whole range of wave heights at all observation times. Therefore, it experiences all possible wave generating conditions. Situations with extreme waves constitute only a small part of this variety. All IDM extreme wave height estimates should therefore be interpreted in terms of synoptic observations, as follows, “Once at a single synoptic observation time during *n* years the wave height *h* can be observed”. The probability of such extreme wave height depends on the total number of synoptic observation times. This means that the IDM method does not produce a distribution of extreme wave heights but determines the quantile, to which the maximum wave corresponds. Another problem is connected to the need for making extrapolations. Usually, an extrapolation is justified up to the probability level of 0.1 %. However, if the original data series is one hundred years long, it can be made up to 0.01%.

For all IDM results in Table 9.1 the parameters $h_{0.5}$ and s of the log-normal distribution are unchanged. Correspondingly, the differences in the estimates for a 50-year and 100-year return wave height resemble the differences in the corresponding probabilities. Using an equivalent independent number of wave records (i.e. $\hat{n} = 10$) instead of correlated wave records (i.e. for $\hat{n} = 1{,}4$) leads to smaller estimates of extreme wave heights, which are closer to the estimates obtained with the help of the AMS method.

Thus, the IDM method, which, in fact, is one of the earliest methods in wave statistics, and which was not intended for use in estimation of extreme wave heights, can, nevertheless, lead to reasonable estimates. In order to successfully use this method it is very important to specify the correct probability of extreme wave heights. It should, however, be remembered that the distribution of extreme wave heights is dependent on the initial distribution. For example, if the initial distribution is of exponential type, then the distribution of extreme (rare) values asymptotically tends to the first limiting distribution (2.5). There are many studies, in which log-normal or Weibull distributions represent the general distribution of wave heights. Both log-normal and Weibull distributions are exponential. In recent years more and more investigators have preferred the log-normal distribution, feeling that it better represents observed and simulated data for mid-latitudes and subtropics where mixed waves dominate.

The POT method

This method is the most popular at present. Extreme waves are observed during storms which alternate with weather windows (See Fig. I.1). In the POT method the sequence of storms is practically treated as a pulse-like random process. The method selects only the highest wave in a storm. This means that it is aimed at estimation of the extreme values.

At the same time, the lack of asymptotic relations in the POT method does not allow a theoretical derivation of quantile h_p. It is dependent on the approximations assumed in (5.1). Furthermore, the method supposes independence of consecutive storms and uses the Poisson distribution for the storm number. This leads to some uncertainty in estimated return period (see Fig. 5.3).

Attempts to consider the consecutive storms correlated and, on the basis of this, to introduce corresponding corrections in the Poisson distribution, do not change the results significantly.
The most “influential” parameter in the POT method is the threshold value of wave height, which discriminates between normal variations of the random process and a storm of interest. For example (see Table 9.1), if the selected threshold changes from 2.5 to 3.6 m, the estimated height of a hundred year wave decreases by 60 cm. Equation (5.10) and fig. 5.2 are instrumental in choosing the optimal threshold value of wave height separating storms that should be selected for further analysis. If the POT method is applied to observed series of wave heights, then any change of the threshold value results in recalculation of the combined distribution parameters using the original time series. Thus, in this case, the choice of the threshold affects the final results less strongly than it does when equation (5.10) is used.

It is noteworthy that studies conducted in other regions of the World Oceans also exhibit strong changes of the POT method output wave height in response to the threshold variations. Bjerke et al. (1990) used a nine-year time series of observations conducted every three hours in the coastal waters of Norway and obtained the following estimates of a hundred year wave height: 16 m for threshold of 3 m and to 14.5 m for threshold of 9 m. Estimates of a hundred year significant wave height along the Atlantic coast of Spain by Rossouw et al. (1995) changed from 13.4 m to 11.7 m due to the threshold variations.

The dependence of the POT method results on the choice of the threshold wave height has been studied in many papers and is well known. Several criteria are proposed for the storm selection [Szabo et al., 1989]. The general rule is the higher the threshold, the smaller the estimated extreme wave height hmax. When the threshold values exceed a certain limit, which is sufficiently high, the POT method extreme wave heights tend to a certain stable value. When the threshold decreases, the POT method estimates approach the IDM method estimates.

A serious shortcoming of the POT method is connected to the uncertainty in the estimates of the return period. This is clearly seen in Table 9.1, and is illustrated in Fig. 5.3. For higher values of the threshold the number of selected storms in the sample may become rather small, therefore the estimates of λ become less accurate, resulting in a broader confidence range for the return period. For example, for the threshold of 2.5 m, 213 storms were selected, and the confidence ranges for 50-year and 100-year wave heights are as small as 6 and 11 years, respectively.

When the threshold progressively increases from 2.5 m to 3.6 m, the number of selected storms decreases from 213 to 15, and the uncertainty in the return period of 50-year and 100-year wave height estimates increases to 48 and 96 years, respectively.

Summarizing, the POT method results in somewhat smaller estimates of extreme wave height in comparison with the AMS method. The higher the threshold, the smaller the final estimate.

The MENU method

This method represents the wave series as a random process. A return period of the extreme wave height h is considered as the expected time of the first up-crossing of this level. As a result, MENU extreme wave height estimates do not differ greatly from the corresponding estimates obtained from the IDM method.

The BOLIVAR method

The method does not assume that consecutive storms are not correlated. It takes into account not only all storms in which wave heights exceeded a certain threshold, but includes data for the strongest storm in each year in the time series.

This means that at least one record for any year is included in the analysis. This procedure makes it possible to utilize asymptotic distributions for maximum wave heights.

The use of the multi-dimensional distribution function (7.3) also makes it possible to extend the analysis from the first to other consecutive maxima that can be recorded at different return periods. It is possible that the second maximum of wave height at return period of 100 years is larger than the estimated 50-year wave height. The ability to produce such estimates is an advantage of the BOLIVAR method.

---oooOooo---