CHAPTER 1

Initial Distribution Method (IDM)

This method estimates the extreme wave height h_{max} of a specified return period as the quantile h_p of the wave, height distribution *F(h)* with probability *p*. Provided the distribution of individual wave heights during the quasi-stationary interval obeys equation (I.1), then

$$h_p = \bar{h}\sqrt{-\frac{4}{\pi}ln(1-p)} \qquad (1.1)$$

For p = 0.001 we obtain h_p = 2.97$\bar{h}$. Thus, the height of wave, which is the highest in a thousand waves, is expected to be almost three times larger than the mean wave height $\bar{h}$. Using approximation (I.5) of long-term wave height distribution, the quantile with probability p can be computed as follows:

$$h_p = h_{0.5}\, exp\left(\frac{U_p}{s}\right) \qquad (1.2)$$

where U_p is the quantile of the standard normal distribution N(0,1). Here quantile h_p should be understood as the wave height which is likely to be observed once (at the standard synoptic observation times) in *T* years.

In applied studies, the period *T* is called “return period”, and the corresponding probability is defined as

$$p = \frac{\Delta t}{24 \cdot 365 \cdot T}$$

where Δt is the interval (in hours) between subsequent observations (say, 6 hours). Then we get p = 0.000684/T. If Δt = 3 *h*, we get p = 0.000342/T.

Table 1.1 shows values of h_p corresponding to different choices of Δt and *T* for $h_{0.5}$ = 1.0 and s = 2.0 in distribution (I.5). It can be seen that the wave height estimates do depend on the recording interval.

Table 1.1
Estimated 1-year and 100-year return period wave heights at synoptic observation times.
Computations for log-normal distribution with $h_{0.5}$ =1.0 m and s = 2.0 and
with three record intervals of 3, 6, and 12 hours

Number of observations per day	*Δt*, hour	$h_{max}/h_{0.5}$	
		***T* = 1 year**	***T* = 100 years**
8	3	5.5	9.4
4	6	5.0	8.8
2	12	4.5	8.1

There are two traditional data sources for the determination of the wave height distribution *F(h)*. Historically, visual observations onboard ships and at the Ocean Weather Stations were the basis for such calculations.

At present the data are mostly provided by instrumental wave observations from automated buoys and through numerical simulations of waves. Regardless of the source of information, the initial distribution method leads to some ambiguity in the estimates of the h_{max} using quantiles (1.1,1.2).

If relation (1.2) is used in (1.1) as an estimate for $\bar{h}$, then the distribution *F(h)* of all individual wave heights *h* during T-year long interval can be represented as the combined distribution

$$F(h) = \int_0^\infty G(h,\bar{h}) f(\bar{h}) d\bar{h} \qquad (1.3)$$

with $G(h,\bar{h})$ being wave height distribution (I.1) over the quasi-stationary interval and $f(\bar{h})$ being the long-term (climatic) probability distribution of mean wave height (I.5).

The average number of individual waves *N* in such a population depends on the mean wave period $\bar{\tau}$. To estimate it, we can assume that the joint distribution of wave heights h and periods τ is governed by multiplication of the individual log-normal distributions

$$f(h,\tau) = f(h) f(\tau \mid h) \qquad (1.4)$$

with parameters $h_{0.5}$ and s for the marginal distribution *f(h)* of wave heights and with parameters $\tau_{0.5}(h)$ and $s_\tau(h)$ for the conditional distribution of wave periods f(τ|h) for a given wave height *h*. Then the probability that $h_k \le h < h_{k+1}$ and $\tau_s \le \tau < \tau_{s+1}$ is equal to

$$p_{ks} = \int_{h_k}^{h_{k+1}} \int_{\tau_s \tau_{s+1}} f(h,\tau) dh d\tau \ ,$$

and the random number of such waves $m_{ks}=p_{ks}N$ is characterized by following the binomial distribution (see Lopatoukhin, Lavrenov et al., 1999]:

$$P_N(m) = C_N^m p^m q^{N-m}, \quad q = 1 - p \qquad (1.5)$$

Here $m = pN$ is the mean number and $\sqrt{Npq}$ is the corresponding r.m.s. deviation.

For small values of probability p, distribution (1.5) tends asymptotically to the Poisson distribution:

$$P(m) = exp\ [-\lambda] \frac{\lambda^m}{m\,!} \qquad (1.6)$$

which has mean value λ, and r.m.s. deviation $\sqrt{\lambda}$.

For example, the median $h_{0.5}$ of mean wave height climatic distribution for the southern part of the Baltic Sea is equal to 0.66 (m), and the shape parameter *s* is equal to 1.81. For the wave periods we have $\tau_{0.5}=3.7(s)$ and s_τ =3.4. Hence the mean number of waves for a year is *N* = 8 400 000. The probability p_{ks} that $1 \le h < 2\ (m)$ and $4 \le \tau < 6(s)$ is 0.11. Thus, using (1.5) we have $m = p_{ks}N \approx$ 924000 waves per year and σ_m = 910. For the probability p_{ks} that $3 \le h < 6(m)$ and $6 \le \tau < 8(s)$, which is equal to 0.0000196, relation (1.6) predicts *m* = 165 and σ_m = 13 (waves).

The initial distribution method results are sensitive to the variation of parameters in extrapolation formulae (1.1) and (1.2). It is particularly sensitive to variations of parameter *s* for small values of probability *p*. Quite often the statistical distribution of wave height in an observed sample does not match closely relations (I.1) and (I.5). This results in significant differences in estimates, which are obtained with the help of various methods. For example, if values of *h* and *s* computed using the formula for statistical moments are, respectively, $h_{0.5}$=0.66 (m) and *s*=1.81, then the estimate of one-hundred year return wave height in the Baltic Sea is h_{max} = 7.3 m. If the median is the same but the parameter *s* is determined using the formula for quantile, which gives $s_{(q)}^*$=1.95, then h_{max}=6.1 m.

The "true" distribution function F(h) is unknown. We use the observation series to obtain an approximation F*(h), the reliability of which depends on the length of the series and the quality of observations. Usually, the time series are long enough. If, for example, the series is 30-40 years long, and if observations are taken 4-8 times a day, the total number of records is 50–100 thousand. Correspondingly, the method can provide quite a narrow confidence range (provided the approximated distribution is really close to the true one).

For the above example the estimates h0.5=0.66 (m) and s=1.81 were obtained using a simulated data series, which was 35 years long. The time step was 6 hours. The total number of readings was N = 51100. Thus, the standard deviation of the estimate of s is τs=0.014. For $\sigma_{h0.5}$ the expected deviation is as small as 0.002. Correspondingly, the 90% confidence range for the one hundred year wave height (hp=7.3 m), which is obtained by extrapolation of distribution (I.4) up to probability p=0.684·10–5, is very narrow, namely 7.2-7.4 m.

It is obvious that the initial distribution method cannot, in principle, represent the true variability of maximum wave heights. Even if we consider approximation (I.1), and, particularly, (I.5) ideal, the parameters $\bar{h}, h_{0.5}, s$ that enter these relations still remain random as they are affected by synoptic, seasonal and extra-annual variability. As a result, one point (i.e. single-value) estimates of extreme wave height have considerable inherent uncertainty. Long-period variability also justifies the use of wider confidence limits for interval estimates.

Hence, marked sensitivity of the initial distribution method to input data quality combined with considerable uncertainties of estimates at small probabilities, as well as adoption of some assumptions regarding the possibility of combining approximated distributions (I.1) and (I.5), suggest that there is a need to further develop the method.

---oooOooo---