CHAPTER 2

Method of annual maximum series (AMS)

This method defines h_{max} as the last term of the ranked independent series of wave heights h. Thus it is a random value with the distribution

$$F_m(x) = [F_h(x)]^n \tag{2.1}$$

which depends on the type and parameters of the original wave height distribution $F_h(x)$ and on the number of data *n.* For large values of *n*, the exact distribution (2.1) of independent, similarly distributed random values tends to one of the three following asymptotic distributions:

$$F(x) = exp(-exp(-x)) \tag{2.2}$$

$$F(x) = \begin{cases} exp(-x^{-\gamma}), & x \ge 0, \gamma > 0 \\ 0, & x < 0 \end{cases} \tag{2,3}$$

$$F(x) = \begin{cases} exp(-(-x)^{\gamma}), & x \le 0, \gamma > 0 \\ 1, & x > 0 \end{cases} \tag{2.4}$$

It is possible to apply a nonlinear transformation to variable *x* in distributions (2.2) – (2.4). Then, depending on the value of γ the distributions will converge to a single Generalized Extreme Value (GEV) distribution (see Leadbetter et al., 1986). Such a transformation is convenient if the type of original distribution (2.1) is not known. Then the selection of limiting distribution has to be made using the series data (see Pilon, Harvey, 1993; Hoybye, Laszlo, 1997).

If the original distribution $F_h(x)$ is of exponential type (such as normal, log-normal, and Weibull types), the distribution (2.1) converges to double-exponential distribution (2.2), known as first-limit or Gumbel distribution, which most often reads as follows:

$$F(x) = exp(-exp[-a(x-b)]) \tag{2.5}$$

Parameters *a* and *b* depend on *n* and $F_h(x)$ as follows

$$F(b) = 1 - 1/n, \qquad a = nf(b) \qquad 2.6$$

where *f(b)* is distribution density $F_h(x)$ at the point b.

For a quasi-stationary interval, and if the Rayleigh distribution (I.1) is satisfied, the following relation holds [Lopatukhin et al., 1991]:

$$a = \frac{\sqrt{\pi \ln n}}{\bar{h}}, \; b = 2\bar{h}\frac{\sqrt{\ln n}}{\pi} \tag{2.7}$$

If n =1000, the median $\tilde{h}_{0.5}$ of distribution (2.5) is equal to

$$\tilde{h}_{0.5} = 3.05\bar{h} \tag{2.8}$$

and main quantiles are as follows:

$$\tilde{h}_{0.05} = 2.73\bar{h}, \; \tilde{h}_{0.95} = 3.61\bar{h}, \; \tilde{h}_{0.99} = 3.95\bar{h} \tag{2.9}$$

This shows that the estimate of h_{max}, which can be obtained using the AMS (formula (2.8)), is biased relative to the h_{max} estimate by the Rayleigh distribution *($h_{1/1000}$ = 2.97$\bar{h}$).* As a random value, the estimate h_{max} should be located within the limits (2.73$\bar{h}$; 3.61$\bar{h}$) in 90% of all cases (for n=1000). Once in a hundred cases the value of h_{max} can exceed 3.95$\bar{h}$.

For the log-normal distribution (I.5) with parameters $h_{0.5}$ and *s,* the parameters in distribution (2.5) read as follows [Hirtzel, 1980, Lopatukhin et al.,1991]:

$$\begin{aligned} a &= \left(\frac{sz}{h_{0.5}}\right) exp\left(-\frac{d}{s}\right), \; b = h_{0.5}\, exp\left(\frac{d}{s}\right), \\ d &= z - (0.918 + \ln z)/z, \; z = \sqrt{2\ln n} \end{aligned} \tag{2.10}$$

The corresponding quantiles are

$$\tilde{h}_p = b - \frac{1}{a}\ln(-\ln p) \tag{2.11}$$

For example:

$$\tilde{h}_{0.5} = b + \frac{0.367}{a}, \tilde{h}_{0.05} = b - \frac{1.097}{a}, \tilde{h}_{0.95} = b + \frac{2.970}{a} \tag{2.12}$$

Let us consider a wave height series in the Baltic Sea that contains four observations per day. The distribution parameters are: $h_{0.5}$=0.66(m) and *s* =1.81. The median $\tilde{h}_{0.5}$ of the annual maximum distribution (2.5) with parameters *a*=1.76(m) and *b*=3.29(m) is 4.1(m) The initial distribution method would result in a one-year return period wave height of 3.9(m). Thus the h_{max} estimate obtained from distribution (2.5) with the annual maximum series method is always shifted relative to the estimate (1.2) in the initial distribution method.

The value h_{max} is random, and in a sample with $n = 365 \cdot 24/\Delta t$ records in 90% of cases it will take a value between 3.3 - 5.6(m), see (2.12). This means that if observations are taken every 6 hours, and if $h_{0.5}$ and s correspond to the basic climatic distribution, the series of annual maximum wave height h_{max} (recorded once at synoptic observation times) will vary within these limits.

Evaluation of wave heights $h_{\max}^{(T)}$ at *T* year return period, is made using extrapolation of the distribution (2.5). The following formula [Leadbetter et al., 1986] is used:

$$P\left\{h_{max}^{(T)} \le h\right\} = P^T\left\{h_{max} \le h\right\} = exp\left[-exp\left[-a\left(h-b-a^{-1}\ln T\right)\right]\right] \qquad (2.13)$$

Distribution (2.13) is of the same type as (2.5) provided

$$a_T = a,\; b_T = b + a^{-1} \ln T \qquad (2.14)$$

Using (2.13) for the above case one can get the median for a hundred year return wave $\tilde{h}_{0.5}^{(100)}$ = 6.8 m. If the initial distribution method was used, the estimate would be 7.3 m.

For the AMS method, the extrapolation to long return periods can be justified only for periods (number of years) T^*, which do not exceed 3-4 lengths of the original data series used for evaluation of the distribution parameters. The estimates of wave heights with return periods of *T* years, for $T>T^*$, can be understood as bounds of corresponding probability interval for distribution (2.13). For example, the 10^4 year return wave height will be equal to the upper 1% percentile of the probability range, which is computed using an ensemble of one hundred series, each of them being one hundred years long [Lopatoukhin, Lavrenov et al., 1999].

Expressions (2.1), (2.7), (2.10), (2.13) are derived assuming independence of random values in the series. However, wave heights do correlate, both within the quasi-stationary interval, and for series made of synoptic time observations. Hence estimates of h_{max} in the Annual Maximum Series method should be corrected accordingly. The first way of correcting them would be to turn to the equivalent number of “conventionally independent” observations:

$$\hat{n} = \frac{\ln \rho}{\alpha} \qquad (2.15)$$

ρ being the threshold correlation coefficient. If the correlation is less than this threshold value, the correlation [see expression (I.3)] can be considered negligible. Parameter α is the decrement in the correlation function (I.3).

This approach can be used for first guess estimates only because parameter α is mostly chosen arbitrarily. Besides, the use of (2.15) is not fully theoretically justified. The estimation of parameters *a* and *b* in expression (2.5) from a sample of annual maximum wave heights h_{max} is used as a rule.

Fig. 2.1 shows a q-q (quantile – quantile) bi-plot of the empirical distribution function *F*(x),* which was computed using 35 annual wave height maxima generated by a model [Boukhanovsky A.V., Lopatoukhin L.J., Rozhkov V.A. 1998]. Empirical values (calculated from a sample) are shown along the x-axis, and theoretical values are shown along the y-axis. Such bi-plots will be used often in this paper. The statistical background for such presentations can be found elsewhere [Denby et al., 1983]. A thin cloud of points along the diagonal suggests that double exponential distribution (2.5) is a good approximation.

Estimates of parameters a* and b* can be made using methods of moments, quantile, maximum likelihood [Boukhanovsky, Davidan et al., 1996]. Also it is possible to use the method of L-moments, which was developed in [Hosking et al., 1985; Hosking, 1988; 1989; 1990]. Each of these methods uses different inputs for computation of a* and b*. Correspondingly, statistical properties of values *a*, b*,* (i.e. means and variances) will also be different. For example, if we use the maximum likelihood method for values A=b, $B=a^{-1}$ that would transform (2.5) into

$$F(h) = exp\left[-exp\left[-\frac{h-A}{B}\right]\right], \qquad (2.16)$$

then we will get the following statistical characteristics of parameters:

$$\sigma_{A^*} = \sqrt{1+\frac{6}{\pi^2}(1-\gamma)^2}\,\frac{B^*}{\sqrt{n}},$$

$$\sigma_{B^*} = \frac{\sqrt{6}}{\pi\sqrt{n}}B^*,$$

$$\rho_{A^*B^*} = 1/\sqrt{1+\frac{\pi^2}{6(1-\gamma)^2}}$$

while the β% uncertainty limit for quantile h_p^* will look as follows:

$$\left|h_p - h_p^*\right| \le U_\beta \sqrt{D_h} \qquad (2.17)$$

Here $\sigma_{A^*}, \sigma_{B^*}, \rho_{A^*B^*}$ are r.m.s. deviations and coefficient of correlation of estimates *A*, B*,* γ = 0.577126, *g=ln(–ln(p)),* U_β is β% quantile of the normal distribution,

$$D_h = \sigma^2_{A^*} + \sigma^2_{B^*} g^2 - 2\rho_{A^*B^*}\sigma_{A^*}\sigma_{B^*} g$$

Thus, an Annual Maximum Method estimate h_{max} = 4.8 (m) of 35 year return wave height in the Baltic Sea is covered by the 95% - confidence limits interval from 4.4 to 5.2 (m). A hundred year return wave height, according to (2.13) is $h^{(100)}_{\max}$ =5.1 (m), and it has a 95% - confidence limit range of 4.6-5.6 (m). The AMS methods uses data series made of all annual wave height maxima.

Such series cannot be too long because we very rarely have observations lasting more than 30 – 40 years. This is why the confidence limits range for AMS h_{max} estimates is sufficiently broad.

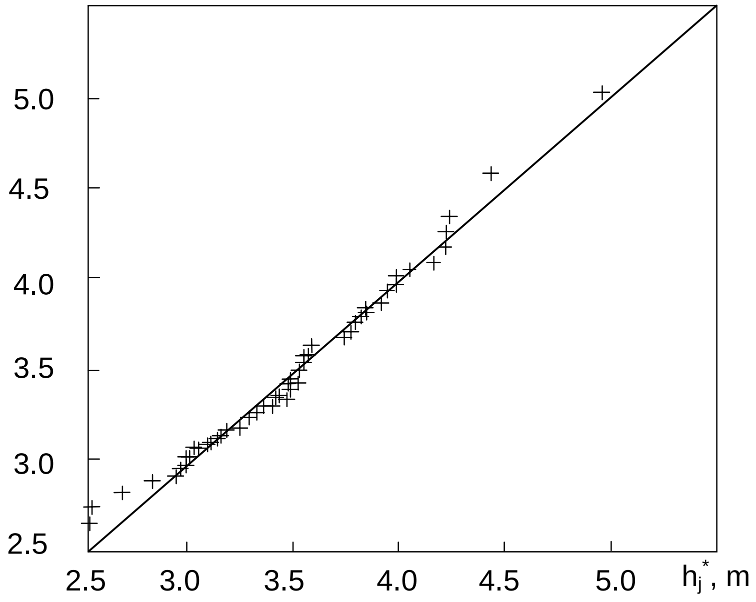

Figure 2.1. Biplot of 35 annual maxima of wave height h_j^* and corresponding quantiles h_j of distribution (2.5). The Baltic Sea

---oooOooo---

hj*, m