CHAPTER 5

Peak Over Threshold Method (POT)

The Initial Distribution or MENU methods require rather long data series for estimation of h_{max} . If the number of years is denoted by *n*, and number of observations per day is denoted by *m* then the total length of the series will be *N*. For example, for *n = 30, m= 4,* $N = 30 \cdot 4 \cdot 365 \approx 44000$. The Annual Maximum Method (AMS) excludes from the analysis those storms that are not the strongest in a particular year but could be the strongest in other years. This is one reason why the Peak-Over-Threshold method is used in estimating the highest wind waves [Muir L.R., El-Shaarovi, 1986]. The method is based on studying the sample of h_{max} for the *k* strongest storms observed during *n* years. The selection of the *k* strongest storms requires the biggest effort in POT and is its most subjective procedure. As a rule, the first step is the selection of a large number of storms, for which h_{max} is determined. In the following step only the strongest storms are considered. Usually, 20-30 storms are taken for a 30-40 year long interval.

It is assumed in the POT method that there is no dependence between wave heights in consecutive storms (i.e. wave heights in different storms are not correlated). Then the distribution function for maximum wave heights can be written as follows:

$$F(h) = \sum_{k=0}^{\infty} [G(h)]^k \, p_k \qquad (5.1)$$

where G(h) is the distribution for wave heights exceeding a predefined threshold *Z* during a year and p_k is distribution within a year of the number of storms during which maximum wave height exceeded *Z*. The Poisson distribution (1.6) is always used for p_k for sufficiently large values of *Z.*

In a particular case, when one can clearly see that storminess during the period of observations underwent considerable year-to-year variability, the geometrical distribution is a good approximation for p_k:

$$p_k = (1-\lambda)\lambda^k, \quad \lambda = 1 - p_0 \qquad (5.2)$$

where p_0 is the probability of occurrence of years when waves are below threshold *Z*.

If the case *(h ≥ Z)* is not rare (i.e. for small values of *Z*), one can use for p_k the truncated normal distribution:

$$p_k = \frac{1}{c\sqrt{2\pi\sigma}} exp\left[-\frac{(k-\lambda)^2}{2\sigma^2}\right], \qquad c = \frac{1}{2}\left[erf\left(\frac{\lambda\sqrt{2}}{2\sigma}\right)+1\right] \qquad (5.3)$$

with parameters λ, σ, c.

Table 5.1 gives, as an example, statistical characteristics of number of strong storms in various seas. They were derived with values of *Z* that were equal to two-three times the annual mean wave height.

Table 5.1.

Probability of occurrence p_k (%), mean values of $\bar{k}$ and r.m.s deviation σ_k of the number of strong storms k in a year

Sea, number of years	pk, %						$\bar{k}$	σ_k
	k=0	*k*=1	*k*=2	*k*=3	*k*=4	*k*=5		
Baltic Sea, 35 years	37	34	17	6	5	1	1.09	1.16
Barents Sea, 41 years	37	27	24	10	2	–	1.15	1.12
Black Sea, 35 years	23	26	31	9	6	5	1.66	1.42
Caspian sea, 39 years	46	21	28	5	–	–	0.92	1.00

The table reveals marked inter-annual variability of strong storm numbers. For example, in 46% of years there were no strong storms in the Caspian Sea. The two last columns in the table contain values of $\bar{k}$ and σ_k that are close to each other. This means that the Poisson distribution is a good approximation for this case.

Summing up the infinite series (5.1) for the distribution of storm numbers (1.6), (5.2), and (5.3) will lead, respectively, to

$$F(h) = exp(-\lambda(1-G(h))) \qquad (5.4)$$

$$F(h) = \frac{1-\lambda}{1-\lambda G(h)} \qquad (5.5)$$

$$F(h) \approx \frac{1}{2c} exp(-\lambda(1-G(h))) \cdot \left[1 - erf\left[\frac{\sqrt{2}}{2}\left(\sigma[1-G(h)] + \frac{\lambda}{\sigma}\right)\right]\right] \quad (5.6)$$

For *G(h)* double exponential see (2.2) or Weibull, see (2.3), distributions are used the most often, as follows:

$$G(h) = exp\left[-exp\left(\frac{h-A}{B}\right)\right] \quad (5.7)$$

$$G(h) = 1 - exp\left[-\left[\frac{h-C}{A}\right]^B\right] \quad (5.8)$$

where *A, B and C* are parameters.

Substituting (5.7) or (5.8) into (5.4)–(5.6) one can obtain at least six types of combined distributions. The Poisson – Gumbel distribution shown in (5.9) below is the most widely used.

$$F(h) = exp\left\{-\lambda\left\{1 - exp\left(-exp\left[\frac{h-A}{B}\right]\right)\right\}\right\} \quad (5.9)$$

It follows from (5.3)–(5.5) that transformation of *G(h)* into *F(h)* is relatively straightforward. At the same time, the actual form of p_k and parameters *(λ,σ)* affect the value of quantile h_p. There is some freedom in the interpretation of distribution *G(h)* as probability *P{h≥Z}.*

One can include in the analysis all wave height observations exceeding *Z* during a single storm (i.e. when there are several cases where $h \geq Z$ during a single storm) or, to represent the storm in the sample by its single maximum wave height *h(t).*

The difference between corresponding empirical distributions *(h > Z)* and *(h+ > Z)* is shown in Fig. 5.1 (q-q bi-plots of (5.7)).

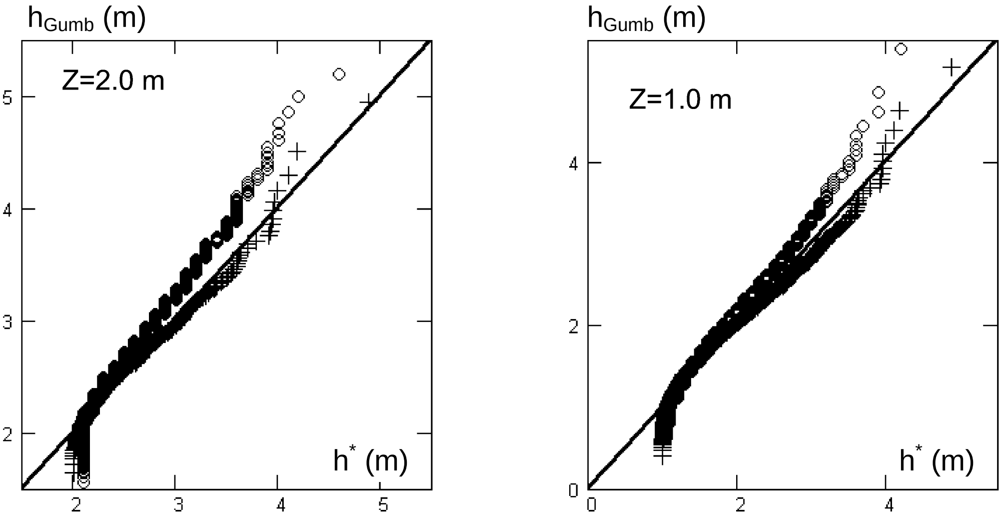

Figure 5.1 Biplots of Gumbel distribution and observed distributions *G(h)* in storms with maximum wave heights exceeding *Z = 2.0 m* and *Z = 1.0 m*. The Baltic Sea.
Straight line shows the theoretical expression. Sign *o* shows all values exceeding a predefined threshold. Sign + shows the single highest waves in a storm (i.e. one wave for a storm).

It can be seen that even for sufficiently small values of *Z (Z = 1(m))* both distributions are approximated fairly well by the Gumbel distribution (except for the utmost “left tail” of data). In the case when Function *G(h)* is determined using wave heights at all synoptic observation times (i.e. more than once in a storm), the “left tail” is given more weight and, hence estimates of long return period waves are lower in comparison with *G(h)* computed using the wave heights h+, which were counted only once and therefore coincided with maxima in the storms.

In computations of wave heights, the sample should include only the highest waves in all selected storms. The wave height h_{max} with return period of *T* years is taken to be equal to the quantile h_p, *p=(1–1/T)%* of distribution *F(h).*

According to (5.9) this distribution depends on the mean number λ of storms in a year, which, in turn, depends on the threshold value *Z*. Thus, h_p (including h_{max} as its particular case) is a function of *Z*. It follows from (5.9) that:

$$h_{max}(Z) = (A-B)ln\left(ln\left(1 + \frac{1}{\lambda T}\left[exp\left(-exp\left(-\frac{Z-A}{B}\right) - 1 - \frac{1}{N}\right)\right]^{-1}\right)\right) \quad (5.10)$$

where *N* is number of storms in *T* years.

The results of computations using relation (5.10) are shown in Fig. 5.2.

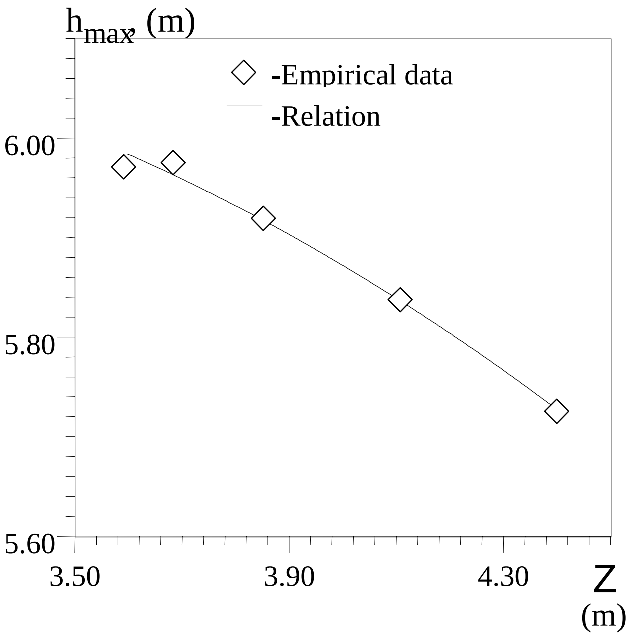

Figure 5.2. h_{max} (at 100 year return period) versus Z for the Gumbel distribution

It follows from (5.10) and Fig. 5.2 that quantile h_p in the POT method decreases as Z increases.

Further, it follows from Table 5.1 that function p_k in relation (5.1) can be computed using the Poisson distribution (1.6). Then it is possible to derive the following relation between T and *F(h):*

$$T = \frac{1}{\lambda F(h)} \tag{5.11}$$

The uncertainty range for estimate h_p computed using relation (5.10) can be derived via (5.11) and (5.9). It is related to the random nature of estimates A* and B* in (5.7) and (5.9) and, as well, to the same random nature of estimates λ* in (1.6). This means that the POT method gives the "true" value of h_{max} that is located in a two - dimensional area of uncertainty range. The first co-ordinate of this area characterises the possible range of estimates h_p* in terms of wave height (due to uncertainty in *A** and *B**) while the other is related to uncertainty in *p** (due to variations of λ* resulting from using data for a limited number of storms). These areas are shown in Fig. 5.3.

Thus, the POT method estimates depend on the choice of approximations for corresponding distributions. Of course, estimates obtained using other methods do also. However, unlike other methods, in the POT approach the uncertainty is connected both with the wave height h_p^* and return period. For example, the 25 year wave height estimate in Fig.5.3 is found to be in the range of 7.2 – 8.4 (m), and return period is in the range of 20-45 years.

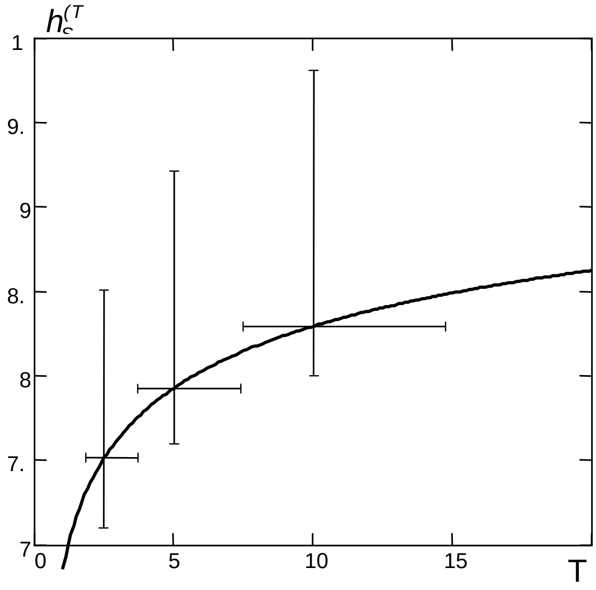

Figure 5.3. Joint uncertainty ranges of POT significant wave height estimates at return periods of 25, 50, and 100 years.

---oooOooo---