

## Chapter 3

# GPD-POT and GEV block maxima

This chapter is devoted to the relation between POT models and Block Maxima (BM). We only consider the classical frameworks where POT excesses are assumed to be GPD, and where the BM follow a GEV distribution. We also study the relation between POT and the classical  $r$  largest statistics context.

Although the BM and  $r$  largest are usually derived from asymptotic considerations, it is quite well-known that the same models result from a temporal aggregation of the Marked Process  $[T_k, X_k]$  as used in POT.

“The distribution of the maximum of a Poisson number of i.i.d. excesses over a high threshold is GEV”. (*Embrechts et al. 1996, chap. 3*).

However, a number of details are most generally omitted. These relate to the possibility that a block has no observations, in which case the maximum does not exist.

### 3.1 GPD-POT to GEV-max

#### 3.1.1 Temporal aggregation of the marked process

In this section we consider the case of a partial observation (or temporal aggregation) for the marked process  $[T_k, X_k]$ . We assume that we have  $B$  disjoint periods or *blocks* with known durations  $w_b$  for  $b = 1, 2, \dots, B$ . We consider the following r.v.s representing the number of events and the maximum of the marks for the block  $b$

$$N_b := \#\{k; T_k \in \text{block } b\}, \quad M_b := \max_{T_k \in \text{block } b} X_k.$$

Observe that  $M_b$  is defined only when the block  $b$  contains at least one observation, a condition which is fulfilled with the probability  $\Pr\{N_b \neq 0\} = 1 - \exp\{-\lambda w_b\}$ . When a block duration is small relative to  $1/\lambda$ , this probability is not that close to one; e.g. for  $\lambda w_b = 1$  we find  $\Pr\{N_b \neq 0\} \approx 0.63$ . Moreover for  $B = 10$  blocks with  $\lambda w_b = 1$ , there is about one chance in a hundred that all blocks have an observation. When  $B$  is large enough, systematically missing observations  $M_b$  due to empty blocks will ineluctably occur, unless  $\lambda$  is large enough. This will be discussed later.

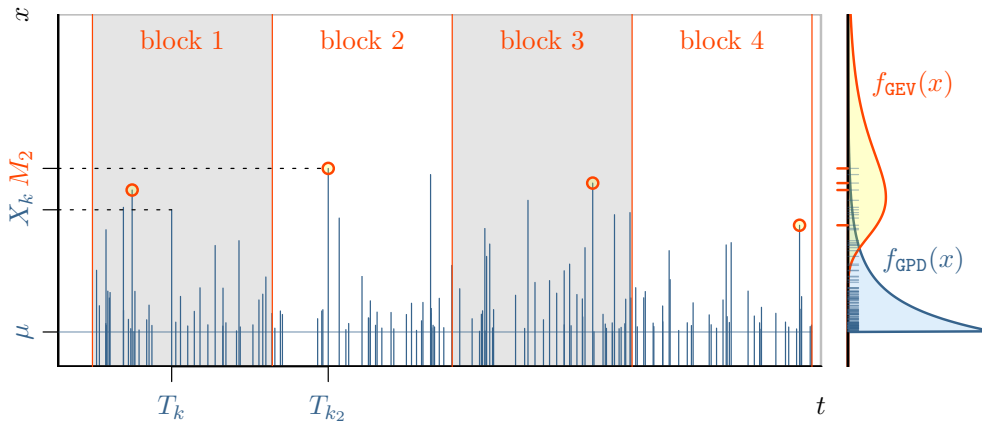


Figure 3.1: Temporal aggregation of the marked process with constant block durations  $w_b \equiv w$ . The distribution of the marks  $X_k$  is  $\text{GPD}(\mu, \sigma, \xi)$  while block maxima  $M_b$  have distribution  $\text{GEV}(\mu^*, \sigma^*, \xi^*)$ .

Since  $N_b$  and the observations  $X_k$  are independent and since  $M_b$  is the maximum of  $N_b$  independent r.v.s we know that

$$\Pr [M_b \leq x | N_b = k] = F_X(x)^k \quad \text{for } k \geq 1, \quad (3.1)$$

which allows the determination of the distribution of  $M_b$  in the following theorem. Since the maxima  $M_b$  corresponding to disjoint blocks are independent, the joint distribution of the maxima results. When  $N_b = 0$  we can set  $M_b := -\infty$  and the joint distribution is of mixed type with a positive probability mass on vectors with some  $M_b$  equal to  $-\infty$ .

It will be convenient to denote by  $\mathcal{I}_{\text{GPD}}(\mu, \sigma, \xi)$  the support of the GPD with parameters  $\mu$ ,  $\sigma$  and  $\xi$ . Similar notations will be used for the GEV distribution. The support of the block maxima  $M_b$  is the same as that of the marks  $X_k$ .

**Theorem 3.1.** *The r.v.s  $M_b$  corresponding to disjoint blocks are independent. The marginal distribution of  $M_b$  is given by*

$$\Pr\{N_b \neq 0, M_b \leq x\} = \exp\{-\lambda w_b S_X(x)\} - \exp\{-\lambda w_b\}. \quad (3.2)$$

If the marks are GPD with  $X_k \sim \text{GPD}(\mu, \sigma, \xi)$ , then for all  $x \in \mathcal{I}_{\text{GPD}}(\mu, \sigma, \xi)$ , we have

$$\Pr\{N_b \neq 0, x \leq M_b \leq x + dx\} = f_{\text{GEV}}(x; \mu_b^*, \sigma_b^*, \xi_b^*) dx \quad (3.3)$$

where the GEV parameters are given by

$$\mu_b^* = \mu + \frac{(\lambda w_b)^\xi - 1}{\xi} \sigma, \quad \sigma_b^* = (\lambda w_b)^\xi \sigma, \quad \xi_b^* = \xi, \quad \text{for } \xi \neq 0, \quad (3.4)$$

or by

$$\mu_b^* = \mu + \log(\lambda w_b) \sigma, \quad \sigma_b^* = \sigma, \quad \xi_b^* = \xi = 0, \quad \text{for } \xi = 0, \quad (3.5)$$

depending on the value of  $\xi$ . In both cases, the likelihood of an observation  $M_b$  is computed as if  $M_b$  comes from a sample of  $\text{GEV}(\mu_b^*, \sigma_b^*, \xi_b^*)$ .

When the blocks have the same duration  $w_b \equiv w$ , the maxima  $M_b$  form an i.i.d. sample of the GEV distribution  $\text{GEV}(\mu^*, \sigma^*, \xi^*)$ .

*Proof.* In the proof we will omit the index  $b$ . Consider  $x$  in the support of the distribution of the marks  $X_k$ . We have

$$\Pr[M \leq x | N \neq 0] = \sum_{k=1}^{\infty} \Pr[M \leq x | N = k] \times \Pr[N = k | N \neq 0] = \sum_{k=1}^{\infty} F_X(x)^k \Pr[N = k | N \neq 0]$$

where (3.1) was used in the second equality. Now, multiplying by  $\Pr\{N \neq 0\}$

$$\begin{aligned} \Pr\{N \neq 0, M \leq x\} &= \sum_{k=1}^{\infty} F_X(x)^k \times \Pr\{N = k\} = \sum_{k=1}^{\infty} F_X(x)^k \times e^{-\lambda w} [\lambda w]^k / k! \\ &= e^{-\lambda w} \sum_{k=1}^{\infty} [\lambda w F_X(x)]^k / k! = e^{-\lambda w} \left[ \exp\{\lambda w F_X(x)\} - 1 \right] \\ &= \exp\{-\lambda w S_X(x)\} - \exp\{-\lambda w\} \end{aligned}$$

which is (3.3). By derivation with respect to  $x$ , we get

$$\Pr\{N \neq 0, x \leq M \leq x + dx\} = \frac{d}{dx} \left\{ e^{-\lambda w S_X(x)} \right\} \times dx. \quad (3.6)$$

From now on let us assume that the marks are  $\text{GPD}(\mu, \sigma, \xi)$  and that  $\mu^*, \sigma^*$  and  $\xi^*$  are given by (3.4) for  $\xi \neq 0$  or (3.5) for  $\xi = 0$ . First, since  $\mu - \sigma/\xi = \mu^* - \sigma^*/\xi^*$  for  $\xi \neq 0$ , it is easy to see that for any vector  $[\mu, \sigma, \xi]^T$  we have

$$\mathcal{I}_{\text{GPD}}(\mu, \sigma, \xi) = \mathcal{I}_{\text{GPD}}(\mu^*, \sigma^*, \xi^*) \subset \mathcal{I}_{\text{GEV}}(\mu^*, \sigma^*, \xi^*)$$

see figure 3.2. Moreover, we have for any  $x \in \mathcal{I}_{\text{GPD}}(\mu, \sigma, \xi)$

$$F_{\text{GEV}}(x; \mu^*, \sigma^*, \xi^*) = \exp\{-S_{\text{GPD}}(x; \mu^*, \sigma^*, \xi^*)\}$$

which can be checked from the closed form expressions. It will thus be enough to prove that for any  $x \in \mathcal{I}_{\text{GPD}}(\mu, \sigma, \xi)$  we have

$$\lambda w S_{\text{GPD}}(x; \mu, \sigma, \xi) = S_{\text{GPD}}(x; \mu^*, \sigma^*, \xi^*) \quad (3.7)$$

indeed, the derivative at the right hand side of (3.6) will thus be the GEV density  $f_{\text{GEV}}(x; \mu^*, \sigma^*, \xi^*)$ . Separating the two cases  $\xi \neq 0$  and  $\xi = 0$ , the verification of (3.7) is simple algebra.  $\square$

**Remark.** Note that (3.3) only holds when  $x$  is in the support  $\mathcal{I}_{\text{GPD}}(\mu, \sigma, \xi)$  and it does not hold over the full superset  $\mathcal{I}_{\text{GEV}}(\mu^*, \sigma^*, \xi^*)$ . It is easy to check that by integrating (3.3) with respect to  $x \in \mathcal{I}_{\text{GPD}}(\mu, \sigma, \xi)$  that we get  $\Pr\{N_b \neq 0\} = 1 - \exp\{-\lambda w_b\}$ .  $\blacksquare$

Note that while  $\lambda$  is related to a time scale (it expresses as an inverse time), no time unit is found in the GEV parameters. The reason is that the time unit is hidden in the block duration  $w$  which is needed to compute the return level curve in time units (typically years).

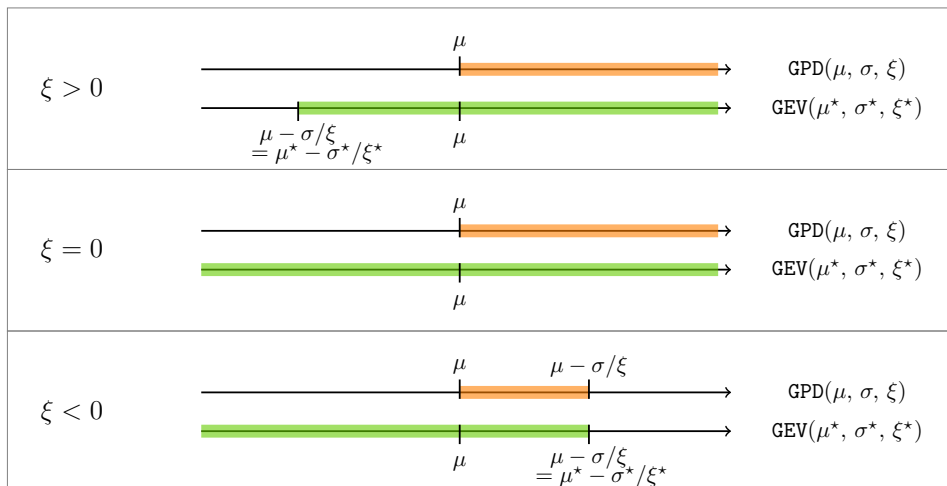


Figure 3.2: Supports  $\mathcal{I}_{\text{GPD}}(\mu, \sigma, \xi)$  and  $\mathcal{I}_{\text{GEV}}(\mu^*, \sigma^*, \xi^*)$ .

### 3.1.2 Links with Extreme Value regression

In the general case where the block duration  $w_b$  is not constant, the distribution of  $M_b$  depends on  $w_b$ . Ignoring the previous derivation, one could have used  $w_b$  as a covariate in an extreme value regression. It is very unlikely that by proceeding in this way we would find the exact relations to the covariate as given in (3.4). In the exponential case  $\xi = 0$ , the exact form of dependence is quite usual

$$\mu_b^* = \beta_0 + \beta_1 \log w_b, \quad \sigma_b^* = \sigma^* \text{ (constant),}$$

with parameters  $\beta_0, \beta_1, \sigma^*$ . Thus the location parameter is related to the log duration of the blocks, which may seem natural. However, when  $\xi \neq 0$  the true relationship would require the links

$$\mu_b^* = \beta_0 + \beta_1 w_b^{\xi^*}, \quad \sigma_b^* = \gamma_1 w_b^{\xi^*}, \quad \xi^* = \xi \text{ (constant),}$$

and the parameters  $\beta_0, \beta_1, \gamma_1$  and  $\xi^*$ . These equations do not fit in the standard framework where each of the three parameters is connected to the covariates through its own link function (Coles 2001, chap. 5). Thus, the very simple situation of a temporal aggregation does not lead to a simple extreme value regression. Note also that as a function of  $w$ , the variations of the term  $w^{\xi^*}$  are large for small values of  $w$ , since in practice  $|\xi^*| \leq 1$ .

## 3.2 GEV-max to GPD-POT

### 3.2.1 Disaggregation for constant block duration

#### Problem

We now assume to be given a sequence  $M_b$  corresponding to disjoint blocks  $b = 1, 2, \dots, B$  with the same duration  $w_b \equiv w$ , so that the  $M_b$  form a sample of a GEV distribution. In other words, we have partial observations of the marked process. We may then estimate the parameters of the underlying marked process and infer on them. However, the marked process embeds four

parameters  $\lambda$ ,  $\mu$ ,  $\sigma$  and  $\xi$ , while the GEV distribution only involves *three* parameters  $\mu^*$ ,  $\sigma^*$  and  $\xi^*$ . Given the vector  $\boldsymbol{\theta}^* := [\mu^*, \sigma^*, \xi^*]^\top$ , there is an infinity of vectors  $\boldsymbol{\theta} = [\lambda, \mu, \sigma, \xi]^\top$  satisfying the relations (3.4). The marked process model that generated the observations  $M_b$  can correspond to any vector provided that all the  $M_b$  lie in the interior of the support  $\mathcal{I}_{\text{GPD}}(\mu, \sigma, \xi)$ . There is an infinity of vectors  $\boldsymbol{\theta} = [\lambda, \mu, \sigma, \xi]^\top$  satisfying these conditions that have the same log-likelihood. The corresponding marked process models can be said *observationally equivalent* with respect to the given sequence of block maxima  $M_b$ .

**Remark.** The observational equivalence is tightly related to the ‘‘POT-stability’’ property of the GPD. By increasing the threshold  $u$  and lowering the rate  $\lambda$  it is possible to maintain the same return level curve. ■

A natural idea to overcome the problem of identifiability is to fix one of the four POT parameters. The following two strategies can be considered

- Choose the rate  $\lambda > 0$ , and then compute or estimate the GPD parameters  $\mu$ ,  $\sigma$  and  $\xi$ .
- Choose the GPD location  $\mu$ , and then compute or estimate the rate  $\lambda$  as well as  $\sigma$  and  $\xi$ .

In the first case, any positive rate  $\lambda > 0$  can be chosen and we simply have a re-parameterisation of the GEV distribution. Taking  $\lambda = 1/w$ , the three GPD parameters of the renewal process become identical to their GEV correspondent, that is:  $\mu = \mu^*$ ,  $\sigma = \sigma^*$  and  $\xi = \xi^*$ .

The second approach is very attractive when the model must be fitted using observations  $M_b$ , since it boils down to a POT estimation from aggregated data as discussed now.

### Fixing $\mu$ : a ‘‘GEV to POT’’ function

The relations (3.4) or (3.5) give the BM parameter vector  $\boldsymbol{\theta}^*$  as a function of the POT parameter  $\boldsymbol{\theta}$ ; we aim to clarify here a possible inverse relation, i.e. the determination of  $\boldsymbol{\theta}$  from  $\boldsymbol{\theta}^*$  for a fixed value of  $\mu$ .

For the GEV context we will denote by  $\Theta^* = \{[\mu^*, \sigma^*, \xi^*]^\top; \sigma^* > 0\}$  the domain of admissible parameters. The notations  $\boldsymbol{\theta}_{(-\lambda)}$  and  $\boldsymbol{\theta}_{(-\mu)}$  are for the vectors obtained by omitting  $\lambda$  or  $\mu$  in the vector  $\boldsymbol{\theta} = [\lambda, \mu, \sigma, \xi]^\top$ . The relations (3.4) giving  $\boldsymbol{\theta}^*$  as a function of  $\boldsymbol{\theta}$  and  $\mu$  can be written as

$$\boldsymbol{\theta}^* = \boldsymbol{\psi}^*(\boldsymbol{\theta}_{(-\mu)}; \mu) \quad (3.8)$$

which can be called a ‘‘POT to GEV’’ transformation. The Jacobian of this transformation is easily computed, see A.3 page 77. The same notations can be used for the Gumbel context and the relations (3.5), provided it is understood then that  $\boldsymbol{\theta}^* = [\mu^*, \sigma^*]^\top$  and  $\boldsymbol{\theta} = [\lambda, \mu, \sigma]^\top$ .

**Theorem 3.2.** *Let  $\boldsymbol{\theta}^* = [\mu^*, \sigma^*, \xi^*]^\top$  be a vector of GEV parameters with  $\sigma^* > 0$ . A solution  $\boldsymbol{\theta}_{(-\mu)} = [\lambda, \sigma, \xi]^\top$  of (3.4) exists if and only if  $\mu$  is an interior point of the support  $\mathcal{I}_{\text{GEV}}(\boldsymbol{\theta}^*)$ . Then the solution  $\boldsymbol{\theta}_{(-\mu)}$  is unique and we may write it as a function of  $\boldsymbol{\theta}^*$  and  $\mu$ , i.e. as  $\boldsymbol{\theta}_{(-\mu)} = \boldsymbol{\psi}(\boldsymbol{\theta}^*; \mu)$ .*

*For a vector of Gumbel parameters  $\boldsymbol{\theta}^* = [\mu^*, \sigma^*]^\top$  with  $\sigma^* > 0$ , a unique solution  $\boldsymbol{\theta}_{(-\mu)}$  of (3.5) exists.*

*Proof.* Consider the GEV case. From (3.4) we have by simple algebra

$$[\lambda w]^{-\xi^*} = 1 + \xi^* (\mu - \mu^*) / \sigma^*. \quad (3.9)$$

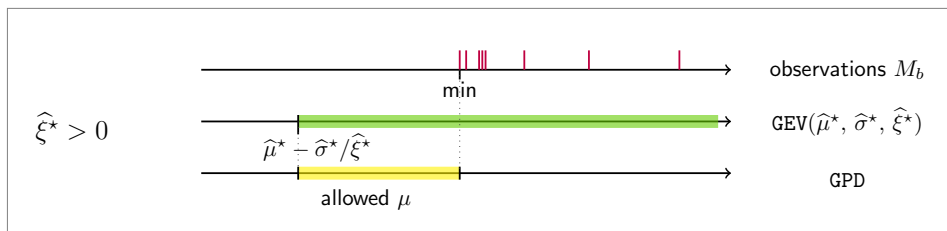


Figure 3.3: The fixed parameter  $\mu$  must lie in the interior of the support  $\mathcal{I}_{\text{GEV}}(\hat{\boldsymbol{\theta}}^*)$ . When  $\hat{\xi}^* > 0$ , we must have  $\mu > \hat{\mu}^* - \hat{\sigma}^*/\hat{\xi}^*$ .

We have  $\lambda > 0$  if and only the right hand side is positive, i.e. if  $\mu$  is located in the interior of the support  $\mathcal{I}_{\text{GEV}}(\boldsymbol{\theta}^*)$ . We may then take the power  $-1/\xi^*$  of each side, leading to  $\lambda w = -\log F_{\text{GEV}}(\mu; \boldsymbol{\theta}^*)$ . We then easily find  $\sigma$  and  $\xi$ . To summarise

$$\lambda = -\frac{1}{w} \log F_{\text{GEV}}(\mu; \boldsymbol{\theta}^*), \quad \sigma = (\lambda w)^{-\xi^*} \sigma^*, \quad \xi = \xi^*. \quad (3.10)$$

The proof is straightforward for the Gumbel case. □

### Fitting BM from POT

Given  $B$  block maxima  $M_b$  we now consider the estimation of a GEV distribution by using a POT model with  $\mu$  fixed. Using the notations of the previous section, we can estimate  $\boldsymbol{\theta}_{(-\mu)}$  rather than the vector  $\boldsymbol{\theta}^*$  of GEV parameters and get this later using the POT to BM transform described in the previous section. More precisely, we can maximise with respect to  $\boldsymbol{\theta}_{(-\mu)}$  the POT likelihood  $L_{\text{POT}}(\boldsymbol{\theta}_{(-\mu)}; \mu)$  where the second argument is meant to recall that  $\mu$  is used as the threshold required in POT.

Not all values of the fixed parameter  $\mu$  can be chosen. The fixed value of  $\mu$  must obviously be such that  $M_b > \mu$  for every block  $b$ , and it must also lie in the interior of the support  $\mathcal{I}_{\text{GEV}}(\hat{\boldsymbol{\theta}}^*)$ , see figure 3.3. Assume that we are given a subset  $\Theta_0^*$  of the parameter space  $\Theta^*$  containing all the parameters  $\boldsymbol{\theta}^*$  that could have generated the observations, and that  $\mu$  is an interior point of  $\mathcal{I}_{\text{GEV}}(\boldsymbol{\theta}^*)$  for all  $\boldsymbol{\theta}^* \in \Theta_0^*$ . Then

$$L_{\text{GEV}}(\boldsymbol{\theta}^*) = L_{\text{POT}}(\boldsymbol{\theta}_{(-\mu)}; \mu), \quad \text{with} \quad \boldsymbol{\theta}_{(-\mu)} = \boldsymbol{\psi}(\boldsymbol{\theta}^*; \mu)$$

holds for all  $\boldsymbol{\theta}^* \in \Theta_0^*$ . It is thus clear that maximising the POT likelihood with respect to  $\boldsymbol{\theta}_{(-\mu)}$  with  $\mu$  fixed and transforming it with the function  $\boldsymbol{\psi}^*$  will lead to the same solution  $\hat{\boldsymbol{\theta}}^*$  as fitting a GEV by ML. In other words

$$\hat{\boldsymbol{\theta}}_{(-\mu)} = \boldsymbol{\psi}(\hat{\boldsymbol{\theta}}^*, \mu) \quad (3.11)$$

and the estimated joint distribution for the observations  $M_b$  will be the same in the two cases.

An advantage of the POT approach lies in the possibility of likelihood concentration seen in chapter 1 (section 1.1.5 page 9). We can fit the model using a two-parameter optimisation involving  $\sigma$  and  $\xi$  while  $\lambda$  is concentrated out through

$$\hat{\lambda} = \frac{B}{\sum_b w_b S_{\text{GPD}}(M_b; \mu, \hat{\sigma}, \hat{\xi})} \quad (3.12)$$