

STK4150/9150 - Environmental and Spatial Statistics

Joint modeling of bulk and tail

Thordis Thorarinsdottir

Spring 2026

Joint modeling of bulk and tail

Environmental variables (rainfall, river flow, wind, pollution) often show:

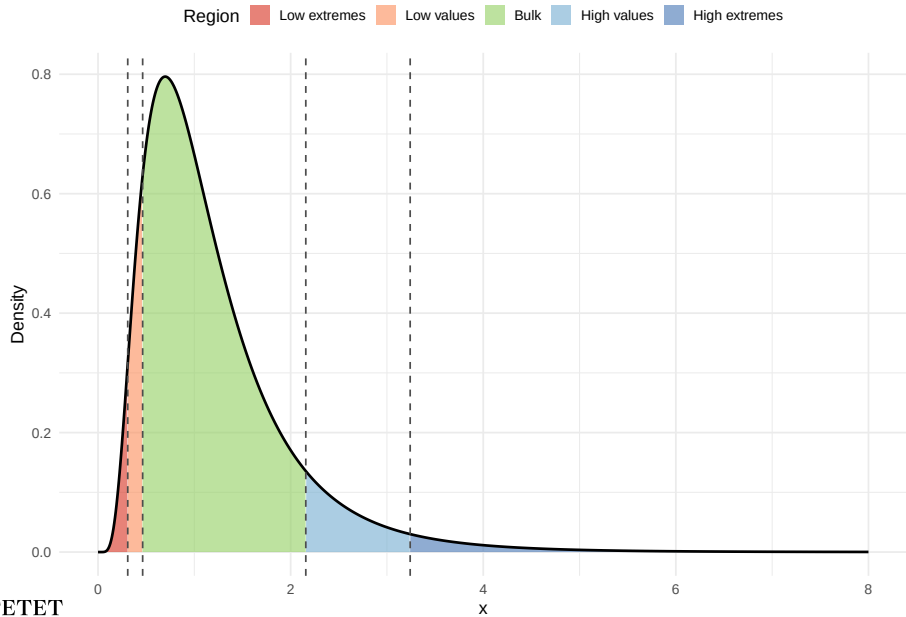
- **Bulk:** typical values
- **Tail:** rare extremes

Classical extreme value methods:

- Block maxima $\Rightarrow$ GEV
- Peaks-over-threshold (POT) $\Rightarrow$ GP for exceedances

Today:

- From tail-only GP to joint bulk–tail models
- Focus: Extended GP model with a bulk distribution B (EGPD)



Motivation: Why joint bulk–tail modeling?

In practice we want:

- Return levels for intermediate and high quantiles
- Risk measures over the entire distribution

Problems with classical PoT:

- Threshold selection is hard
- Below-threshold model often ad hoc

Idea: build a single parametric model:

- GP-consistent in the tail
- Flexible in the bulk

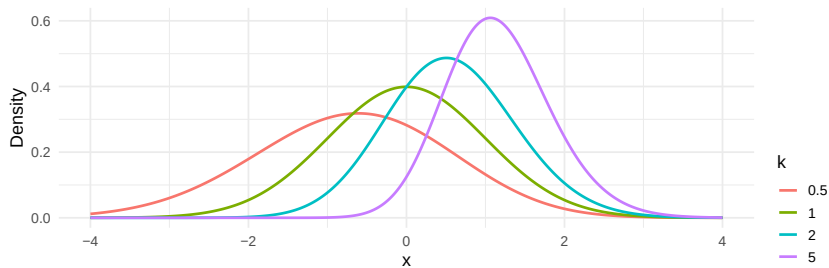
Overall ideas

How: We add more parameters in specific ways

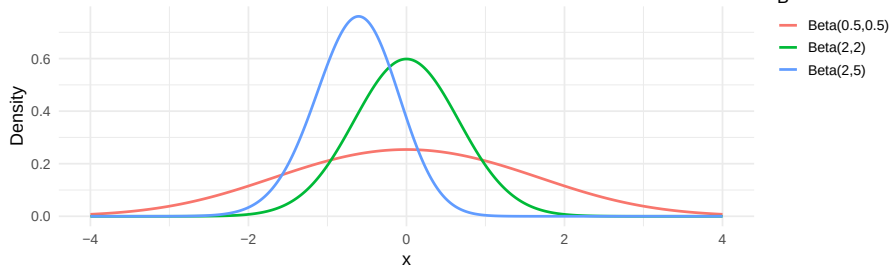
Two observations that the EGPD is built on:

1. For any interval $I \subseteq \mathbb{R}$, if F is a cumulative distribution function (cdf) on I and $k > 0$, then F^k is also a cdf on I .
2. Let $B : [0, 1] \rightarrow [0, 1]$ be non-decreasing and right-continuous with $B(0) = 0$ and $B(1) = 1$ (that is, a cdf on $[0, 1]$). Then for any cdf F on I , the composition $G(x) = B(F(x))$ is again a cdf on I .

Density of $F(x)^k$ for selected k , with F the standard normal cdf



Density of $B(F(x))$ with B a beta cdf and F the standard normal cdf



Recap: Generalized Pareto (GP) for exceedances

For a large threshold u , exceedances

$$Y = X - u \mid X > u$$

are approximately GP with cdf

$$F_{\text{GP}}(y) = 1 - \left(1 + \xi \frac{y}{\sigma}\right)^{-1/\xi}, \quad y \geq 0, \quad 1 + \xi \frac{y}{\sigma} > 0,$$

where ξ is the shape parameter and $\sigma > 0$ is the scale parameter.

The tail behavior is controlled by the shape parameter:

- $\xi > 0$: heavy tail (Pareto-type)
- $\xi = 0$: exponential limit
- $\xi < 0$: finite upper endpoint at $-\sigma/\xi$

Design goals for bulk–tail joint models

We seek a family F fulfilling:

1. **Tail correctness (EVT compatibility)**
 - Tail index driven by ξ (as in GP).
2. **Bulk flexibility**
 - Extra parameters for a flexible bulk.
3. **Single, smooth distribution**
 - No artificial joining at a threshold; F is a proper cdf on I .
4. **Modeling convenience**
 - Easy interpretation
 - Can be used in regression, spatial and hierarchical models.

GP building block in standardized form

We introduce a standardized GP building block with scale 1:

$$H_{\xi}(z) = \begin{cases} 1 - (1 + \xi z)^{-1/\xi}, & \xi \neq 0, \\ 1 - e^{-z}, & \xi = 0, \end{cases} \quad z \geq 0, \quad 1 + \xi z > 0.$$

Note: This is just F_{GP} with scale parameter fixed at 1.

For our exceedance Y , we will use $z = y/\sigma$.

We then modify H_{ξ} in a controlled way to obtain a bulk-tail model.

A simple extended GP (Type 0)

First, a simple extension of the standardized GP is

$$F_0(y) = H_{\xi}^{\varepsilon}\left(\frac{y}{\sigma}\right), \quad y \geq 0,$$

with parameters:

- ξ : tail index (as in GP)
- $\sigma > 0$: scale
- $\varepsilon > 0$: extra shape for the bulk / lower tail

Note:

- If $\varepsilon = 1$, we recover the standard GP.
- This provides some bulk flexibility while preserving the GP tail index.

General bulk–tail model with B

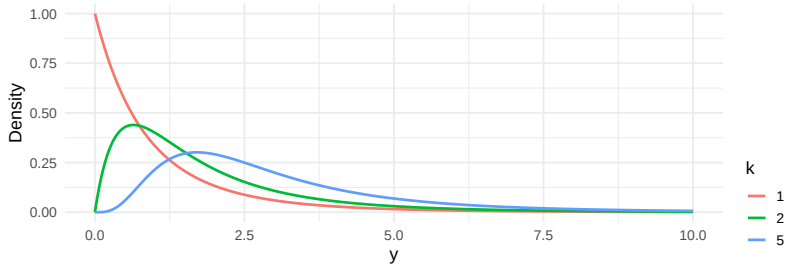
Add a bulk cdf B on $[0, 1]$ with density b .

Define the extended GP (EGPD) model:

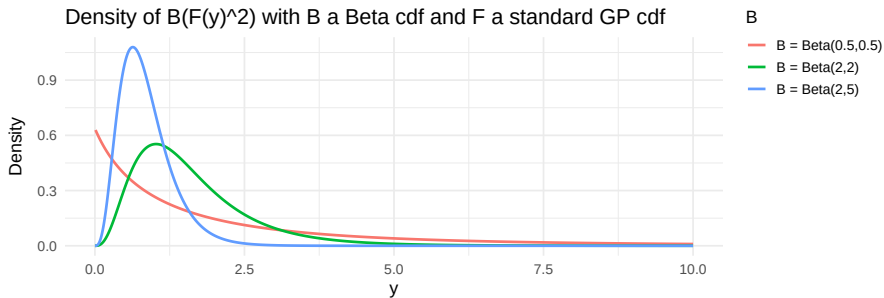
$$F(y) = B\left(H_{\xi}^{\varepsilon}\left(\frac{y}{\sigma}\right)\right), \quad y \geq 0.$$

- $B : [0, 1] \rightarrow [0, 1]$ is
 - non-decreasing,
 - right-continuous,
 - with $B(0) = 0, B(1) = 1$.
- B provides extra flexibility in the bulk: Choice of B (e.g., Beta cdf, Bernstein polynomials, truncated distributions) shapes the body.
- Special case $B(u) = u$ gives back the simple EGPD $F_0(y)$.

Density of $F(y)^k$ where F is standard GP cdf



Density of $B(F(y)^2)$ with B a Beta cdf and F a standard GP cdf



Distribution and density

Let $z = y/\sigma$ and write $U(z) = H_\xi^\varepsilon(z)$.

The cdf is given by

$$F(y) = B(U(z)), \quad U(z) = H_\xi^\varepsilon(z).$$

We obtain the density by the chain rule:

$$f(y) = \frac{\varepsilon}{\sigma} h_\xi(z) H_\xi^{\varepsilon-1}(z) b(H_\xi^\varepsilon(z)), \quad y \geq 0,$$

where $h_\xi(z)$ is the GP density for H_ξ .

Tail behavior

As $y \rightarrow$ upper endpoint (or ∞ if $\xi \geq 0$),

$$1 - F(y) = 1 - B(H_\xi^\varepsilon(z)) \approx b(1) \varepsilon [1 - H_\xi(z)].$$

That is, the tail behavior is still governed by ξ (it is EVT-consistent), provided $b(1)$ is finite and positive.

Interpretation of parameters

- ξ (shape / tail index)
 - Controls tail heaviness as in GP.
- σ (scale)
 - Stretches/compresses the y -axis.
- ε (body parameter)
 - $\varepsilon = 1$: reduces to GP if $B(u) = u$.
 - $\varepsilon < 1$: more mass near the lower tail; lighter middle range.
 - $\varepsilon > 1$: less mass near very small values; heavier intermediate range.
- B (bulk cdf on $[0, 1]$)
 - Provides additional bulk shape beyond ε .
 - Different parametrizations (e.g. Beta-family, Bernstein polynomials) yield a variety of shapes.

Quantile function and simulation

Let B^{-1} denote the quantile function of B on $[0, 1]$, and H_{ξ}^{-1} the GP quantile.

To simulate Y from F :

1. Draw $U \sim \text{Unif}(0, 1)$.
2. Compute

$$W = B^{-1}(U).$$

3. Set

$$Z = H_{\xi}^{-1}\left(W^{1/\varepsilon}\right), \quad Y = \sigma Z.$$

The quantile function is thus:

$$Q(p) = \sigma H_{\xi}^{-1}\left(\left[B^{-1}(p)\right]^{1/\varepsilon}\right), \quad 0 < p < 1.$$

Example: B as Beta(2, 5) on $[0, 1]$

We now fix the bulk cdf $B(u)$ to be Beta(2, 5) with density

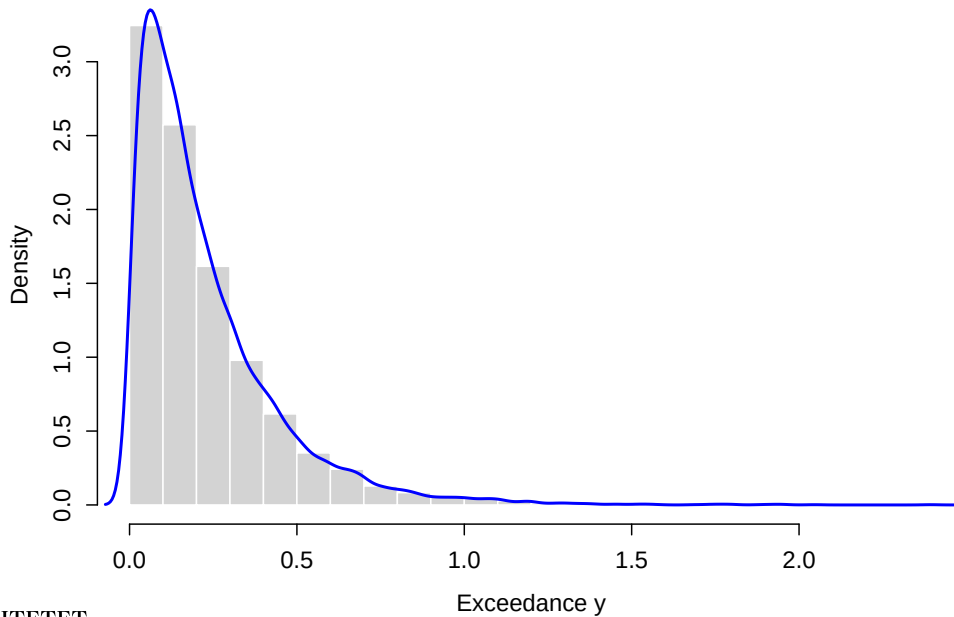
$$b(u) = \frac{\Gamma(2+5)}{\Gamma(2)\Gamma(5)} u^{2-1} (1-u)^{5-1} = 30 u(1-u)^4, \quad 0 < u < 1.$$

The EGPD cdf becomes

$$F(y) = B\left(H_{\xi}^{\varepsilon}\left(\frac{y}{\sigma}\right)\right) = \text{Beta}\left(H_{\xi}^{\varepsilon}\left(\frac{y}{\sigma}\right); 2, 5\right).$$

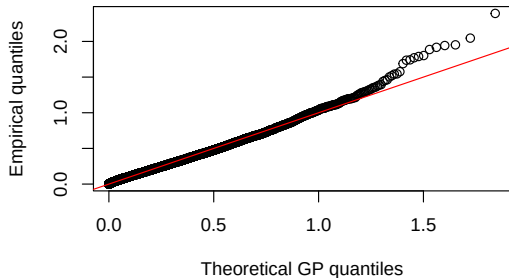
This gives a EGPD model with:

- Tail index ξ
- Scale σ
- Body parameter ε
- Fixed bulk shape governed by Beta(2, 5)

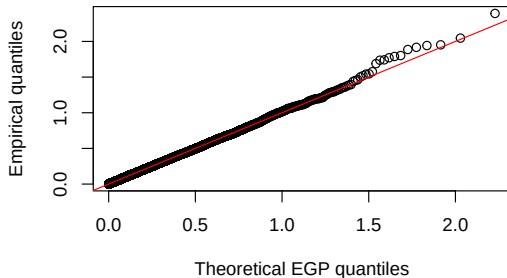


Comparison of GPD and EGDP fits

QQ plot: GPD fit



QQ plot: EGPD (B = Beta(2,5)) fit



AIC_gpd	AIC_egp	BIC_gpd	BIC_egp
-9732.448	-9981.435	-9718.027	-9959.804