Robust and efficient estimation for the Generalized Extreme-Value distribution

Università Ca'Foscari Venezia

Nathan Huet
Joint work with Ilaria Prosdocimi

September 11, 2025

Study of Extreme Values

Why? model, predict, understand, anticipate, or manage extreme phenomena such as heavy precipitation, devastating floods, stock market crashes...

Flood in Netherlands, 1953 (photo from *Watersnoodmuseum*).

Extreme Value Theory

Focus: observations outside the mass center of the distribution, *i.e.* in the tail of the distribution

Working assumptions on Z a random element

- convergence in distribution of maxima, *i.e.*

$$\lim_{n\to+\infty} \mathscr{L}\left(\frac{\max_{i=1}^{n} Z_i - b_n}{a_n}\right) = \mathscr{L}(X),$$

  with $Z_i \overset{i.i.d.}{\sim} Z$.

- convergence in distribution of excesses, *i.e.*

$$\lim_{t\to+\infty} \mathscr{L}(Z/t \mid \|Z\| \geq t) = \mathscr{L}(Z_\infty).$$

Convergence of maxima

Fisher–Tippett–Gnedenko theorem: if

$$\lim_{n\to+\infty} \mathscr{L}\Big(\frac{\max_{i=1}^n Z_i - b_n}{a_n}\Big) = \mathscr{L}(X)$$

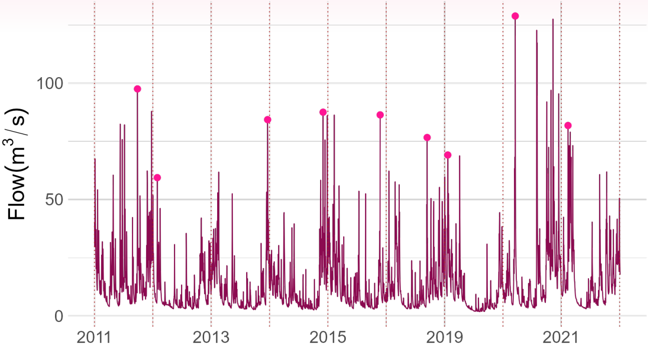
with non-degenerate X, then X follows a Generalized Extreme-Value (GEV) distribution, i.e.

$$\mathbb{P}(X \leq x) = \exp\Big(- \Big(1 + \xi\Big(\frac{x-\mu}{\sigma}\Big)^{-1/\xi}\Big)\Big)\mathbb{1}\{x \in D_{\mu,\sigma,\xi}\},$$

with $D_{\mu,\sigma,\xi}$ given by

$$D_{\mu,\sigma,\xi} = \begin{cases} [\mu - \sigma/\xi, +\infty[, & \text{if } \xi > 0; \\ \mathbb{R}, & \text{if } \xi = 0; \\ ]-\infty, \mu - \sigma/\xi], & \text{if } \xi < 0. \end{cases}$$

Block Maxima



- sample in practice : block maxima (that follows a GEV distribution) ⇝ **limited sample size**

⇝ estimating the parameters of a GEV distribution via maximum likelihood involves high variance and high sensitivity to outliers or very large observations.

⇝ **robust estimation methods**

Minimum density power divergence

[Basu et al., 1998]

Density Power Divergence $d_\alpha(g, f)$ between f and g:

$$d_\alpha(g, f) = \int_{\mathscr{X}} \left(f^{1+\alpha}(x) - \left(1 + \frac{1}{\alpha}\right) g(x) f^\alpha(x) + \frac{1}{\alpha} g^{1+\alpha}(x) \right) dx.$$

Minimum Density Power Divergence θ_α for a parametric density model $\mathscr{F} = \{f(x; \theta), x \in \mathscr{X}, \theta \in \Theta\}$:

$$\theta_\alpha \in \text{argmin}_{\theta \in \Theta} d_\alpha(g, f(\cdot; \theta)).$$

Minimum density power divergence estimator

Let $X_1, \ldots, X_n$ i.i.d. random element defined on $\mathscr{X}$. Denote by g_n their empirical density function.

A **Minimum Density Power Divergence Estimator** (MDPDE) $\hat{\theta}_\alpha \in \Theta$ is defined as

$$\hat{\theta}_\alpha \in \mathrm{argmin}_{\theta \in \Theta} d_\alpha(g_n, f(\cdot;\theta))$$

- for $\alpha \to 0$, the MDPDE is the MLE (efficient)
- for $\alpha = 1$, the MDPDE is the L^2-estimator (robust)

$\rightsquigarrow$ for $\alpha \in]0, 1[$ compromise between efficiency and robustness

MDPDE for GEV

◦ $X_1, ..., X_n$ i.i.d. GEV random variables; g_n their empirical density function.

◦ density model $\mathscr{F} = \{f(x;\mu,\sigma,\xi), x \in \mathbb{R}, (\mu,\sigma,\xi) \in \mathbb{R}\times]0,+\infty[\times\mathbb{R}\}$ with

$$f(x;\mu,\sigma,\xi) = \frac{1}{\sigma}\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-(\xi+1)/\xi} \exp\left(-\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)^{-1/\xi}\right)\right)\mathbb{1}\{x \in D_{\mu,\sigma,\xi}\}.$$

◦ MDPDE $(\hat{\mu}_\alpha, \hat{\sigma}_\alpha, \hat{\xi}_\alpha)$ for GEV

$$(\hat{\mu}_\alpha, \hat{\sigma}_\alpha, \hat{\xi}_\alpha) \in \operatorname{argmin}_{(\mu,\sigma,\xi)\in\mathbb{R}\times]0,+\infty[\times\mathbb{R}} d_\alpha(g_n, f(\cdot;\mu,\sigma,\xi))$$

◦ this approach has already been considered in an extreme framework in [Juárez and Schucany, 2004]

Asymptotic Normality

Theorem. Let (μ_0, σ_0, ξ_0) be the target parameters. Suppose $\xi_0 > -(1+\alpha)/(2+\alpha)$, for fixed $\alpha > 0$. Then, there exists a consistent sequence of MDPDE $\{(\hat{\mu}_\alpha, \hat{\sigma}_\alpha, \hat{\xi}_\alpha)\}$ for (μ_0, σ_0, ξ_0). In addition,

$$\sqrt{n}(\hat{\mu}_\alpha - \mu_0, \hat{\sigma}_\alpha - \sigma_0, \hat{\xi}_\alpha - \xi_0)^\top \xrightarrow{d} \mathcal{N}\left(0, J_\alpha^{-1}(\mu_0, \sigma_0, \xi_0) K_\alpha(\mu_0, \sigma_0, \xi_0) J_\alpha^{-1}(\mu_0, \sigma_0, \xi_0)\right),$$

as $n \to +\infty$.

- for $\alpha \to 0$, we obtain the classic restriction $\xi_0 > -1/2$ for the asymptotic normality of the MLE [Bücher and Segers, 2017].
- for $\alpha > 0$, the region on which the asymptotic normality holds is enlarged as compared to the MLE.

Experiments

Comparison of four estimators: MLE, MDPDE (with $\alpha = 0.05$), MDPDE (with $\alpha = 0.1$), MQE [Lin et al., 2024]

Contaminated model: $(1-\varepsilon)GEV(\mu_0, \sigma_0, \xi_0) + \varepsilon GEV(\mu_1, \sigma_1, \xi_1)$.

- true parameters: $\mu_0 = 0, \sigma_0 = 1, \xi_0 \in \{-0.3, 0, 0.3\}$
- contamination on scale parameter σ_1 and shape parameter ξ_1, **one at a time**;
- proportion of contamination : $\varepsilon = 0.1$; sample size : $n = 100$; number of replication : $d = 200$

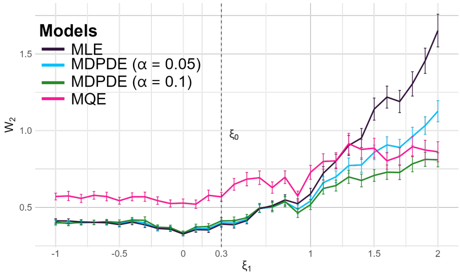
Performance measured according to the Wasserstein 2-distance

$$W_2(F_0, \hat{F}_0) = \left(\int_0^1 \left(F_0^{\leftarrow}(p) - \hat{F}_0^{\leftarrow}(p) \right)^2 dp \right)^{1/2},$$

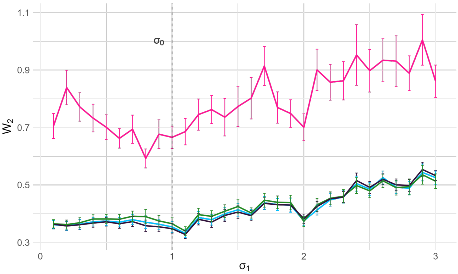
where $F_0^{\leftarrow}$ is the true quantile function and $\hat{F}_0^{\leftarrow}$ the empirical quantile function estimated by each model.

Experiments : positive shape parameter

true : $\xi_0 = 0.3, \sigma_0 = 1, \mu_0 = 0$
contamination : $\mu_1 = 0, \varepsilon = 0.1$
model : $(1-\varepsilon)GEV(\mu_0, \sigma_0, \xi_0) + \varepsilon GEV(\mu_1, \sigma_1, \xi_1)$



(a) $\sigma_1 = \sigma_0 = 1$, varying ξ_1

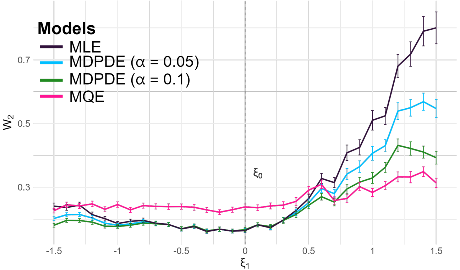


(b) $\xi_1 = \xi_0 = 0.3$, varying σ_1

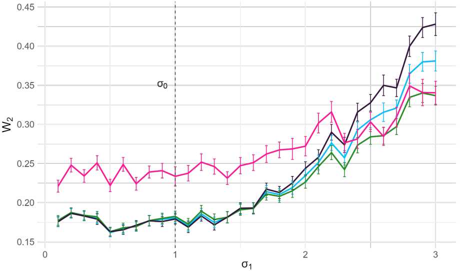
Figure: Average Wasserstein distance across various contaminated models.

Experiments : zero shape parameter

true : $\xi_0 = 0, \sigma_0 = 1, \mu_0 = 0$
contamination : $\mu_1 = 0, \varepsilon = 0.1$
model : $(1-\varepsilon)GEV(\mu_0, \sigma_0, \xi_0) + \varepsilon GEV(\mu_1, \sigma_1, \xi_1)$



(a) $\sigma_1 = \sigma_0 = 1$, varying ξ_1



(b) $\xi_1 = \xi_0 = 0$, varying σ_1

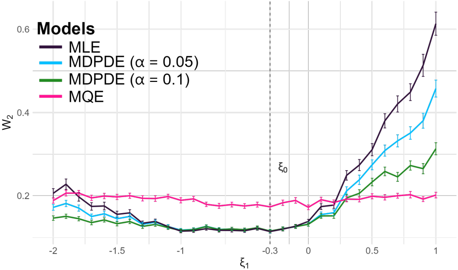
Figure: Average Wasserstein distance across various contaminated models.

Experiments : negative shape parameter

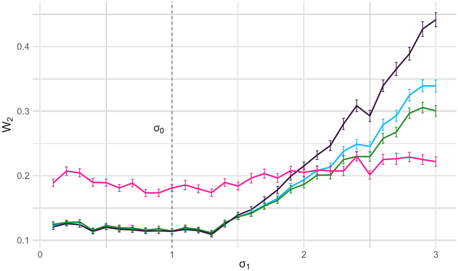
true : $\xi_0 = -0.3, \sigma_0 = 1, \mu_0 = 0$

contamination : $\mu_1 = 0, \varepsilon = 0.1$

model : $(1 - \varepsilon)GEV(\mu_0, \sigma_0, \xi_0) + \varepsilon GEV(\mu_1, \sigma_1, \xi_1)$



(a) $\sigma_1 = \sigma_0 = 1$, varying ξ_1

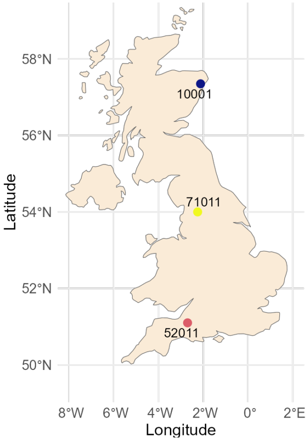


(b) $\xi_1 = \xi_0 = -0.3$, varying σ_1

Figure: Average Wasserstein distance across various contaminated models.

Application: flood frequency analysis in the UK

provided by the National River Flow Archive



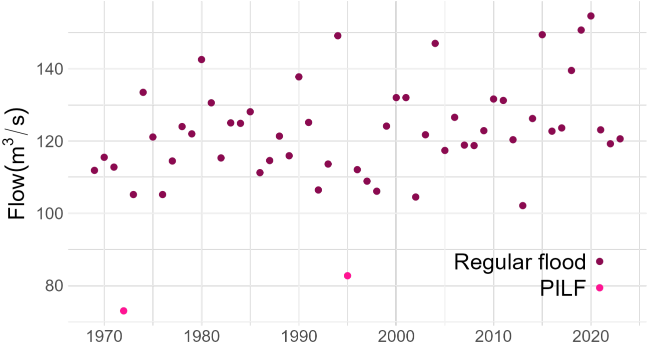
- 10001 : Ythan River — 1939-2023 — PILFs: 0.04
- 52011 : Cary River — 1965-2023 — PILFs: 0.03
- 71011 : Ribble River — 1970-2023 — PILFs: 0.04

- Data : annual maximum river flows

Potentially Influential Low Floods (PILFs)

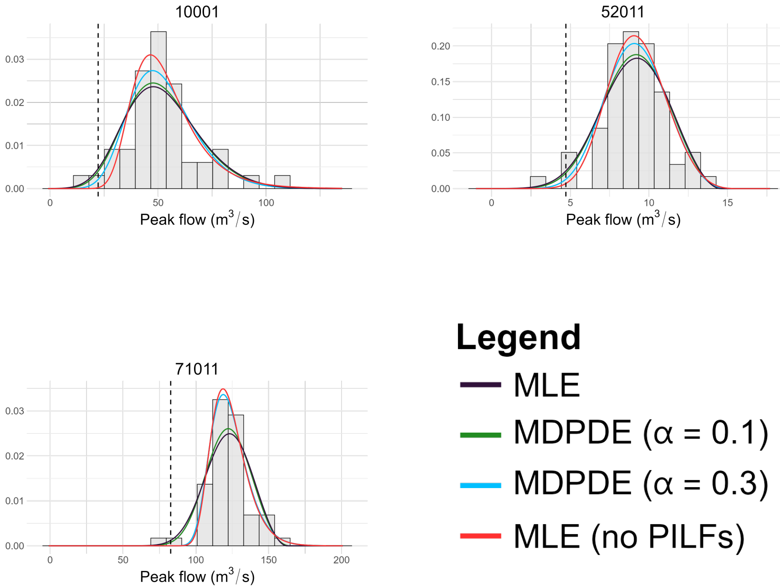
Why MDPDE? Presence of PILFs
$\rightarrow$ must be removed [England et al., 2018]

⚠ reduce the sample size even more



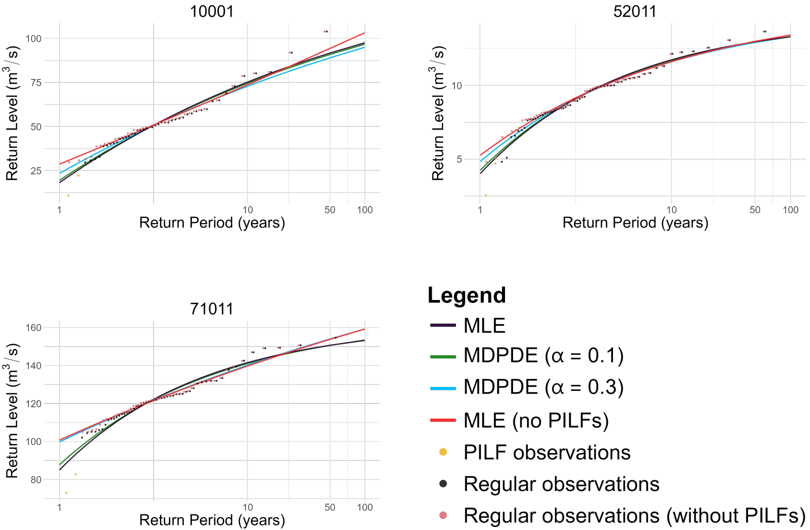
Comparison: MLE, MDPDE ($\alpha = 0.1$), MDPDE ($\alpha = 0.3$), MLE without the PILFs

Density plots



Return curves

Value expected once every Y years



Future works/Open questions

- data-driven method to choose α for a good compromise between efficiency and robustness?
- extension to non-stationary case
- other real-world applications?

References

- A. Basu, I. R. Harris, N. L. Hjort, and M. Jones., *obust and efficient estimation by minimising a density power divergence*, Biometrika, 1998;
- A. Bücher and J. Segers, *On the maximum likelihood estimator for the generalized extreme-value distribution*, Extremes, 2017;
- J. F. England Jr, T. A. Cohn, B. A. Faber, J. R. Stedinger, W. O. Thomas Jr, A. G. Veilleux, J. E. Kiang and R. R. Mason Jr, *Guidelines for determining flood flow frequency—Bulletin 17C*, US Geological Survey, 2018;
- S. F. Juárez and W. R. Schucany, *Robust and efficient estimation for the generalized pareto distribution*, Extremes, 2004;
- S. Lin, A. Kong and R. Azencott, *Multi-Quantile Estimators for the parameters of Generalized Extreme Value distribution*, arXiv, 2024.

Thank you for your attention!

Appendix

Involved quantities

(just for completeness)

Denote by $S(x;\mu,\sigma,\xi)$ and $i(x;\mu,\sigma,\xi)$ the score function and the information of the GEV distribution. Define the 3×3 matrices K_α and J_α as

$$K_\alpha(\mu,\sigma,\xi) = \int_{S_{\mu,\sigma,\xi}} S(x;\mu,\sigma,\xi)S^\top(x;\mu,\sigma,\xi)f^{1+2\alpha}(x;\mu,\sigma,\xi)dx - U_\alpha(\mu,\sigma,\xi)U_\alpha^\top(\mu,\sigma,\xi),$$

where

$$U_\alpha(\mu,\sigma,\xi) = \begin{bmatrix} \int_{S_{\mu,\sigma,\xi}} S_\mu(x;\mu,\sigma,\xi)f^{1+\alpha}(x;\mu,\sigma,\xi)dx \\ \int_{S_{\mu,\sigma,\xi}} S_\sigma(x;\mu,\sigma,\xi)f^{1+\alpha}(x;\mu,\sigma,\xi)dx \\ \int_{S_{\mu,\sigma,\xi}} S_\xi(x;\mu,\sigma,\xi)f^{1+\alpha}(x;\mu,\sigma,\xi)dx \end{bmatrix},$$

and

$$J_\alpha(\mu,\sigma,\xi) = \int_{S_{\mu,\sigma,\xi}} S(x;\mu,\sigma,\xi)S^\top(x;\mu,\sigma,\xi)f^{1+\alpha}(x;\mu,\sigma,\xi)dx.$$

Influence Function

Sensitivity Curve. For a sample statistic T,

$$SC_n(x) = \frac{T(X_1, ..., X_{n-1}, x) - T(X_1, ..., X_{n-1})}{(1/n)}.$$

Influence Function. For a sample statistic T,

$$IF(x) := \lim_{n \to +\infty} SC_n(x).$$

Example: for T the mean,

$$\begin{aligned} SC_n(x) &= \frac{mean_n(X_1, ..., X_{n-1}, x) - mean_{n-1}(X_1, ..., X_{n-1})}{(1/n)} \\ &= x - mean_{n-1}(X_1, ..., X_{n-1}) \\ &\to x - \mathbb{E}[X] = IF(x). \end{aligned}$$

Influence function of MDPDE for GEV

Theorem. Let $\theta_0 := (\mu_0, \sigma_0, \xi_0)$ be the target parameters. Suppose $\xi_0 > -(1+\alpha)/(2+\alpha)$, for fixed $\alpha > 0$. Then, the influence function of the MDPDE is given by

$$IF_\alpha(x, \theta_0) = J_\alpha^{-1}(\theta_0)\left[S(x;\theta_0)f^\alpha(x;\theta_0) - U_\alpha(\theta_0)\right],$$

and is bounded for $\alpha > 0$.

- Advantage over the MLE which has unbounded influence function.

- Decomposition:

$$IF_\alpha(x, \theta_0) = \Big(IF_{\alpha,\mu}(x, \theta_0), IF_{\alpha,\sigma}(x, \theta_0), IF_{\alpha,\xi}(x, \theta_0)\Big)^\top.$$

Illustration influence function

$\xi_0 = -0.3, \sigma_0 = 1, \mu_0 = 0$

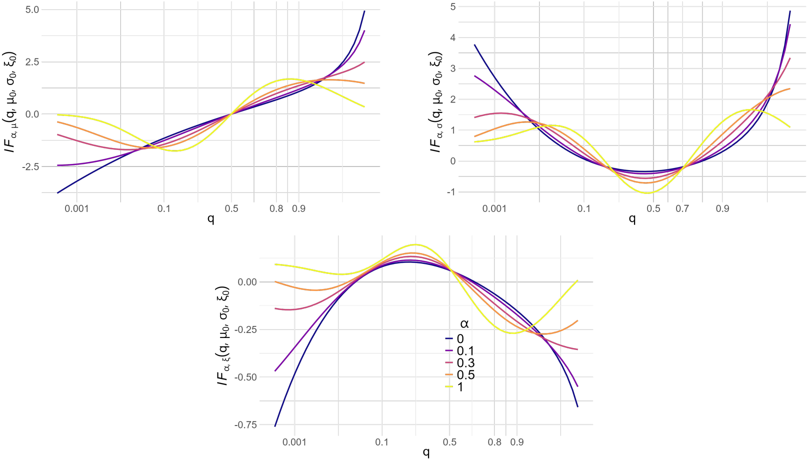


Figure: y-axis : componentwise MDPD influence functions.
x-axis : quantile level at which the functions are evaluated.

Illustration influence function

$\xi_0 = 0.3, \sigma_0 = 1, \mu_0 = 0$

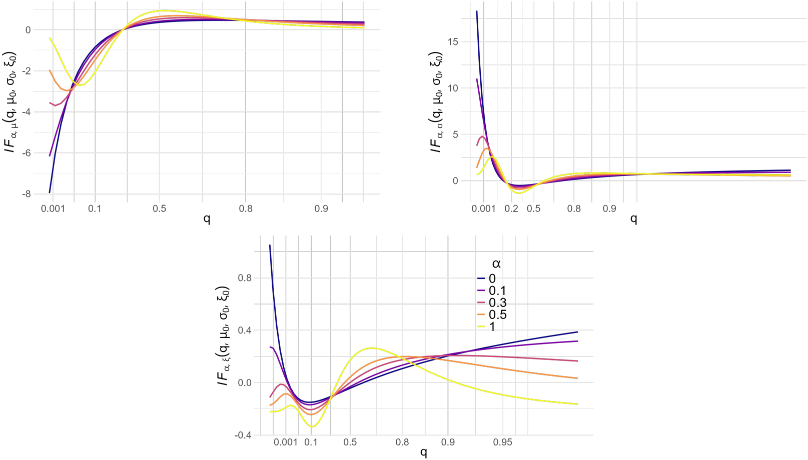


Figure: y-axis : componentwise MDPD influence functions.
x-axis : quantile level at which the functions are evaluated.