

# Extrapolating into the Extremes with Minimum Distance Estimation

Alexis Boulin<sup>\*</sup> Erik Haufs<sup>†</sup>

November 26, 2025

Understanding complex dependencies and extrapolating beyond observations are key challenges in modeling environmental space-time extremes. To address this, we introduce a simplifying approach that projects a wide range of multivariate exceedance problems onto a univariate peaks-over-threshold problem. In this framework, an estimator is computed by minimizing the  $L_2$ -distance between the empirical distribution function of the data and the theoretical distribution of the model. Asymptotic properties of this estimator are derived and validated in a simulation study. We evaluated our estimator in the EVA (2025) conference Data Challenge as part of Team Bochum's submission. The challenge provided precipitation data from four runs of LENS2, an ensemble of long-term weather simulations, on a  $5 \times 5$  grid of locations centered at the grid point closest to Asheville, NC. Our estimator achieved a top-three rank in two of six competitive categories and won the overall preliminary challenge against ten competing teams.

*Keywords.* Extreme value analysis, Peaks-over-threshold, Minimum distance estimation, Spatio-temporal extremes, Robust inference, Environmental statistics

*MSC subject classifications.* Primary 62G32, 62G05,  
Secondary 62F12, 62P12.

## Contents

<b>1. Introduction</b>	<b>2</b>
<b>2. Methods</b>	<b>3</b>
2.1. From events to univariate GPDs . . . . .	4
2.2. Peaks-over-Threshold Framework . . . . .	5

---

<sup>\*</sup>Ruhr-Universität Bochum, Fakultät für Mathematik. Email: [alexis.boulin@rub.de](mailto:alexis.boulin@rub.de)

<sup>†</sup>Ruhr-Universität Bochum, Fakultät für Mathematik. Email: [erik.haufs@rub.de](mailto:erik.haufs@rub.de)

2.3. Minimum Distance Estimation (MDE) . . . . .	6
<b>3. Application to the Data Challenge</b>	<b>9</b>
3.1. Residual-based estimation of $\varsigma$ . . . . .	10
3.2. On the choice of threshold . . . . .	11
<b>4. Conclusion</b>	<b>12</b>
<b>A. Proofs for Section 2</b>	<b>16</b>
<b>B. Efficiency comparison to MLE</b>	<b>21</b>
<b>C. Additional results for the challenge</b>	<b>24</b>
<b>D. Simulation Study</b>	<b>26</b>
<b>E. Covariance Calculations with Mathematica</b>	<b>29</b>

## 1. Introduction

Extreme value statistics is frequently concerned with the problem of estimating some extreme occurrence probability. Classical models are often limited to either a spatial *or* a temporal analysis of an extreme event.

In environmental statistics, however, the interest lies in the spatio-temporal extent of an extreme event: for example, the severeness of a flooding is influenced by the cumulative precipitation over different time scales (hourly to several days) as well as the regional spread of extreme rain. The tasks in the EVA 2025 data challenge try to capture such complexities. All posed tasks may be phrased in the quite general setting of estimating exceedance probabilities. They have been designed in a way to incorporate multivariate methods, targeting at the simultaneous exceedance probability of precipitation at multiple distinct locations on a grid, as well as capturing the temporal dependence structure by asking for exceedances over consecutive days. A detailed description of the challenge design is provided in EVA2025-Data-Challenge-Team 2025. The data challenge consisted of a preliminary and a competition challenge, each subdivided into three target probabilities, which will be denoted by (P1) – (P3) and (C1) – (C3) in the following. The targets refer to a simulated data set of daily precipitation at 25 locations, organized into a  $5 \times 5$  grid. At each location, 165 years of daily observations are recorded; four simulation runs of this model were presented.

Our contribution may be seen as two-fold, first methodological, and secondly, theoretical. On the methodological level, we provide a strategy how to project all, up to 25-dimensional, problems into a simplifying univariate framework, which we exploit to estimate the target probabilities. Within the univariate setting, an  $L_2$ -minimum distance estimation is employed to extrapolate from less extreme threshold exceedances to the target exceedance probabilities. Latter idea emerged from a diagnostic tool frequently used in the evaluation of multivariate extreme value methods to compare theoretical

survival probabilities to empirical ones, see, e.g., Kiriliouk and Zhou 2022, Fig. 3, Li et al. 2024, Fig. 7 or Buriticá et al. 2025, Fig. 4.

Our proposed minimum distance approach directly fits a parametric model to the empirical survival probabilities and uses this model for extrapolation, similar to a peaks-over-threshold approach. Latter is widely applied throughout extreme value statistics, with GPD modeling dating back as far as Pickands 1975 and with statistical properties studied by Balkema and Haan 1974; Davison 1984; Davison and Smith 1990, among many others.

Our major theoretical contribution lies in the introduction of an  $L_2$  Minimum-Distance Estimator, for which we provide asymptotic theory as well as an exhaustive efficiency comparison to maximum likelihood estimation.

Minimum distance estimation strategies are widely used in different areas of statistics Drossos and Philippou 1980; Clarke and Heathcote 1994; Ozturk 1997, increasingly popular in situations demanding robust estimation. Notable contributions to the  $L_2$ -MDE may be attributed to, for instance, Hettmansperger et al. 1994. Even within extreme value statistics, they are a commonly applied tool, see for instance Berghaus et al. 2013; Broadwater and Chellappa 2009; Dietrich and Hüsler 1996; Yilmaz et al. 2021. In particular for the GPD, the MLE has been long criticized to yield non-robust estimates and attempts have been made to provide more robust estimation strategies, such as Jurez and Schucany 2004.

Probably closest to our proposed approach is the work of Chen et al. 2017. However, they only minimize their distance on the GPD sample and do not minimize a distance related to the whole empirical survival functions. Besides, we prove asymptotic normality and provide explicit formulae for the limiting covariance matrix.

The remainder of this paper is organized as follows. In Section 2, we formalize the problems of the data challenge and detail the construction and properties of the estimators, including two ways for the construction of normal confidence intervals. These tools are then applied to the problems of the challenge in Section 3, forming the core of this paper.

## 2. Methods

We consider the general task of estimating probabilities of rare spatio-temporal events in multivariate (precipitation) data.

Let  $(\Omega, \mathcal{A}, \mathbb{P})$  be a probability space. In this paper, all random variables are assumed to be defined on this space. Denote with  $(\xi_t^{(\ell)})_t \subset \mathbb{R}^d, t = 1, \dots, n; \ell = 1, \dots, 4$  the ( $d = 25$ -dimensional) vector of daily precipitation at day  $t$  in the simulation run  $\ell$ . The time series  $(\xi_t^{(\ell)})_t$  are regarded as independent observations of a stochastic process  $(\xi_t)_t$  with observation length  $n = 365 \cdot 165$ . The  $i^{\text{th}}$  coordinate of  $\xi_t$  shall be  $\xi_{t,i}$  and the  $i^{\text{th}}$  smallest entry of  $\xi_t$  shall be  $\xi_{t,i:d}$ . Each task of the EVA 2025 Data Challenge corresponds to estimating an expectation (target) of the form

$$T_A := \mathbb{E} \left[ \sum_{t=1}^n \mathbf{1}_{A_t} \right] = n \cdot \mathbb{P}(A_t),$$

where  $A_t = A_t((\xi_s)_{s \in \mathcal{S}(t)})$  is a generic spatio-temporal event describing the properties of  $\xi_s$  in a temporal neighborhood  $\mathcal{S}(t)$  of time  $t$ . The equality  $T_A = n \cdot \mathbb{P}(A_t)$  holds, of course, only under stationarity over  $t$ , which we want to assume throughout.

*Example 1* (Challenge events). We consider the following events: for a threshold  $q > 0$ , put

$$\begin{aligned} A_t^{(P1)}(q) &:= \left\{ \sum_{i=1}^d \xi_{t,i} > q \right\} & A_t^{(C1)}(q) &:= \left\{ \xi_{t,1:d} > q \right\} \\ A_t^{(P2)}(q) &:= \left\{ \xi_{t,3:5} > q \right\} & A_t^{(C2)}(q) &:= \left\{ \xi_{t,6:d} > q \right\} \\ A_t^{(P3)}(q) &:= \left\{ \xi_{t-1,3:5} \leq q, \xi_{t,3:5} \wedge \xi_{t+1,3:5} > q \right\} & A_t^{(C3)}(q) &:= \left\{ \xi_{t-1,3:d} \leq q, \xi_{t,3:d} \wedge \xi_{t+1,3:d} > q \right\} \end{aligned} \quad (1)$$

The events (P1) – (P3) and (C1) – (C3) denote the targets of the preliminary and competition challenge, respectively.

Our subsequent approach is threefold. First, for a given  $A_t$ , we show how to construct univariate random variables  $X_t$  with  $\mathbb{P}(A_t(q)) = \mathbb{P}(X_t > q)$ , phrasing the targets as one-dimensional threshold exceedances. Next, we employ a peaks-over-threshold approach to model high threshold exceedances via a generalized Pareto distribution. Finally, we construct a  $L_2$  Minimum Distance Estimator for  $\mathbb{P}(A_t)$ , whose asymptotic properties we leverage for inference. The remainder of this section formalizes the described steps.

## 2.1. From events to univariate GPDs

All events  $A(q)$  of the data challenge have in common that they may be expressed as a function of a univariate parameter  $q \geq 0$ , to be called *threshold*. Further, a common property is  $\mathbb{P}(A(q)) \rightarrow 0$  as  $q \rightarrow \infty$ . Contrary, the limit  $q \rightarrow 0$  may differ between the events: for P1, P2, C1 and C2, we have  $\mathbb{P}(A(q)) \rightarrow 1$  as  $q \rightarrow 0$  and, for P3 and C3,  $\mathbb{P}(A(q)) \rightarrow p \in [0, 1]$ . The qualitative difference of possible functions  $q \mapsto \mathbb{P}(A(q))$  is depicted in the left-hand side of Figure 1. During the data challenge, we are interested in modeling  $\mathbb{P}(A(q))$  for large values of  $q$ , based on the observed counts  $\#\{t : A_t(q)\}$ . Assume that  $\mathbb{P}(A(q)) \leq \mathbb{P}(A(q'))$  for sufficiently large  $q' < q$ , the counts  $\#\{t : A_t(q)\}$  are monotonically decreasing in  $q$  (at sufficiently large  $q$ ); which is not restrictive for the challenge events. Thus, it appears natural to interpret the rescaled counts as the empirical survival function of a random variable. We want to construct such a random variable next and motivate a distributional assumption for it.

Recall that  $q \mapsto \mathbb{P}(A(q))$  maps to  $[0, 1]$  with  $\mathbb{P}(A(q)) \downarrow 0$  as  $q \rightarrow \infty$  for  $q$  sufficiently large. Thus, we put

$$F(q) = 1 - \sup_{w \geq q} \mathbb{P}(A_t(w)), \quad q \geq 0 \quad (2)$$

with  $F(q) = 1 - \mathbb{P}(A_t(q))$  for  $q$  sufficiently large. See the right-hand side of Figure 1 for (qualitative) examples of  $1 - F$ . Note that  $F$  is a cumulative distribution function on  $[0, \infty)$  with a point mass of  $1 - \sup_{w \geq 0} \mathbb{P}(A_t(w))$  at  $q = 0$ .<sup>1</sup> Now introduce iid random

---

<sup>1</sup>such a distribution, also known as a zero-inflated distribution, may occur, for instance, when modeling daily precipitation sums

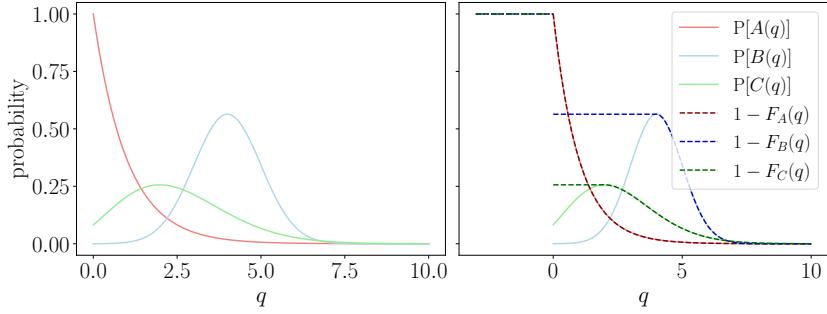


Figure 1: Left: qualitatively possible curves of  $q \mapsto \mathbb{P}(E(q))$  for different generic events  $E \in \{A, B, C\}$ . Event  $A$  qualitatively corresponds to P1, P2, C1 and C2;  $B$  (if  $p = 0$ ) and  $C$  (if  $p > 0$ ) to P3 and C3. Right: event probability curves  $q \mapsto \mathbb{P}(E(q))$  together with their corresponding survival function  $1 - F_E(q) = \sup_{w \geq q} \mathbb{P}(A_t(w))$  as constructed in Equation (2).

variables  $X_t \sim F, t = 1, \dots, n$ , and denote with  $t_1, \dots, t_k$  the time points of exceedances,  $\{t_1, \dots, t_k\} := \{t : X_t > u\}$ . The exceedances shall be  $Y_j := X_{t_j} - u, j = 1, \dots, k$ . By construction, we have for  $x := q - u > 0$ ,

$$\mathbb{P}(Y_j > x) = \frac{\mathbb{P}(X_t > q)}{\mathbb{P}(X_t > u)} = \frac{\mathbb{P}(A(q))}{\mathbb{P}(A(u))}.$$

This demonstrates how to express the target probability  $\mathbb{P}(A(q))$  via a univariate random variable  $Y$ . The term  $\mathbb{P}(X_t > u)$  may simply be approximated via its empirical probability; the term  $\mathbb{P}(Y_j > x)$  is a typical threshold exceedance probability. Threshold exceedances are usually modeled by a GPD, as the following section elaborates.

## 2.2. Peaks-over-Threshold Framework

When analyzing exceedances of a random variable  $X$  above a high threshold  $u$ , univariate extreme value theory motivates the use of Generalized Pareto Distribution (GPD) as a limiting model for excesses. The GPD cumulative distribution function is defined for any  $x \geq 0$  by

$$F_\theta(x) = 1 - S_\theta(x) = 1 - \left(1 + \gamma \frac{x}{\sigma}\right)_+^{-1/\gamma}, \quad \theta = (\gamma, \sigma) \in \Theta \subset \mathbb{R} \times (0, \infty),$$

where  $(z)_+ = \max(z, 0)$ ,  $\gamma$  is a shape parameter and  $\sigma$  is a positive scale parameter.

We now motivate why and how to model the multivariate events  $A_t$  of Example 1 with a GPD: let  $X_1, X_2, \dots, X_n \stackrel{\text{iid}}{\sim} F$  denote the random univariate variables constructed as described in the previous section. For a high threshold  $u$ , and excess level  $x \geq 0$ , define the conditional distribution of excesses:

$$F_u(x) := \mathbb{P}(X_1 - u \leq x \mid X_1 > u), \quad x \geq 0.$$

The Pickands-Balkema-de Haan theorem [Pickands 1975](#); [Balkema and Haan 1974](#) gives conditions, so that for sufficiently large  $u$ ,  $F_u(x) \approx F_\theta(x)$  where  $F_\theta$  is the GPD with shape parameter  $\gamma \in \mathbb{R}$  and  $\sigma > 0$ . In particular, exceedance probabilities admit the classical approximation

$$\mathbb{P}(X_1 > u + x) \approx \mathbb{P}(X_1 > u) \cdot S_\theta(x) \quad (3)$$

The exceedance probability  $\mathbb{P}(X_1 > u)$  can be estimated empirically by  $\hat{\mathbb{P}}(X_1 > u) = \frac{1}{n} \sum_{t=1}^n \mathbf{1}(X_t > u)$ , so it remains to focus on estimating the GPD parameters  $\theta = (\gamma, \sigma)$ .

*Example 2.* To illustrate how multivariate tail behavior leads to a univariate POT limit, we begin with a concrete example. Let  $\xi_t \sim Y$  for any  $t = 1, \dots, n$  where  $Y$  is a  $d$ -dimensional real random vector. Suppose that the  $d$ -dimensional random vector  $Y = (Y_1, \dots, Y_d)$  follows max-linear model, i.e.,  $Y_j = \max_{a=1}^K A_{ja} Z_a$ ,  $j = 1, \dots, d$  with nonnegative coefficients  $A_{ja}$  and independent  $\alpha$ -Fréchet  $Z_a$ ,  $a = 1, \dots, K$  with  $K \in \mathbb{N}_{\geq 1}$ . Consider the failure region  $C(x) = \{y : \max(y_1, \dots, y_n) > x\}$ , which is standard in environmental applications such as wind speed or extreme rainfall. Then, classical multivariate extreme value theory yields the approximation for exceedance of  $X = \max\{Y_1, \dots, Y_d\}$  for large  $x$

$$\mathbb{P}\{X > x\} \approx \nu_Y(C(x)) = \frac{1}{x^\alpha} \sum_{a=1}^K \max_{j=1, \dots, d} A_{ja}^\alpha,$$

where  $\nu_Y$  is the so-called exponent measure of the max-linear random vector  $\xi$ , see, e.g., [Kiriliouk and Zhou 2022](#), Equation (9), [Boulin and Bücher 2025](#), Lemma 2.7. Consequently, the tail of  $X$  is regularly varying with index  $\alpha$ , so exceedances over a high threshold follow a GPD limit with shape parameter  $\gamma = 1/\alpha > 0$  and scale parameter  $\sigma = \frac{1}{\alpha} \left( \sum_{a=1}^K \max_{j=1, \dots, d} A_{ja}^\alpha \right)^{1/\alpha}$ . Similar representation can be done for failure region such as  $\{\mathbf{y} : \min(y_1, \dots, y_d) > x\}$ ,  $\{\mathbf{y} : \sum_{j=1}^d v_j y_j > x\}$  where  $v_j > 0$ ,  $j = 1, \dots, d$  or, more generally, for any homogeneous transformation  $h : [0, \infty)^d \rightarrow \mathbb{R}$  of a regularly varying random vector  $\mathbf{Y}$  such that  $\nu_{\mathbf{Y}} \circ h^{-1}$  is not the null-measure using [Kulik and Soulier 2020](#), Proposition 2.1.12. This illustrates how multivariate regular variation naturally leads to a univariate GPD limit for exceedances of suitable transformations.

### 2.3. Minimum Distance Estimation (MDE)

Recall  $X_1, X_2, \dots, X_n \stackrel{\text{iid}}{\sim} F$ , fix a high threshold  $u = u_n \in \mathbb{R}$ , and define the (random) number of exceedances as  $K = K_n := \sum_{i=1}^n \mathbf{1}_{\{X_i > u\}}$ . Denote by  $\{i_1, \dots, i_K\} \subset \{1, \dots, n\}$  the indices of those observations that exceed the threshold. For each  $j = 1, \dots, K$ , we define the excesses above  $u$  by  $Y_j := X_{i_j} - u$ . Thus, put  $Y_1, \dots, Y_K \in (0, \infty)$  to denote the observed exceedances, which form the basis for GPD modeling. Define the empirical survival function to be

$$\hat{S}_K(x) := \frac{1}{K} \sum_{j=1}^K \mathbf{1}(Y_j > x).$$

To construct the MDE, we will minimize the distance functional

$$d(F, G) := \int_0^\infty [F(x) - G(x)]^2 dx,$$

for any two functions  $F, G \in L_2$ . If  $\gamma < 2$ , then  $S_\theta \in L_2$  and we can we define

$$\hat{\theta}_K^{\text{MDE}} := \arg \min_{\theta \in \Theta} J_K(\theta), \quad J_K(\theta) = d(\hat{S}_K, S_\theta).$$

By the plug-in principle, we obtain an estimator for the second term of the desired exceedance probability (3),  $\hat{S}_K^{\text{MDE}}(x) = S_{\hat{\theta}_K^{\text{MDE}}}(x)$ . To avoid technical complications with dependence within exceedances, only approximate validity of the GPD law, and  $K$  being random, we will derive asymptotic properties for

$$\hat{\theta}_k^{\text{MDE}} := \arg \min_{\theta \in \Theta} J_k(\theta), \quad J_k(\theta) = d(\hat{S}_k, S_\theta), \quad \hat{S}_k(x) := \frac{1}{k} \sum_{j=1}^k \mathbb{1}(Z_j > x),$$

with  $Z_1, \dots, Z_k \sim \text{GPD}(\theta_0)$  iid and  $k \in \mathbb{N}$  deterministic. The sample  $(Z_j)_j$  shall be thought of approximating  $(Y_j)_j$ . The results below are stated under quite restrictive conditions, but sufficient to apply the estimator to the challenge. Most importantly, see Section 3 to verify  $\gamma \in [0, 1]$ . In particular, they allow us to formulate the MDE as a  $Z$ -estimator.

*Condition 1.* The random variables  $Z_1, Z_2, \dots$  are iid and follow exactly a  $\text{GPD}(\theta_0)$  distribution, with  $\theta_0 \in \text{Interior}(\Theta)$ , where  $\Theta \subset (0, 1) \times (0, \infty)$  is a compact and convex subset.

**Lemma 1** (MDE is a  $Z$ -estimator). *Let  $(z_1, \dots, z_k) \in (0, \infty)^k$  and  $\Theta \subset (0, 1) \times (0, \infty)$  compact. Any local minimizer of the optimization problem*

$$\arg \min_{\theta \in \Theta} d\left(\frac{1}{k} \sum_{j=1}^k \mathbb{1}(z_j > x), S_\theta\right)$$

satisfies

$$\Psi_k(\theta) := \frac{1}{k} \sum_{j=1}^k \psi(z_j, \theta) = 0,$$

where, for  $x \in (0, \infty)$ ,  $\psi(x, \theta) = (\psi_\gamma(x, \theta) \quad \psi_\sigma(x, \theta))^\top$  and

$$\begin{aligned} \psi_\gamma(x, \theta) &= \frac{\sigma}{2(\gamma - 2)^2} + (\gamma - 1)^{-2}\gamma^{-2} \left[ -\gamma^2\sigma + \left(\frac{\sigma}{\sigma + \gamma x}\right)^{1/\gamma} \left\{ \gamma(\gamma\sigma + (2\gamma - 1)x) \right. \right. \\ &\quad \left. \left. - (\gamma - 1)(\sigma + \gamma x) \log\left(\frac{\gamma x}{\sigma} + 1\right) \right\} \right] \\ \psi_\sigma(x, \theta) &= -\frac{1}{2(\gamma - 2)} - (\gamma - 1)^{-1}\sigma^{-1} \left[ (\sigma + x) \left(\frac{\sigma}{\sigma + \gamma x}\right)^{1/\gamma} - \sigma \right]. \end{aligned}$$

By standard  $Z$ -estimator theory, we conclude two major results.

**Theorem 2** (Consistency). *Assume Condition 1. Then any sequence  $\hat{\theta}_k^{\text{MDE}} = \hat{\theta}_k^{\text{MDE}}(Z_1, \dots, Z_k)$  with  $J_k(\hat{\theta}_k^{\text{MDE}}) \leq J_k(\theta_0) + o_{\mathbb{P}}(1)$ , as  $k \rightarrow \infty$ ,*

$$\hat{\theta}_k^{\text{MDE}} \xrightarrow{\mathbb{P}} \theta_0.$$

Furthermore, for  $x \geq 0$ ,

$$\hat{S}_k^{\text{MDE}}(x) = S_{\hat{\theta}_k^{\text{MDE}}}(x) \xrightarrow{\mathbb{P}} S_{\theta_0}(x) := S_0(x).$$

**Theorem 3** (Asymptotic Distribution). *Assume Condition 1. Then any consistent estimator sequence  $\hat{\theta}_k^{\text{MDE}}$  with  $\Psi_k(\hat{\theta}_k^{\text{MDE}}) = o_{\mathbb{P}}(1/\sqrt{k})$ , as  $k \rightarrow \infty$*

$$\sqrt{k}(\hat{\theta}_k^{\text{MDE}} - \theta_0) \xrightarrow{\mathcal{D}} Z \sim \mathcal{N}_2(0, \Sigma_{\theta_0}),$$

with  $\Sigma_{\theta_0}$  being explicitly given in the proof and visualized in Figure 7. Second, for  $x \geq 0$ ,

$$\sqrt{k}(\hat{S}_{\text{MDE}}(x) - S_0(x)) \xrightarrow{\mathcal{D}} \mathbb{G}(x) := \nabla_{\theta} S_0(x)^{\top} Z.$$

where  $\mathbb{G}(x)$  is a mean-zero Gaussian Process with covariance kernel

$$\text{Cov}(\mathbb{G}(x), \mathbb{G}(x')) = \nabla_{\theta} S_0(x)^{\top} \Sigma_{\theta_0} \nabla_{\theta} S_0(x').$$

In particular,  $\varsigma_{\theta}^2(x) := \text{Cov}(\mathbb{G}(x), \mathbb{G}(x))$  satisfies

$$\begin{aligned} & (3(3-2\gamma)^3(\gamma-6)^2(\gamma-3)^2\gamma^4(\sigma+\gamma x)^2) \cdot \varsigma_{\theta}^2(x) \\ &= 4\left(\frac{\gamma x}{\sigma}+1\right)^{-2/\gamma} \left[ 4\gamma^2(\gamma+2)^2 \left( \gamma \left( \gamma \left( 30\gamma^2 - 266\gamma + (\gamma-2)^2(\sigma+\gamma x) \log \left( \frac{\gamma x}{\sigma} + 1 \right) + 857 \right) \right. \right. \right. \\ & \quad \left. \left. \left. - 1208 \right) + 639 \right) x^2 \cdot \left\{ - \left( \left( \gamma \left( \gamma \left( 2\gamma \left( 4\gamma^2 - 58\gamma + 243 \right) - 683 \right) + 452 \right) - 639 \right) (\gamma-2)^2 \right. \right. \\ & \quad \left. \left. \cdot (\sigma+\gamma x) \log \left( \frac{\gamma x}{\sigma} + 1 \right) \right) - 4\gamma(\gamma+2) \left( 2\gamma \left( \gamma \left( 8\gamma^2 - 62\gamma + 223 \right) - 415 \right) + 639 \right) x \right\} \right] \end{aligned}$$

The relevance in Theorem 3 lies in the explicit formula for  $\varsigma_{\theta}^2(x)$ . Utilizing the variance estimator  $\hat{\varsigma}_k^2(x) := \hat{\varsigma}_{\hat{\theta}_k^{\text{MDE}}}^2(x)$ , it allows to construct an asymptotically valid level- $\alpha$  confidence interval for  $S_{\theta}(x)$  via

$$\mathcal{C}_{\theta}(x) := [\hat{S}_{\text{MDE}}(x) \pm \sqrt{k} \cdot q_{1-\alpha/2} \cdot \hat{\varsigma}_k(x)], \quad (4)$$

where  $q_{1-\alpha/2}$  denotes the  $1 - \alpha/2$ -quantile of the standard normal distribution.

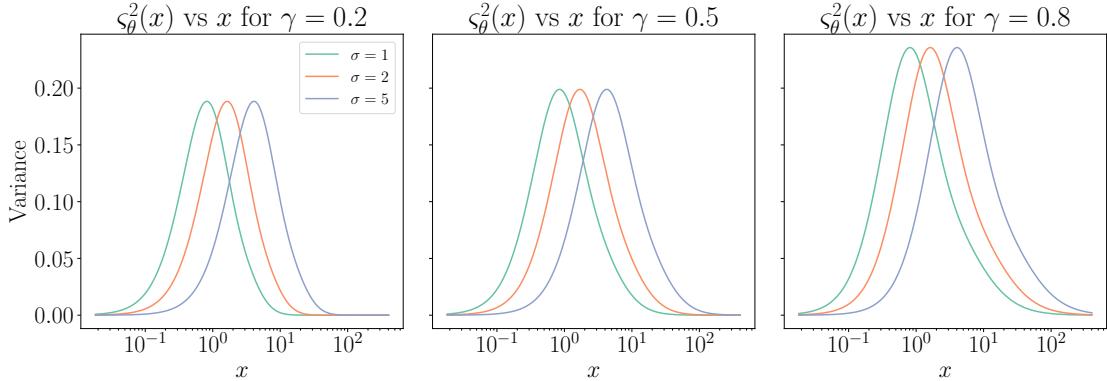


Figure 2: A depiction of the variance functions  $x \mapsto \xi_\theta^2(x)$  for different choices of  $\gamma$  and  $\sigma$ .

### 3. Application to the Data Challenge

We now apply the methodology introduced in Section 2 to the three preliminary (P1) – (P3) and three competition targets (C1) – (C3) from the EVA 2025 Data Challenge. Recall that each target corresponds to an expected number of spatio-temporal exceedances in a 25-dimensional grid over 165 years of daily observations. In particular, we are interested in the estimation of

$$A_t^{(P1)}(85), \quad A_t^{(P2)}(4.3), \quad A_t^{(P3)}(2.5), \quad A_t^{(C1)}(1.7), \quad A_t^{(C2)}(5.7), \quad A_t^{(C3)}(5),$$

with  $A_t^{(T)}(q)$  as in Equation (1). For each target, we are provided four model runs  $(\xi_t^{(\ell)})_t$ , assumingly independent copies of  $(\xi_t)_t$ . Every copy leads to events  $A_t^{(T),\ell}(q)$ , with equal probability  $\mathbb{P}(A_t^{(T),\ell}(q)) = \mathbb{P}(A_t^{(T),r}(q))$  for all  $\ell, r = 1, 2, 3, 4$ .

Following the construction of Section 2.3, we denote  $F^{(T)}(q) = 1 - \sup_{w \geq q} \mathbb{P}(A_t^{(T)}(w))$  and put  $X_t^{(T),\ell} \sim F^{(T)}$ . Further, fix a threshold  $u \in (0, \infty)$ , to be chosen later. We put  $k_\ell := |\{t \in [n] : X_t^{(T),\ell} > u\}|$  and denote the indices (time points) of exceedances in copy  $\ell$  with  $\{i_1, \dots, i_{k_\ell}\} = \{t \in [n] : X_t^{(T),\ell} > u\}$ . The exceedances themselves shall be the random variables  $Y_j^{(T),\ell} := X_{i_j}^{(T),\ell} - u$  for  $j = 1, \dots, k_\ell$ . Their survival functions are given by

$$\hat{S}_k^{(T),\ell}(x) := \frac{1}{k_\ell} \sum_{j=1}^{k_\ell} \mathbf{1}(Y_j^{(T),\ell} > x).$$

With  $Y_j^{(T),\ell}$  being independent over  $\ell$ , we can obtain a more refined estimator for  $S(x)$  by averaging over the four runs, leading to

$$\hat{S}_k^{(T)}(x) := \frac{1}{k} \sum_{\ell=1}^4 \sum_{j=1}^{k_\ell} \mathbf{1}(Y_j^{(T),\ell} > x),$$

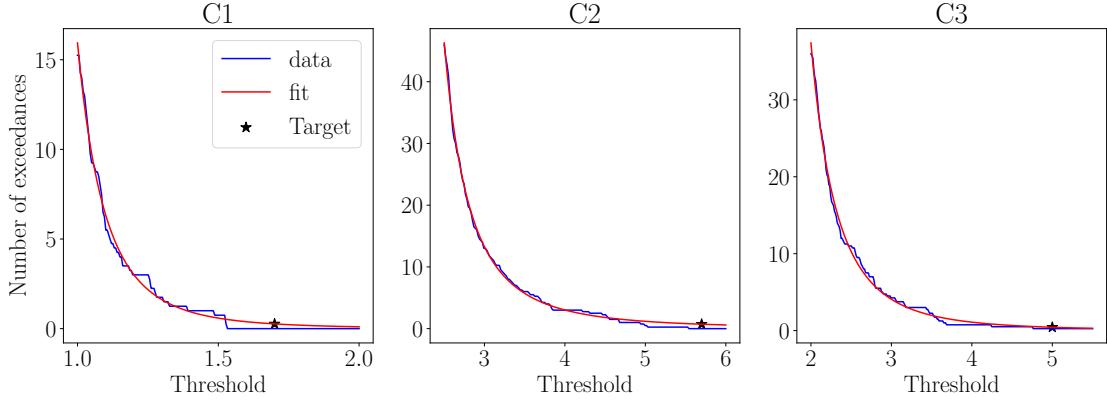


Figure 3: Average of empirical targets  $\hat{T}_A(q) = \sum_{t=1}^n \mathbb{1}(A_t(q))$  (blue) and corresponding MDE fits  $\hat{T}_A^{\text{MDE}}(q)$  (red), targets C1 – C3, defined in Example 1. The black  $\star$  denotes the target of the respective challenge.

where  $k = k_1 + k_2 + k_3 + k_4$ . The graphs of  $\hat{S}_k^{(\text{T})}(x)$  are depicted together with their MDE fits in 3 for targets C1 – C3, the preliminary targets P1 – P3 are shown in the appendix, Figure 12. In contrast to the theory in Section 2.3, we fit a three-parametric GPD,  $\vartheta = (\gamma, \mu, \sigma)$  to the survival function

$$\hat{S}_\vartheta(x) := \left(1 + \gamma \frac{x - \mu}{\sigma}\right)_+^{-1/\gamma}, \quad x \geq \mu, \quad \vartheta = (\gamma, \mu, \sigma) \in \Xi \subset \mathbb{R}^2 \times (0, \infty), \quad (5)$$

allowing for a more flexible model.

One of the practical advantages of the proposed method is its simplicity, which goes hand in hand with its computational speed: a few seconds on a standard laptop.

### 3.1. Residual-based estimation of $\varsigma$

At the time of the data challenge, Theorem 3 was not available, so that the estimator  $\hat{\varsigma}_k$  as a function of  $\hat{\theta}$ , as used in Equation (4), could not be utilized. We therefore employed a residual-based approach for the estimation of  $\varsigma_\vartheta^2(x)$ , which shall be shortly discussed. Simulation results strongly suggest to prefer the plug-in estimator  $\hat{\varsigma}_k^2(x)$ .

Consider the absolute residuals between the empirical survival function and the MDE fit

$$\hat{r}_k(x) := |\hat{S}_k(x) - \hat{S}_k^{\text{MDE}}(x)|.$$

Inspired by heteroscedastic regression, e.g. Glejser 1969, the residuals  $\hat{r}_k(x)$  were treated as approximations to the standard deviation  $\varsigma_k(x)$ . We therefore fit a parametric model to the residuals, see Figure 4, and choose, for the sake of simplicity, the one-parametric model

$$\varsigma_\phi(x) := \phi \cdot \sqrt{\hat{S}_k^{\text{MDE}}(x)}. \quad (6)$$

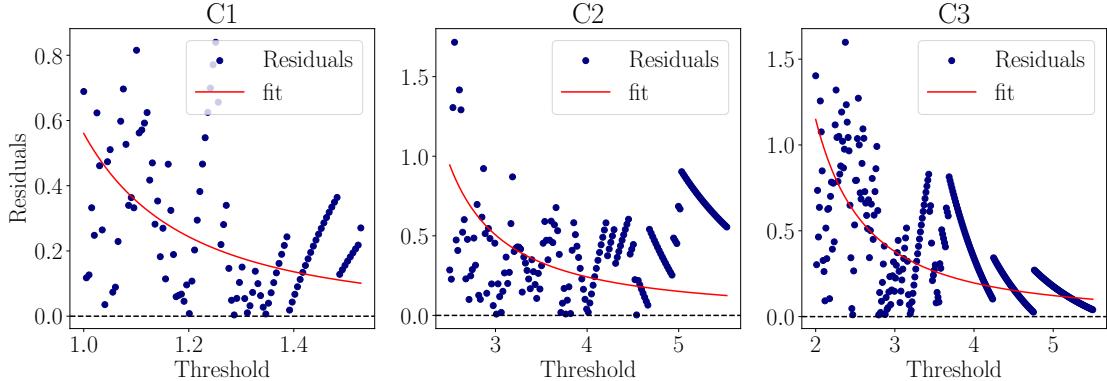


Figure 4: Residuals and fitted  $\tilde{\zeta}_k^2(x)$  for target C1 – C3. The other residual fits have been postponed to Appendix C, Figure 13.

The parameter  $\phi$  is finally determined via MDE,

$$\hat{\phi}_k := \arg \min_{\phi > 0} \int_0^\infty [\hat{r}_k(x) - \zeta_\phi(x)]^2 dx.$$

### 3.2. On the choice of threshold

All previous sections assumed a known threshold  $u = u_n$ . As it has to be chosen in practice, we investigate how the estimated value  $\tilde{S}_k^{\text{MDE}}(x)$  changes with  $u$ , the results are depicted in Figure 5 and 14. It is particularly notable that each of the six targets shows a similar behavior with varied threshold  $u$ :

- For very low thresholds, the GPD is fitted to the whole bulk of data, not only to extreme exceedances. Consequently, the GPD model does not describe the data well and we see large estimated values.
- For intermediate thresholds, a close-to-zero prediction is visible in all six targets. This may indicate that the intermediate parts of these distributions are not heavy-tailed enough to capture the actual behaviour of extreme observations adequately.
- For high thresholds, a stabilizing region may be identified, over which the estimated values do not change too much. A zoomed-in version of the “stabilized regions” is shown in Figures 6 and 15.

As it still remains open how to choose the threshold within the stabilized region adequately, we circumvented this issue in two different ways: for the preliminary challenge, we chose a threshold  $u$ , such that the resulting fit appears to describe the empirical survival probabilities adequately, based on visual diagnostics. For the competition targets, we identified a stabilized region  $[u_1, u_2]$  via visual diagnostics and averaged the predictions over this interval. The choices of thresholds are summarized in Table 1, together

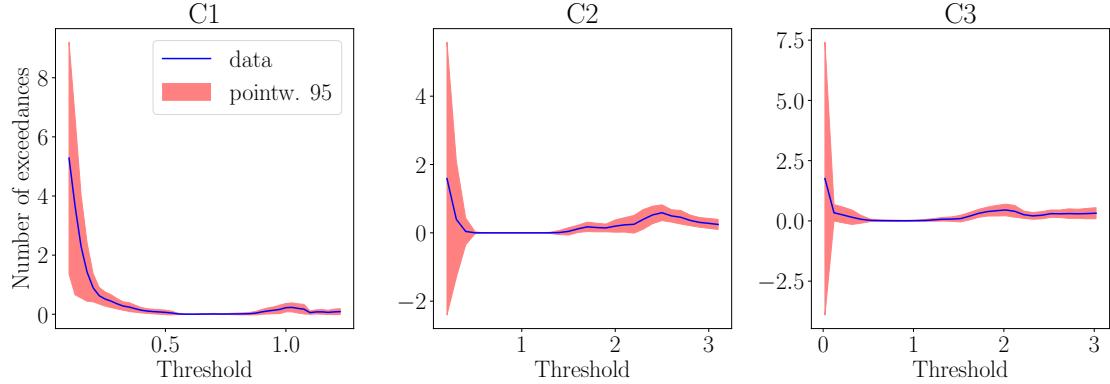


Figure 5: Development of the estimator  $\hat{T}_A^{\text{MDE}}(q)$  over the threshold  $u$  for target C1 – C3. The extreme threshold  $q = x + u$  is the target threshold and remains constant.

with the point estimates and confidence intervals. Note that the parameter values of the competition targets are only exemplary parameters – they change over the interval the estimators are being averaged over.

*Remark 1* (On the choice of threshold). There exists a vast literature reviewing tools for choosing the threshold sequence  $u_n$  in practice, motivated either by theory Fougères et al. 2015; Scarrott and MacDonald 2012; Northrop et al. 2017 or by applications Pan et al. 2022; Solari et al. 2017; Durán-Rosal et al. 2022, among many others. The discussion about which method should be preferred remains ongoing, and no clear consensus has emerged, as recently highlighted in a contributed abstract at the EVA 2025 conference Belzile et al. 2025. We do not aim to contribute further to this debate and simply advise practitioners to adopt their preferred threshold-selection method.

## 4. Conclusion

In this work, we presented a very simple approach to estimating spatio-temporal exceedance probabilities in multivariate precipitation data, as posed by the EVA 2025 Data Challenge. By reducing each task to a univariate peaks-over-threshold problem, we demonstrate how to efficiently avoid the whole field of multivariate extreme value statistics – even though all tasks targeted this particular field of study.

Within the peaks-over-threshold problem, we investigate an  $L_2$  estimator for the parameters of a generalized Pareto distribution. Despite estimation via minimum distance is not novel at all, the particular  $L_2$  distance under consideration has, to the best of our knowledge, not been studied theoretically for the GPD distribution so far. We derive asymptotic theory for the  $L_2$ -MDE by embedding it into the classical framework of  $Z$ -estimation.

Further, we provide a detailed comparison of our  $L_2$ -MDE to the classical maximum likelihood estimator in the appendix. Monte Carlo simulations indicate superiority of

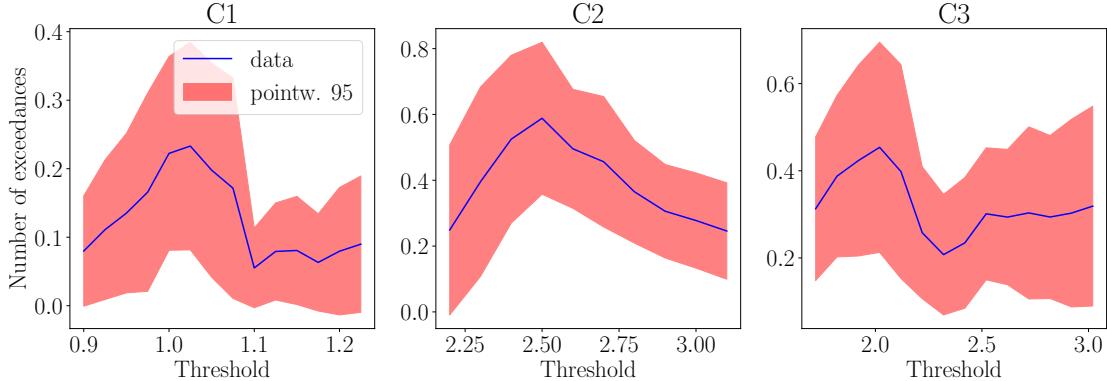


Figure 6: Zoom-in: development of the estimator  $\hat{T}_A^{\text{MDE}}(q)$  over the threshold  $u$  for target C1 – C3. The extreme threshold  $q = x + u$  is the target threshold and remains constant.

Table 1: Summary of estimation results for targets (P1) – (P3) and (C1) – (C3). The target level  $q = x + u$  is split into an excess  $x$  and a threshold  $u$ , with choices of  $u$  in third column. Consecutive columns contain the  $L_2$ -MDE estimated parameters of the  $\text{GPD}(\gamma, \mu, \sigma)$  distribution, the point estimate for the target probability and a corresponding 95% CI.

Target	$q$	$u$ $[(u_1, u_2)]$	$(\hat{\gamma}, \hat{\mu}, \hat{\sigma})$	Point estimate	95% CI
P1	85	43	(0.168, 20.71, 1.615)	0.3218	(0.1336, 0.5101)
P2	4.3	2.1	(0.199, 0.914, 0.070)	0.4160	(0.2970, 0.5351)
P3	2.5	1.2	(0.072, -0.0755, 0.125)	0.1978	(0.1080, 0.2875)
C1	1.7	[42, 70]	(0.257, 0.671, 0.012)	0.1281	(0.0186, 0.2377)
C2	5.7	[2, 4]	(0.383, 1.738, 0.020)	0.4405	(0.2064, 0.6746)
C3	5.0	[0.9, 1.3]	(0.302, 1.065, 0.034)	0.3144	(0.1291, 0.4996)

MLE over  $L_2$ -MDE, thus we cautiously recommend using former in application settings. Notwithstanding the simplicity of our approach, we managed to be competitive to stringent and complex modeling approaches from multivariate extreme value analysis, in the application to the EVA2025 data challenge.

Beyond the scope of this competition, our approach may be considered as an unconventional, but successful alternative to problems in EVT, its main idea being universally applicable to a large scope of different problems. Besides, the discussed properties of the  $L_2$ -MDE may shed new light on the usage of minimum distance estimators for peaks-over-threshold problems.

## Acknowledgments

The authors would like to express their thanks to Dan Cooley, Ben Shaby (Colorado State), Jennifer Wadsworth (Lancaster) and Emily Hector (Michigan) for organizing the data challenge of the 14<sup>th</sup> International Conference on Extreme Value Analysis 2025. Finally, valuable hints from and discussions with Axel Bücher are gratefully acknowledged.

## Declarations

### Funding

This work has been supported by the integrated project “Climate Change and Extreme Events – ClimXtreme Module B Statistics Phase II” (project B3.3, grant number 01LP2323L) funded by the German Federal Ministry of Education and Research (BMBF). Further financial support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation; Project-ID 520388526; TRR 391: Spatio-temporal Statistics for the Transition of Energy and Transport) is gratefully acknowledged. Erik Haufs is grateful for support by the Studienstiftung des deutschen Volkes and by the Ruhr University Research School, funded by Germany’s Excellence Initiative [DFG GSC 98/3]. This work used resources of the Deutsches Klimarechenzentrum (DKRZ) granted by its Scientific Steering Committee (WLA) under project ID bb1152.

### Conflict of interest

The authors do not declare any conflicts of interest.

### Data availability statement

The data supporting the findings of this study are from the EVA 2025 data challenge EVA2025-Data-Challenge-Team [2025](#), and are available on request from the the organizers.

### Author contribution

AB and EH participated together in the Data Challenge, conducting exploratory data analysis, simulations, and developed approaches. EH suggested the construction in Section 2. AB and EH wrote the manuscript.

## References

- Balkema, A. A. and Haan, L. de (1974). “Residual Life Time at Great Age”. In: *The Annals of Probability* 2.5. DOI: [10.1214/aop/1176996548](https://doi.org/10.1214/aop/1176996548).

- Belzile, L. et al. (June 2025). “Choosing the Threshold in Extreme Value Analysis”. In: *Contributed Session Abstracts, EVA 2025, Chapel Hill*. Contributed Session 5.2. Extreme Value Analysis (EVA) Conference. Chapel Hill, NC, USA.
- Berghaus, B. et al. (2013). “Minimum distance estimators of the Pickands dependence function and related tests of multivariate extreme-value dependence”. In: *J. SFdS* 154.1, pp. 116–137.
- Boulin, A. and Bücher, A. (2025). “Structured linear factor models for tail dependence”. In: *arXiv preprint arXiv:2507.16340*.
- Broadwater, J. B. and Chellappa, R. (2009). “Adaptive threshold estimation via extreme value theory”. In: *IEEE Transactions on signal processing* 58.2, pp. 490–500.
- Buriticá, G. et al. (2025). “Modeling extreme events: Univariate and multivariate data-driven approaches: EVA (2023) Conference Data Challenge: Team genEVA”. In: *Extremes* 28.1, pp. 75–99.
- Chen, P. et al. (2017). “Minimum Distance Estimation for the Generalized Pareto Distribution”. In: *Technometrics* 59.4, pp. 528–541. DOI: [10.1080/00401706.2016.1270857](https://doi.org/10.1080/00401706.2016.1270857). eprint: <https://doi.org/10.1080/00401706.2016.1270857>.
- Clarke, B. R. and Heathcote, C. R. (1994). “Robust estimation of k-component univariate normal mixtures”. In: *Annals of the Institute of Statistical Mathematics* 46.1, pp. 83–93. DOI: [10.1007/bf00773595](https://doi.org/10.1007/bf00773595).
- Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton: Princeton University Press, pp. 479–480.
- Davison, A. C. and Smith, R. L. (July 1990). “Models for Exceedances Over High Thresholds”. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 52.3, pp. 393–425. DOI: [10.1111/j.2517-6161.1990.tb01796.x](https://doi.org/10.1111/j.2517-6161.1990.tb01796.x).
- Davison, A. C. (1984). “Modelling Excesses over High Thresholds, with an Application”. In: *Statistical Extremes and Applications*. Ed. by J. T. de Oliveira. Dordrecht: Springer Netherlands, pp. 461–482. DOI: [10.1007/978-94-017-3069-3\\_34](https://doi.org/10.1007/978-94-017-3069-3_34).
- Dietrich, D. and Hüsler, J. (1996). “Minimum distance estimators in extreme value distributions”. In: *Communications in Statistics - Theory and Methods* 25.4, pp. 695–703. DOI: [10.1080/03610929608831725](https://doi.org/10.1080/03610929608831725). eprint: <https://doi.org/10.1080/03610929608831725>.
- Drossos, C. A. and Philippou, A. N. (1980). “A note on minimum distance estimates”. In: *Ann. Inst. Statist. Math.* 32.1, pp. 121–123. DOI: [10.1007/BF02480318](https://doi.org/10.1007/BF02480318).
- Durán-Rosal, A. M. et al. (2022). “A mixed distribution to fix the threshold for Peak-Over-Threshold wave height estimation”. In: *Scientific Reports* 12.1, p. 17327.
- EVA2025-Data-Challenge-Team (2025). *EVA2025 Data Challenge*. Introduction to the EVA2025 Data Challenge Special Issue.
- Fougères, A.-L. et al. (2015). “Bias correction in multivariate extremes”. In: *The Annals of Statistics* 43.2, pp. 903–934. DOI: [10.1214/14-AOS1305](https://doi.org/10.1214/14-AOS1305).
- Glejser, H. (Mar. 1969). “A New Test for Heteroskedasticity”. In: *Journal of the American Statistical Association* 64.325, pp. 316–323. DOI: [10.1080/01621459.1969.10500976](https://doi.org/10.1080/01621459.1969.10500976).

- Hettmansperger, T. et al. (Oct. 1994). “Minimum distance estimators”. In: *Journal of Statistical Planning and Inference* 41.3, pp. 291–302. DOI: [10.1016/0378-3758\(94\)90025-6](https://doi.org/10.1016/0378-3758(94)90025-6).
- Jurez, S. F. and Schucany, W. R. (Sept. 2004). “Robust and Efficient Estimation for the Generalized Pareto Distribution”. In: *Extremes* 7.3, pp. 237–251. DOI: [10.1007/s10687-005-6475-6](https://doi.org/10.1007/s10687-005-6475-6).
- Kiriliouk, A. and Zhou, C. (2022). “Estimating probabilities of multivariate failure sets based on pairwise tail dependence coefficients”. In: *arXiv preprint arXiv:2210.12618*.
- Kulik, R. and Soulier, P. (2020). *Heavy-Tailed Time Series*. Springer Series in Operations Research and Financial Engineering. New York, NY: Springer.
- Li, M. et al. (Oct. 2024). “A wee exploration of techniques for risk assessments of extreme events: EVA (2023) conference data challenge: wee extremes group”. In: *Extremes* 28.1, pp. 101–121. DOI: [10.1007/s10687-024-00500-5](https://doi.org/10.1007/s10687-024-00500-5).
- Northrop, P. J. et al. (2017). “Cross-validatory extreme value threshold selection and uncertainty with application to ocean storm severity”. In: *Journal of the Royal Statistical Society Series C: Applied Statistics* 66.1, pp. 93–120.
- Ozturk, O. (June 1997). “Generalised weighted Cramer-von Mises distance estimators”. In: *Biometrika* 84.2, pp. 283–294. DOI: [10.1093/biomet/84.2.283](https://doi.org/10.1093/biomet/84.2.283).
- Pan, X. et al. (2022). “Peaks-over-threshold model in flood frequency analysis: a scoping review”. In: *Stochastic Environmental Research and Risk Assessment* 36.9, pp. 2419–2435.
- Pickands, J. (1975). “Statistical Inference Using Extreme Order Statistics”. In: *The Annals of Statistics* 3.1. DOI: [10.1214/aos/1176343003](https://doi.org/10.1214/aos/1176343003).
- Scarrott, C. and MacDonald, A. (2012). “A review of extreme value threshold estimation and uncertainty quantification”. In: *REVSTAT-Statistical journal* 10.1, pp. 33–60.
- Solari, S. et al. (2017). “Peaks Over Threshold (POT): A methodology for automatic threshold estimation using goodness of fit p-value”. In: *Water Resources Research* 53.4, pp. 2833–2849.
- Vaart, A. W. v. d. (Oct. 1998). *Asymptotic Statistics*. DOI: [10.1017/cbo9780511802256](https://doi.org/10.1017/cbo9780511802256).
- Wolfram Research, Inc. (n.d.). *Mathematica, Version 14.1*. Champaign, IL, 2024.
- Yilmaz, A. et al. (2021). “Comparison of different estimation methods for extreme value distribution”. In: *Journal of Applied Statistics* 48.13–15. PMID: 35707070, pp. 2259–2284. DOI: [10.1080/02664763.2021.1940109](https://doi.org/10.1080/02664763.2021.1940109).

## A. Proofs for Section 2

*Proof of Lemma 1.* Recall that the gradient  $\nabla J_k$  vanishes for any local minimizer. Thus, we will consecutively calculate the gradient. Denote by  $f_\theta(x)$  the density of a GPD( $\theta$ ), i.e.,

$$f_\theta(x) = \frac{1}{\sigma} \left(1 + \frac{\gamma}{\sigma}x\right)^{-\frac{1}{\gamma}-1},$$

For any  $x \geq 0$ , the partial derivatives of  $F_\theta$  are given by:

$$\begin{aligned}\frac{\partial}{\partial \gamma} F_\theta(x) &= -\left(1 + \frac{\gamma}{\sigma}x\right)^{-1/\gamma} \left(\frac{\log\left(1 + \frac{\gamma x}{\sigma}\right)}{\gamma^2} - \frac{\frac{1}{\gamma\sigma}x}{1 + \frac{\gamma}{\sigma}x}\right) \\ &= -\frac{1}{\gamma^2\sigma} f_\theta(x) \left((1 + \gamma\frac{x}{\sigma})(\log\left(1 + \gamma\frac{x}{\sigma}\right) - 1)\right), \\ \frac{\partial}{\partial \sigma} F_\theta(x) &= -\frac{x}{\sigma^2} \left(1 + \frac{\gamma}{\sigma}x\right)^{-1/\gamma-1} = -\frac{1}{\sigma} x f_\theta(x).\end{aligned}$$

To compute the gradient of  $J_k$ , we want to swap the integral and the derivative sign. Since  $\Theta$  is compact, there exist constants  $0 < \gamma_{\min} < \gamma < \gamma_{\max} < 1$  and  $0 < \sigma_{\min} < \sigma < \sigma_{\max} < \infty$  such that the rectangle  $(\gamma_{\min}, \gamma_{\max}) \times (\sigma_{\min}, \sigma_{\max})$  contains  $(\gamma, \sigma)$  and lies within a neighborhood of  $\Theta$ . For the partial derivative with respect to  $\gamma$ , observe that for any  $\theta \in \Theta$ ,

$$\left| \frac{\partial}{\partial \gamma} F_\theta(x) \right| \leq g_1(x),$$

with,

$$g_1(x) = \left(1 + \frac{\gamma_{\max}}{\sigma_{\max}}x\right)^{-1/\gamma_{\max}} \left(\frac{\log\left(1 + \frac{\gamma_{\max}}{\sigma_{\min}}x\right)}{\gamma_{\min}^2} + \frac{\frac{x}{\gamma_{\min}\sigma_{\min}}}{1 + \frac{\gamma_{\min}}{\sigma_{\max}}x}\right),$$

where we use that  $\gamma \mapsto (1 + \gamma x/\sigma)^{-1/\gamma}$  is non-decreasing and  $\sigma > 0$ . This can be seen by defining

$$A(\gamma, \sigma, x) := \left(1 + \frac{\gamma x}{\sigma}\right)^{-1/\gamma}, \quad B(\gamma, \sigma, x) := \frac{\log\left(1 + \frac{\gamma x}{\sigma}\right)}{\gamma^2} - \frac{\frac{x}{\gamma\sigma}}{1 + \frac{\gamma x}{\sigma}}.$$

Then:

$$\left| \frac{\partial}{\partial \gamma} F_\theta(x) \right| = |A(\gamma, \sigma, x) \cdot B(\gamma, \sigma, x)| \leq A(\gamma, \sigma, x) \cdot |B(\gamma, \sigma, x)|.$$

We now bound both terms. Due to  $\gamma \mapsto (1 + \gamma x/\sigma)^{-1/\gamma}$  being non-decreasing,

$$A(\gamma, \sigma, x) \leq \left(1 + \frac{\gamma_{\max}}{\sigma_{\max}}x\right)^{-1/\gamma_{\max}}.$$

Second, for  $B(\gamma, \sigma, x)$ , observe that

$$B(\gamma, \sigma, x) = \frac{1}{\gamma^2} \phi\left(1 + \frac{\gamma}{\sigma}x\right),$$

where

$$\phi: (0, \infty) \rightarrow (0, \infty), \quad x \mapsto \phi(x) = \log(x) - 1 + \frac{1}{x}.$$

Since  $\phi(\cdot)$  is increasing on  $(1, \infty)$  we obtain by noticing that  $(1 + \gamma x/\sigma) \leq (1 + \gamma_{\max} x/\sigma_{\min})$ :

$$|B(\gamma, \sigma, x)| \leq \frac{\log\left(1 + \gamma_{\max} \cdot x/\sigma_{\min}\right)}{\gamma_{\min}^2} + \frac{x/(x/\sigma_{\min})}{1 + \gamma_{\min} \cdot x/\sigma_{\max}}.$$

Combining the bounds, we conclude:

$$\left| \frac{\partial}{\partial \gamma} F_\theta(x) \right| \leq g_1(x),$$

where:

$$g_1(x) := \left( 1 + \frac{\gamma_{\max}}{\sigma_{\max}} x \right)^{-1/\gamma_{\max}} \left( \frac{\log \left( 1 + \frac{\gamma_{\max}}{\sigma_{\min}} x \right)}{\gamma_{\min}^2} + \frac{\frac{x}{\gamma_{\min} \sigma_{\min}}}{1 + \frac{\gamma_{\min}}{\sigma_{\max}} x} \right).$$

We now show that  $g_1(x)$  is integrable. Indeed

$$\int_0^\infty \left( 1 + \frac{\gamma_{\max}}{\sigma_{\max}} x \right)^{-1/\gamma_{\max}} \log \left( 1 + \frac{\gamma_{\max}}{\sigma_{\min}} x \right) dx = \frac{\sigma_{\max} \gamma_{\max}}{(\gamma_{\max} - 1)^2} < \infty,$$

where  $0 < \gamma_{\max} < 1$ . Also,

$$\int_0^\infty \left( 1 + \frac{\gamma_{\max}}{\sigma_{\max}} x \right)^{-1/\gamma_{\max}} \left( \frac{\frac{x}{\gamma_{\min} \sigma_{\min}}}{1 + \frac{\gamma_{\min}}{\sigma_{\max}} x} \right) dx < \infty.$$

Same arguments applies for  $\frac{\partial}{\partial \sigma} F_\theta$ , therefore, by Lebesgue's theorem, we may interchange integration and differentiation and obtain  $\hat{\theta}_k^{\text{MDE}}$  as the (possibly not unique) root of

$$\partial_\theta J_k(\theta) = \int_0^\infty \frac{\partial}{\partial \theta} [\hat{F}_k(x) - F_\theta(x)]^2 dx$$

We conclude

$$\psi(x, \theta) = \int_0^x \left\{ \frac{\partial}{\partial \theta} F_\theta(y) \right\} dy - \int_0^\infty \int_0^u \left\{ \frac{\partial}{\partial \theta} F_\theta(y) \right\} dy dF_\theta(u) \quad (7)$$

by partial integration in analogy to Clarke and Heathcote 1994. The assertion follows from evaluating the appearing integrals, see also Appendix E: the inner integrals are

$$\begin{aligned} \int_0^x \left\{ \frac{\partial}{\partial \gamma} F_\theta(y) \right\} dy &= \frac{\left( \frac{\sigma}{\sigma + \gamma x} \right)^{1/\gamma} (\gamma(\gamma\sigma + (2\gamma - 1)x) - (\gamma - 1)(\sigma + \gamma x) \log \left( \frac{\gamma x}{\sigma} + 1 \right)) - \gamma^2 \sigma}{(\gamma - 1)^2 \gamma^2}, \\ \int_0^x \left\{ \frac{\partial}{\partial \sigma} F_\theta(y) \right\} dy &= -\frac{(\sigma + x) \left( \frac{\sigma}{\sigma + \gamma x} \right)^{1/\gamma} - \sigma}{(\gamma - 1)\sigma}, \end{aligned}$$

whereas the outer integrals yield

$$\begin{aligned} \int_0^\infty \int_0^u \left\{ \frac{\partial}{\partial \gamma} F_\theta(y) \right\} dy dF_\theta(u) &= -\frac{\sigma}{2(\gamma - 2)^2}, \\ \int_0^\infty \int_0^u \left\{ \frac{\partial}{\partial \sigma} F_\theta(y) \right\} dy dF_\theta(u) &= \frac{1}{2(\gamma - 2)}. \end{aligned}$$

□

*Proof of Theorem 2.* This is a direct consequence of Vaart 1998, Theorem 5.7, it remains to verify its conditions. First, we need to show that

$$\sup_{\theta \in \Theta} \|J_k(\theta) - J(\theta)\| \xrightarrow{\mathbb{P}} 0.$$

For this, we have by Cauchy-Schwartz and the triangle inequality

$$\begin{aligned} |J_k(\theta) - J(\theta)| &= \left| \int_0^\infty (\hat{S}_k(x) - S_\theta(x))^2 - (S_{\theta_0}(x) - S_\theta(x))^2 dx \right| \\ &= \left| \int_0^\infty (\hat{S}_k(x) - S_{\theta_0}(x)) [\hat{S}_k(x) + S_{\theta_0}(x) - 2S_\theta(x)] dx \right| \\ &= \left\langle \hat{S}_k - S_{\theta_0}, \hat{S}_k + S_{\theta_0} - 2S_\theta \right\rangle_{L_2}^2 \\ &\leq \|\hat{S}_k - S_{\theta_0}\|_{L_2}^2 \cdot \|\hat{S}_k - S_{\theta_0} + 2[S_{\theta_0} - S_\theta]\|_{L_2}^2 \\ &\leq \|\hat{S}_k - S_{\theta_0}\|_{L_2}^2 \cdot \left\{ \|\hat{S}_k - S_{\theta_0}\|_{L_2} + 2\|S_{\theta_0} - S_\theta\|_{L_2} \right\}^2. \end{aligned}$$

Consequently,

$$\sup_{\theta \in \Theta} |J_k(\theta) - J(\theta)| \leq \|\hat{S}_k - S_{\theta_0}\|_{L_2}^2 \cdot \left\{ \|\hat{S}_k - S_{\theta_0}\|_{L_2} + 2 \sup_{\theta \in \Theta} \|S_{\theta_0} - S_\theta(x)\|_{L_2} \right\}^2.$$

For fixed  $\theta_0 \in \Theta$ , we have

$$\sup_{\theta \in \Theta} \|S_{\theta_0} - S_\theta\|_{L_2}^2 = \sup_{\theta \in \Theta} \int_0^\infty [S_{\theta_0}(x) - S_\theta(x)]^2 dx =: C_{\theta_0} < \infty,$$

where the latter holds due to the compactness of  $\Theta \subset (0, 1) \times (0, \infty)$ . Now, by Fubini,

$$\begin{aligned} \mathbb{E} \left[ \|\hat{S}_k - S_{\theta_0}\|_{L_2}^2 \right] &= \int_0^\infty \mathbb{E} \left[ \left[ \frac{1}{k} \sum_{j=1}^k \mathbf{1}(Z_j > x) - S_{\theta_0}(x) \right]^2 \right] dx = \int_0^\infty \text{Var} \left[ \frac{1}{k} \sum_{j=1}^k \mathbf{1}(Z_j > x) \right] dx \\ &= \frac{1}{k} \int_0^\infty S_{\theta_0}(x)[1 - S_{\theta_0}(x)] dx =: \frac{V_{\theta_0}}{k} < \infty, \end{aligned}$$

as  $\hat{S}_k(x)$  is a rescaled  $\text{Bin}(k, S_{\theta_0}(x))$  random variable, pointwise. The latter  $V_{\theta_0} < \infty$  again holds due to  $\Theta \subset (0, 1) \times (0, \infty)$ . By Chebyshev's inequality,

$$\|\hat{S}_k - S_{\theta_0}\|_{L_2}^2 = o_{\mathbb{P}}(1).$$

We conclude

$$\sup_{\theta \in \Theta} |J_k(\theta) - J(\theta)| \leq o_{\mathbb{P}}(1) \cdot \{o_{\mathbb{P}}(1) + 2\sqrt{C_{\theta_0}}\}^2 = o_{\mathbb{P}}(1),$$

so that the claimed convergence in probability is indeed fulfilled. It remains to check the second condition

$$\inf_{\theta: d(\theta, \theta_0) \geq \epsilon} \|J(\theta)\| > 0 = \|J(\theta_0)\|,$$

which directly follows from the identifiability of the GPD.

We conclude the second assertion by applying the continuous mapping theorem.  $\square$

*Proof of Theorem 3.* We apply Vaart 1998, Theorem 5.21. Having established Lemma 1, the well-behaved derivatives of  $\psi$  and the strong assumptions on  $\Theta$  make most conditions easy to verify, see sufficient conditions in Vaart 1998, Theorem 5.41. The only condition requiring major effort is the boundedness of the second-order partial derivatives of  $\psi$  in a neighborhood around  $\theta_0$ , which we show later. It remains to calculate the limiting covariance matrix  $\Sigma_{\theta_0}$ . Following Vaart 1998, we have  $\Sigma_{\theta} = U_{\theta}^{-1}V_{\theta}(U_{\theta}^{\top})^{-1}$  with

$$U_{\theta} = \mathbb{E} \begin{bmatrix} \partial_{\gamma}\psi_{\gamma}(Y, \theta) & \partial_{\sigma}\psi_{\sigma}(Y, \theta) \\ \partial_{\sigma}\psi_{\gamma}(Y, \theta) & \partial_{\sigma}\psi_{\sigma}(Y, \theta) \end{bmatrix}, \quad V_{\theta} = \mathbb{E} \begin{bmatrix} \psi_{\gamma}(Y, \theta)^2 & \psi_{\gamma}(Y, \theta)\psi_{\sigma}(Y, \theta) \\ \psi_{\gamma}(Y, \theta)\psi_{\sigma}(Y, \theta) & \psi_{\sigma}(Y, \theta)^2 \end{bmatrix}.$$

Evaluating these matrices, we find

$$U_{\theta} = \begin{pmatrix} \frac{(\gamma-6)\sigma}{2(\gamma-2)^3(\gamma+2)} & \frac{6-\gamma}{4(\gamma-2)^2(\gamma+2)} \\ \frac{6-\gamma}{4(\gamma-2)^2(\gamma+2)} & \frac{1}{4\sigma-\gamma^2\sigma} \end{pmatrix}.$$

and

$$V_{\theta} = \begin{pmatrix} \frac{(8\gamma^5-148\gamma^4+918\gamma^3-2587\gamma^2+3416\gamma-1719)\sigma^2}{12(\gamma-3)^2(\gamma-2)^4(2\gamma-3)^3} & \frac{(\gamma(\gamma(-4(\gamma-15)\gamma-285)+548)-369)\sigma}{12(\gamma-2)^3(2\gamma^2-9\gamma+9)^2} \\ \frac{(\gamma(\gamma(-4(\gamma-15)\gamma-285)+548)-369)\sigma}{12(\gamma-2)^3(2\gamma^2-9\gamma+9)^2} & \frac{\gamma(2\gamma-17)+29}{12(\gamma-3)(\gamma-2)^2(2\gamma-3)} \end{pmatrix}.$$

Concluding,

$$\begin{aligned} \Sigma_{\theta} &= U_{\theta}^{-1}V_{\theta}(U_{\theta}^{\top})^{-1} \\ &= \begin{pmatrix} \frac{4(\gamma-2)^4(\gamma(\gamma(2\gamma(4\gamma^2-58\gamma+243)-683)+452)-639)}{3(\gamma-6)^2(\gamma-3)^2(2\gamma-3)^3} & \frac{4(\gamma-2)^2(\gamma(\gamma(2\gamma(4\gamma^2-50\gamma+207)-791)+778)-387)\sigma}{3(\gamma-6)(\gamma-3)^2(2\gamma-3)^3} \\ \frac{4(\gamma-2)^2(\gamma(\gamma(2\gamma(4\gamma^2-50\gamma+207)-791)+778)-387)\sigma}{3(\gamma-6)(\gamma-3)^2(2\gamma-3)^3} & \frac{4(\gamma(\gamma(8\gamma^3-84\gamma^2+374\gamma-843)+944)-423)\sigma^2}{3(\gamma-3)^2(2\gamma-3)^3} \end{pmatrix} \end{aligned} \quad (8)$$

Details of this calculation have been performed in **Mathematica** Wolfram Research, Inc. n.d., which is elaborated in Appendix E. The entries of  $\Sigma_{\theta}$  for selected values of  $\theta$  are depicted in Figure 7.

It remains to show the second assertion, to which we apply the delta method. As  $\nabla_{\theta}S_0(x) \neq 0$  for fixed  $x$ , the convergence

$$\sqrt{k}(\hat{S}_{\text{MDE}}(x) - S_0(x)) \xrightarrow{\mathcal{D}} \mathbb{G}(x) := \nabla_{\theta}S_0(x)^{\top}Z$$

obviously holds. The covariance structure of  $\mathbb{G}(x)$  is a direct consequence. Finally, the explicit structure of  $\varsigma_{\theta}^2(x)$  immediately follows from evaluating  $\varsigma_{\theta}^2(x) = \nabla_{\theta}F_{\theta}(x)^{\top}\Sigma_{\theta}\nabla_{\theta}F_{\theta}(x)$ .

*Proof of the boundedness of the second-order partial derivatives.* Within the interior of  $\Theta$ , the second order derivatives are continuous in a neighborhood around  $\theta_0$ . It therefore

suffices to show the integrability of the second-order partial derivatives at  $\theta_0$ . We have

$$\begin{aligned} \int_0^\infty \frac{\partial^2}{\partial \sigma^2} \psi(x, \theta) dF_\theta(x) &= -\frac{1}{(\gamma^2 - 4)\sigma} \\ \int_0^\infty \frac{\partial^2}{\partial \sigma \partial \gamma} \psi(x, \theta) dF_\theta(x) &= -\frac{G_{3,3}^{3,2}\left(1 \mid \begin{array}{c} -2, -\frac{2}{\gamma}, -1 \\ -2, -2, 0 \end{array}\right)}{(\gamma - 1)\gamma^4 \Gamma(1 + 2/\gamma)} - \frac{(\gamma^3 - 6\gamma^2 + 4\gamma - 8)}{4(\gamma - 1)\gamma(\gamma - 2)^2(\gamma + 2)} \\ \int_0^\infty \frac{\partial^2}{\partial \gamma^2} \psi(x, \theta) dF_\theta(x) &= \frac{2(\gamma^2 + \gamma - 1)}{(\gamma - 1)^2 \gamma^4 \Gamma(2/\gamma)} G_{3,3}^{3,2}\left(1 \mid \begin{array}{c} -2, -\frac{2}{\gamma}, -1 \\ -2, -2, 0 \end{array}\right) \sigma \\ &\quad + \frac{(2\gamma^5 - 11\gamma^4 + 4\gamma^3 + 10\gamma^2 + 20\gamma - 16)}{2(\gamma - 1)^2 \gamma^2 (\gamma - 2)^3 (\gamma + 2)} \sigma, \end{aligned}$$

where  $G_{3,3}^{2,3}$  denotes the Meijer-G function. As all previous expressions are finite for all  $\theta \in \Theta$ , we conclude the assertion.  $\square$

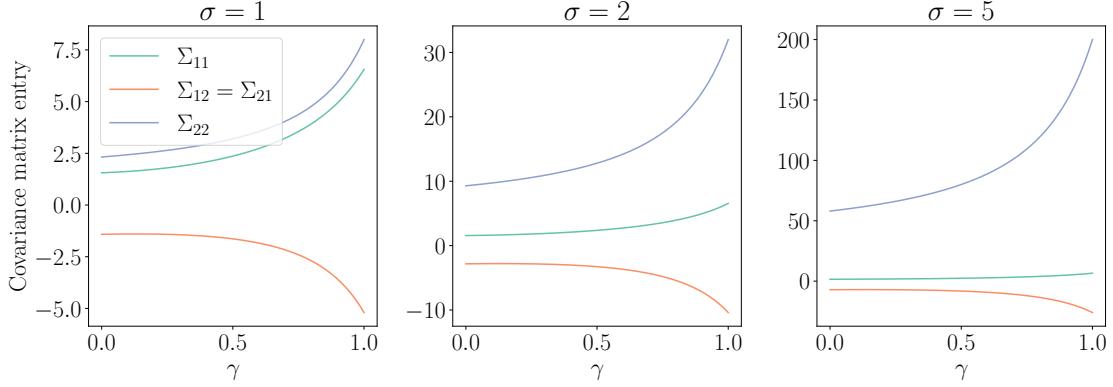


Figure 7: Entries of the covariance matrix  $\gamma \mapsto \Sigma_\theta$  for  $\sigma = 1, 2, 5$ , corresponding to the display (8). The absolute values  $(\Sigma_\theta)_{ij}$  grow with  $\gamma$  and  $\sigma$ .

## B. Efficiency comparison to MLE

A common estimator of the GPD parameters in the peaks-over-threshold method is the maximum-likelihood estimator (MLE). We therefore treat this well-known estimator as a benchmark to compare our MDE against. By standard assumptions, the MLE  $\hat{\theta}_k^{\text{MLE}}$  is asymptotically normal with limiting covariance being equal to the inverse Fisher information,

$$\sqrt{k}(\hat{\theta}_k^{\text{MLE}} - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_{\theta_0}^{\text{MLE}})$$

with

$$\Sigma_{\theta}^{\text{MLE}} = \begin{pmatrix} (\gamma + 1)^2 & -((\gamma + 1)\sigma) \\ -((\gamma + 1)\sigma) & 2(\gamma + 1)\sigma^2 \end{pmatrix}.$$

Consequently, evaluating  $\varsigma_\theta^{2,\text{MLE}}(x) = \nabla_\theta F_\theta(x)^\top \Sigma_\theta \nabla_\theta F_\theta(x)$ , we see

$$\begin{aligned} \gamma^4(\sigma + \gamma x)^2 \varsigma_\theta^{2,\text{MLE}}(x) &= (\gamma + 1) \left( \frac{\gamma x}{\sigma} + 1 \right)^{-2/\gamma} \left[ (\gamma + 1)(2\gamma + 1)\gamma^2 x^2 + (\sigma + \gamma x) \log \left( \frac{\gamma x}{\sigma} + 1 \right) \right. \\ &\quad \left. \cdot \left\{ ((\gamma + 1)(\sigma + \gamma x) \log \left( \frac{\gamma x}{\sigma} + 1 \right) - 2\gamma(2\gamma + 1)x \right\} \right]. \end{aligned}$$

By the Cramér-Rao-bound Cramér 1946, Pages 479-480, we cannot expect to achieve  $\varsigma_\theta^2(x) < \varsigma_\theta^{2,\text{MLE}}(x)$ , but we can nevertheless compare these estimators in terms of asymptotic variance. Investigating Figure 9, we observe only small differences in limiting variance between MLE and MDE. This is more illustrative in the relative efficiency  $\varsigma_\theta^2(x)/\varsigma_\theta^{2,\text{MLE}}(x)$ , which appears to be well below 2 for all choices of  $\gamma$ , see Figure 10. Of particular interest is the limiting behavior  $\lim_{x \rightarrow 0} \varsigma_\theta^{2,\text{MLE}}(x)/\varsigma_\theta^2(x)$  and  $\lim_{x \rightarrow \infty} \varsigma_\theta^{2,\text{MLE}}(x)/\varsigma_\theta^2(x)$ , as

$$\begin{aligned} \lim_{x \rightarrow 0} \varsigma_\theta^{2,\text{MLE}}(x)/\varsigma_\theta^2(x) &= \frac{2(8\gamma^5 - 84\gamma^4 + 374\gamma^3 - 843\gamma^2 + 944\gamma - 423)}{3(\gamma - 3)^2(\gamma + 1)(2\gamma - 3)^3} \\ \lim_{x \rightarrow \infty} \varsigma_\theta^{2,\text{MLE}}(x)/\varsigma_\theta^2(x) &= -\frac{4(\gamma - 2)^4 (\gamma (\gamma (2\gamma (4\gamma^2 - 58\gamma + 243) - 683) + 452) - 639)}{3(3 - 2\gamma)^3(\gamma - 6)^2(\gamma - 3)^2(\gamma + 1)^2}. \end{aligned}$$

Notably, these limits are independent from  $\sigma$  and never exceed 2 for all choices of  $\gamma$ , as it is visualized in Figure 11.

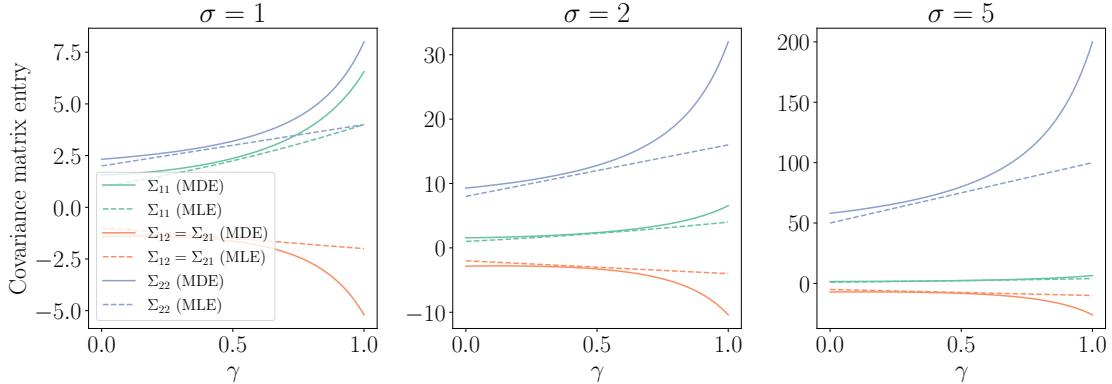


Figure 8: Entries of the limiting covariance matrices  $\Sigma_\theta^{\text{MLE}}$  and  $\Sigma_\theta$  as a function of  $\gamma$ , shown for  $\sigma = 1, 2, 5$  (from left to right). The diagonal entries of  $\Sigma_\theta^{\text{MLE}}$  are below those of  $\Sigma_\theta$ .

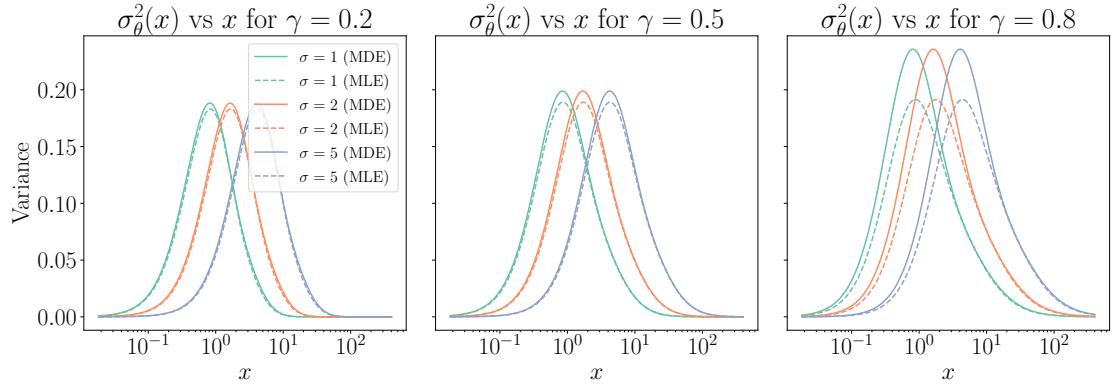


Figure 9: Variances of the survival function,  $\varsigma_\theta^2(x)$ ,  $\varsigma_\theta^{2,\text{MLE}}(x)$ . The variances approach zero for  $x \rightarrow \infty$  and  $x \rightarrow 0$ , with their peak depending on  $\gamma$  and  $\sigma$ .

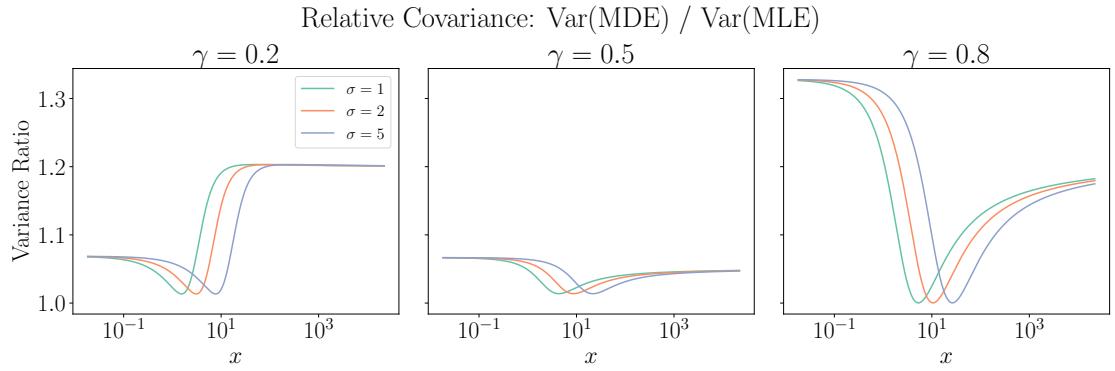


Figure 10: Relative variances of the survival function,  $\varsigma_\theta^2(x)/\varsigma_\theta^{2,\text{MLE}}(x)$  for particular choices of  $\gamma = 0.2, 0.5, 0.8$  and  $\sigma = 1, 2, 5$ . Note that  $\sigma$  only shifts the ratio horizontally, leaving the limits  $x \rightarrow 0$  and  $x \rightarrow \infty$  unaffected.

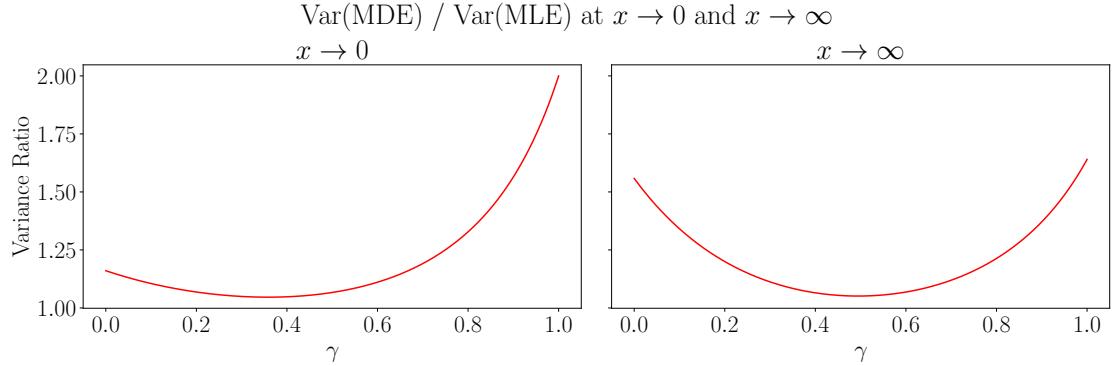


Figure 11: Limiting ratios  $\lim_{x \rightarrow 0} \varsigma_\theta^{2,\text{MLE}}(x)/\varsigma_\theta^2(x)$  and  $\lim_{x \rightarrow \infty} \varsigma_\theta^{2,\text{MLE}}(x)/\varsigma_\theta^2(x)$  as a function of  $\gamma$ . Note that the ratio never exceeds 2.

## C. Additional results for the challenge

In this section, we provide additional results on the data challenge. All plots (Figures 12 – 15) have their correspondence to a Figure in the main paper. For the sake of brevity, we therefore omit a lengthy discussion of each.

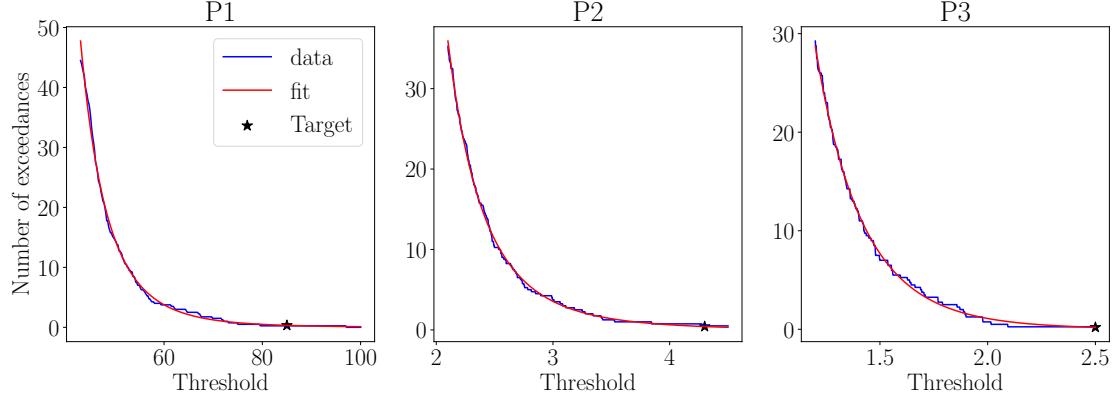


Figure 12: Average of empirical survival functions  $\hat{S}_k(x)$  (blue) and corresponding MDE fits  $\hat{S}_k^{\text{MDE}}(x)$  (red), targets P1 – P3. The black  $\star$  denotes the target of the respective challenge.

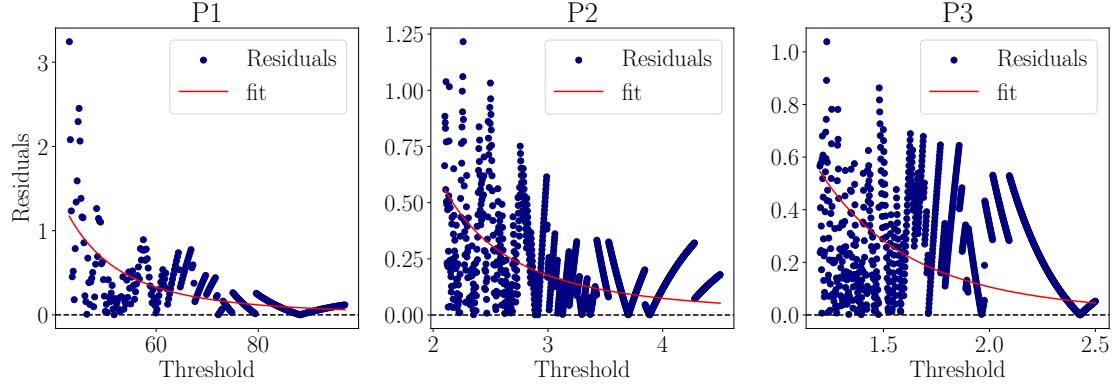


Figure 13: Residuals and fitted  $\tilde{\zeta}_k^2(x)$  for target P1 – P3.

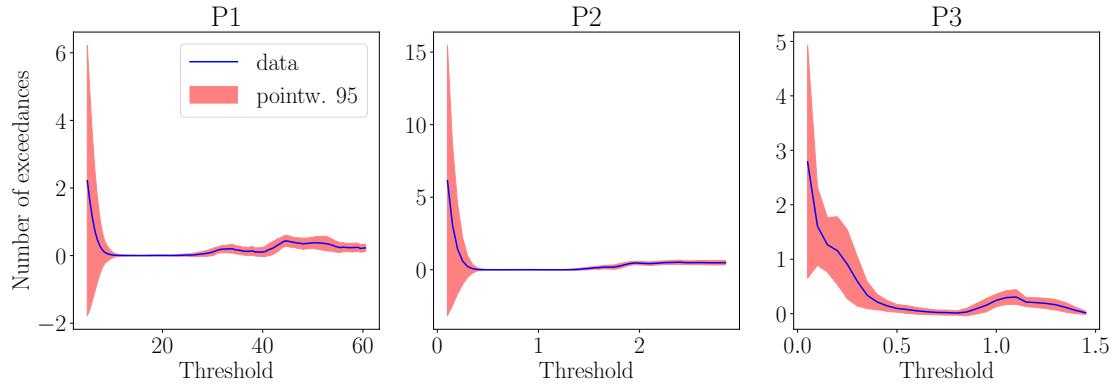


Figure 14: Development of the estimator  $\hat{S}^{\text{MDE}}(q)$  over the threshold  $u$  for target P1 – P3. The extreme threshold  $q = x + u$  is the target threshold and remains constant.

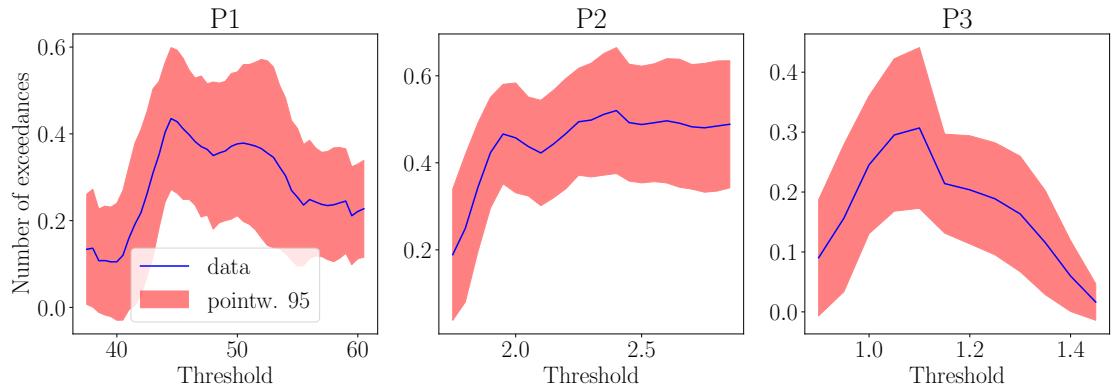


Figure 15: Zoom-in: development of the estimator  $\hat{S}^{\text{MDE}}(q)$  over the threshold  $u$  for target C1 – C3. The extreme threshold  $q = x + u$  is the target threshold and remains constant.

## D. Simulation Study

We perform a short Monte Carlo simulation by drawing  $n$  iid samples from a  $\text{GPD}(\gamma, 1)$  distribution. In particular, we are interested in the comparison of MSE, variance and bias for different choices of  $n$  and  $\gamma$ . Thus, we vary  $n = 10, 50, 100$  and  $\gamma \in \{0.05, 0.1, \dots, 0.6\}$ . Each MSE (variance, bias) is approximated via 1000 repetitions of each simulation experiment. The results are depicted in Figure 16.

It is notable that all MSEs are dominated by the variance. Due to the optimality in variance, the theoretical result is confirmed that the MLE outperforms the MDE in terms of MSE for most combinations of  $\gamma$  and  $n$ . The trends in MSE of GPD parameters directly translate to trends in MISE of survival function estimation, shown in Figure 17.

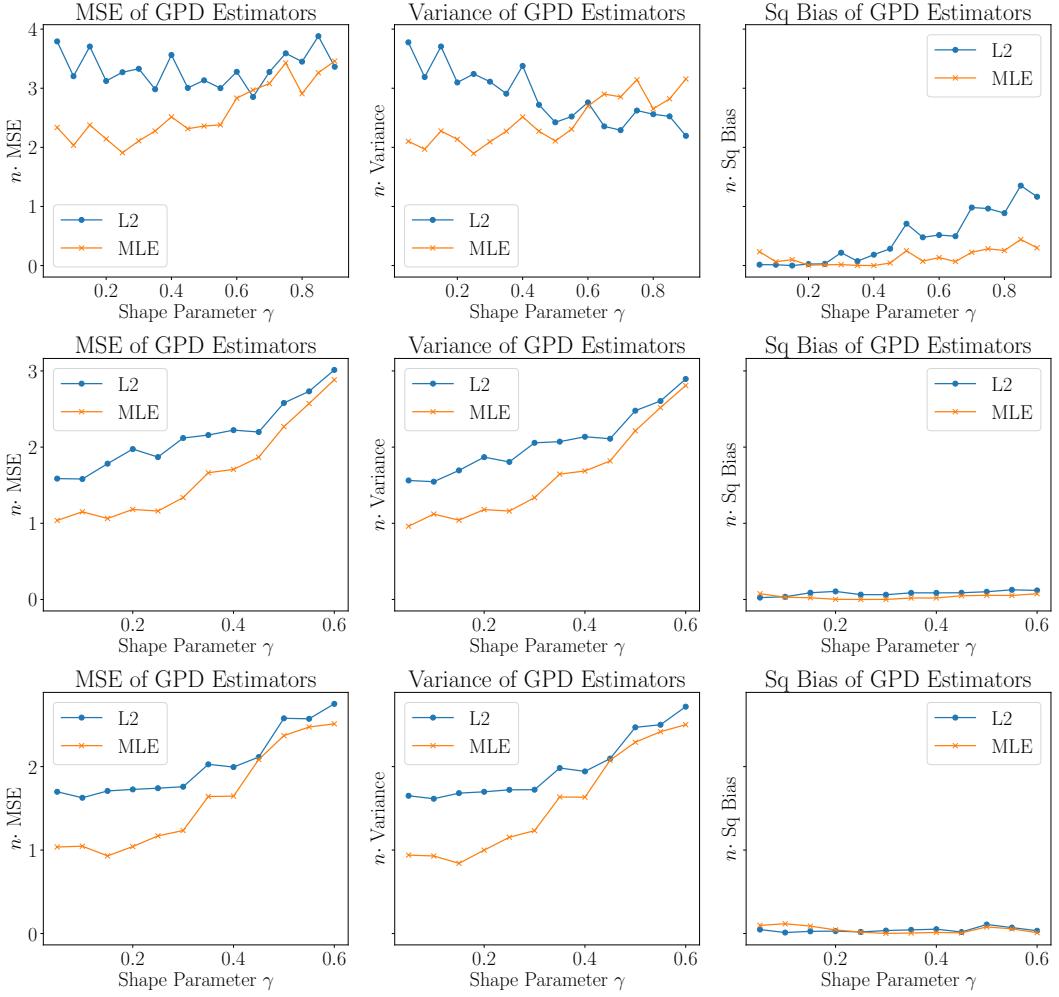


Figure 16: MSE, variance and squared bias for the shape parameter  $\gamma$ . estimation. For each point,  $n$  samples from a  $\text{GPD}(\gamma, 1)$  have been drawn and the MSE was computed over 1000 repetitions. From top to bottom:  $n = 10, 50, 100$ . We observe that all MSEs are variance-dominated and that MLE always outperforms MDE, but the MDE differences remain small.

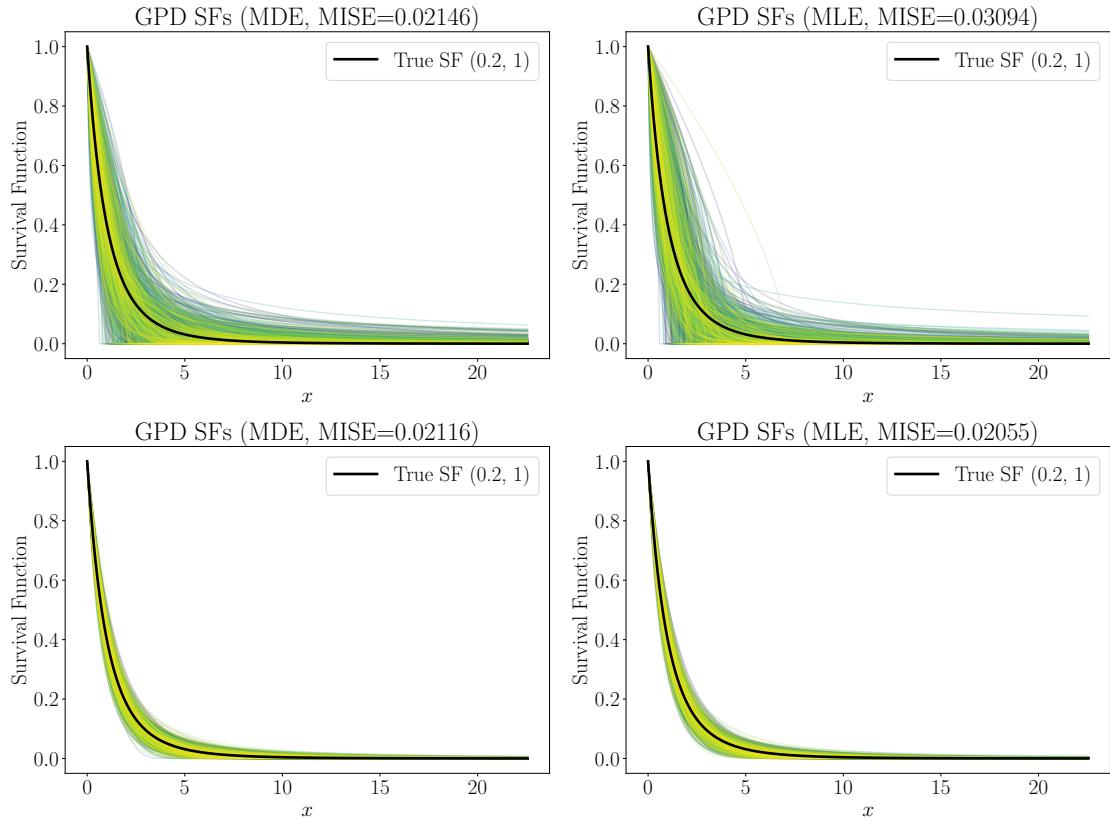


Figure 17: Estimated survival functions  $S_{\hat{\theta}}(x)$  around true  $S_{\theta_0}(x)$ ,  $\theta_0 = (0.2, 1)$ . Top:  $n = 10$ , bottom:  $n = 100$ . The MISE is computed over the shown domain on 1000 repetitions.

## E. Covariance Calculations with Mathematica

### Defining the estimating equations

```
In[1]:= F[x_] := 1 - (1 + γ / σ * x)^(−1 / γ)
In[2]:= f[x_] := D[F[x], x]
In[3]:= gpd = ProbabilityDistribution[f[x], {x, 0, ∞}, Assumptions → {σ > 0, γ > 0, γ < 1}];
```

Gamma derivative and estimating equation:

```
In[4]:= D[F[x], γ]
Out[4]= - $\left(1 + \frac{x \gamma}{\sigma}\right)^{-1/\gamma} \left(-\frac{x}{\gamma \left(1 + \frac{x \gamma}{\sigma}\right) \sigma} + \frac{\text{Log}\left[1 + \frac{x \gamma}{\sigma}\right]}{\gamma^2}\right)$ 
In[5]:= Integrate[D[F[y], γ], {y, 0, x}, Assumptions → 0 < γ < 1 && σ > 0 && x > 0]
Out[5]= 
$$\frac{-\gamma^2 \sigma + \left(\frac{\sigma}{x \gamma + \sigma}\right)^{\frac{1}{\gamma}} \left(\gamma \left(x (-1 + 2 \gamma) + \gamma \sigma\right) - (-1 + \gamma) (x \gamma + \sigma) \text{Log}\left[1 + \frac{x \gamma}{\sigma}\right]\right)}{(-1 + \gamma)^2 \gamma^2}$$

In[6]:= Expectation[
$$\frac{-\gamma^2 \sigma + \left(\frac{\sigma}{x \gamma + \sigma}\right)^{\frac{1}{\gamma}} \left(\gamma \left(x (-1 + 2 \gamma) + \gamma \sigma\right) - (-1 + \gamma) (x \gamma + \sigma) \text{Log}\left[1 + \frac{x \gamma}{\sigma}\right]\right)}{(-1 + \gamma)^2 \gamma^2}, x \approx \text{gpd}]$$

Out[6]= 
$$-\frac{\sigma}{2 (-2 + \gamma)^2}$$

In[7]:= PsiGam[x_] := 
$$\frac{-\gamma^2 \sigma + \left(\frac{\sigma}{x \gamma + \sigma}\right)^{\frac{1}{\gamma}} \left(\gamma \left(x (-1 + 2 \gamma) + \gamma \sigma\right) - (-1 + \gamma) (x \gamma + \sigma) \text{Log}\left[1 + \frac{x \gamma}{\sigma}\right]\right)}{(-1 + \gamma)^2 \gamma^2} + \frac{\sigma}{2 (-2 + \gamma)^2}$$

```

Sigma derivative and estimating equation:

```
In[8]:= Integrate[D[F[y], σ], {y, 0, x}, Assumptions → 0 < γ < 1 && σ > 0 && x > 0]
Out[8]= 
$$-\frac{\sigma + (x + \sigma) \left(\frac{\sigma}{x \gamma + \sigma}\right)^{\frac{1}{\gamma}}}{(-1 + \gamma) \sigma}$$

In[9]:= Expectation[-
$$\frac{-\sigma + (x + \sigma) \left(\frac{\sigma}{x \gamma + \sigma}\right)^{\frac{1}{\gamma}}}{(-1 + \gamma) \sigma}, x \approx \text{gpd}]$$

Out[9]= 
$$\frac{1}{2 (-2 + \gamma)}$$

```

$$\text{In}[1]:= \text{PsiSig}[x] := -\frac{-\sigma + (x + \sigma) \left(\frac{\sigma}{x \gamma + \sigma}\right)^{\frac{1}{\gamma}}}{(-1 + \gamma) \sigma} - \frac{1}{2 (-2 + \gamma)}$$

## Computing $V$

$$\text{In}[2]:= \text{Expectation}[\text{PsiGam}[x]^2, x \approx \text{gpd}]$$

$$\text{Out}[2]= \frac{(-1719 + 3416 \gamma - 2587 \gamma^2 + 918 \gamma^3 - 148 \gamma^4 + 8 \gamma^5) \sigma^2}{12 (-3 + \gamma)^2 (-2 + \gamma)^4 (-3 + 2 \gamma)^3}$$

$$\text{In}[3]:= \text{Expectation}[\text{PsiSig}[x]^2, x \approx \text{gpd}]$$

$$\text{Out}[3]= \frac{29 - 17 \gamma + 2 \gamma^2}{12 (-3 + \gamma) (-2 + \gamma)^2 (-3 + 2 \gamma)}$$

$$\text{In}[4]:= \text{Expectation}[\text{PsiSig}[x] * \text{PsiGam}[x], x \approx \text{gpd}]$$

$$\text{Out}[4]= -\frac{(369 - 548 \gamma + 285 \gamma^2 - 60 \gamma^3 + 4 \gamma^4) \sigma}{12 (-2 + \gamma)^3 (9 - 9 \gamma + 2 \gamma^2)^2}$$

$$\text{In}[5]:= \mathbf{V} =$$

$$\left\{ \left\{ \frac{(-1719 + 3416 \gamma - 2587 \gamma^2 + 918 \gamma^3 - 148 \gamma^4 + 8 \gamma^5) \sigma^2}{12 (-3 + \gamma)^2 (-2 + \gamma)^4 (-3 + 2 \gamma)^3}, -\frac{(369 - 548 \gamma + 285 \gamma^2 - 60 \gamma^3 + 4 \gamma^4) \sigma}{12 (-2 + \gamma)^3 (9 - 9 \gamma + 2 \gamma^2)^2} \right\},$$

$$\left\{ -\frac{(369 - 548 \gamma + 285 \gamma^2 - 60 \gamma^3 + 4 \gamma^4) \sigma}{12 (-2 + \gamma)^3 (9 - 9 \gamma + 2 \gamma^2)^2}, \frac{29 - 17 \gamma + 2 \gamma^2}{12 (-3 + \gamma) (-2 + \gamma)^2 (-3 + 2 \gamma)} \right\};$$

$$\text{In}[5]:= \text{Simplify}[\text{MatrixForm}[\mathbf{V}]]$$

$$\text{Out}[5]/\text{MatrixForm}=$$

$$\left( \begin{array}{cc} \frac{(-1719 + 3416 \gamma - 2587 \gamma^2 + 918 \gamma^3 - 148 \gamma^4 + 8 \gamma^5) \sigma^2}{12 (-3 + \gamma)^2 (-2 + \gamma)^4 (-3 + 2 \gamma)^3} & -\frac{(369 - 548 \gamma + 285 \gamma^2 - 60 \gamma^3 + 4 \gamma^4) \sigma}{12 (-2 + \gamma)^3 (9 - 9 \gamma + 2 \gamma^2)^2} \\ -\frac{(369 - 548 \gamma + 285 \gamma^2 - 60 \gamma^3 + 4 \gamma^4) \sigma}{12 (-2 + \gamma)^3 (9 - 9 \gamma + 2 \gamma^2)^2} & \frac{29 - 17 \gamma + 2 \gamma^2}{12 (-3 + \gamma) (-2 + \gamma)^2 (-3 + 2 \gamma)} \end{array} \right)$$

## Computing $U$

$$\text{In}[6]:= \text{Expectation}[\text{D}[\text{PsiGam}[x], \gamma], x \approx \text{gpd}]$$

$$\text{Out}[6]= \frac{(-6 + \gamma) \sigma}{2 (-2 + \gamma)^3 (2 + \gamma)}$$

In[8]:= **Expectation[D[PsiGam[x], σ], x ≈ gpd]**

$$\text{Out}[8]= -\frac{-6 + \gamma}{4 (-2 + \gamma)^2 (2 + \gamma)}$$

In[9]:= **Expectation[D[PsiSig[x], γ], x ≈ gpd]**

$$\text{Out}[9]= -\frac{-6 + \gamma}{4 (-2 + \gamma)^2 (2 + \gamma)}$$

In[10]:= **Expectation[D[PsiSig[x], σ], x ≈ gpd]**

$$\text{Out}[10]= -\frac{1}{(-4 + \gamma^2) \sigma}$$

$$\text{In[11]:= } \mathbf{U} = \left\{ \left\{ \frac{(-6 + \gamma) \sigma}{2 (-2 + \gamma)^3 (2 + \gamma)}, -\frac{-6 + \gamma}{4 (-2 + \gamma)^2 (2 + \gamma)} \right\}, \left\{ -\frac{-6 + \gamma}{4 (-2 + \gamma)^2 (2 + \gamma)}, -\frac{1}{(-4 + \gamma^2) \sigma} \right\} \right\};$$

In[12]:= **Simplify[MatrixForm[U]]**

$$\text{Out}[12]//\text{MatrixForm}= \begin{pmatrix} \frac{(-6+\gamma) \sigma}{2 (-2+\gamma)^3 (2+\gamma)} & -\frac{-6+\gamma}{4 (-2+\gamma)^2 (2+\gamma)} \\ -\frac{-6+\gamma}{4 (-2+\gamma)^2 (2+\gamma)} & \frac{1}{4 \sigma - \gamma^2 \sigma} \end{pmatrix}$$

## Computing $\Sigma$

In[13]:= **Sigma = FullSimplify[Inverse[U].V.Inverse[Transpose[U]]];**

In[14]:= **MatrixForm[Sigma]**

$$\text{Out}[14]//\text{MatrixForm}= \begin{pmatrix} \frac{4 (-2+\gamma)^4 (-639+\gamma (452+\gamma (-683+2 \gamma (243-58 \gamma+4 \gamma^2))))}{3 (-6+\gamma)^2 (-3+\gamma)^2 (-3+2 \gamma)^3} & \frac{4 (-2+\gamma)^2 (-387+\gamma (778+\gamma (-791+2 \gamma (207-50 \gamma+4 \gamma^2)))) \sigma}{3 (-6+\gamma) (-3+\gamma)^2 (-3+2 \gamma)^3} \\ \frac{4 (-2+\gamma)^2 (-387+\gamma (778+\gamma (-791+2 \gamma (207-50 \gamma+4 \gamma^2)))) \sigma}{3 (-6+\gamma) (-3+\gamma)^2 (-3+2 \gamma)^3} & \frac{4 (-423+\gamma (944+\gamma (-843+374 \gamma-84 \gamma^2+8 \gamma^3))) \sigma^2}{3 (-3+\gamma)^2 (-3+2 \gamma)^3} \end{pmatrix}$$

## Computing the Covariance of $G(x)$

```
In[45]:= First[FullSimplify[{D[F[x], \gamma], D[F[x], \sigma]}.Sigma.{D[F[x], \gamma], {D[F[x], \sigma]}}]]
```

$$\text{Out}[45]= \left(4 \left(1+\frac{x \gamma}{\sigma}\right)^{-2/\gamma} \left(4 x^2 \gamma^2 (2+\gamma)^2 \left(639+\gamma \left(-1208+\gamma \left(857-266 \gamma+30 \gamma^2\right)\right)\right)+(-2+\gamma)^2 \left(x \gamma+\sigma\right) \text{Log}\left[1+\frac{x \gamma}{\sigma}\right] \left(-4 x \gamma (2+\gamma) \left(639+2 \gamma \left(-415+\gamma \left(223-62 \gamma+8 \gamma^2\right)\right)\right)-(-2+\gamma)^2 \left(-639+\gamma \left(452+\gamma \left(-683+2 \gamma \left(243-58 \gamma+4 \gamma^2\right)\right)\right)\right) \left(x \gamma+\sigma\right) \text{Log}\left[1+\frac{x \gamma}{\sigma}\right]\right)\right)/\left(3 \left(3-2 \gamma\right)^3 (-6+\gamma)^2 (-3+\gamma)^2 \gamma^4 \left(x \gamma+\sigma\right)^2\right)$$

```
In[46]:= sigma[x_, \gamma_, \sigma_]:=
```

$$\left(4 \left(1+\frac{x \gamma}{\sigma}\right)^{-2/\gamma} \left(4 x^2 \gamma^2 (2+\gamma)^2 \left(639+\gamma \left(-1208+\gamma \left(857-266 \gamma+30 \gamma^2\right)\right)\right)+(-2+\gamma)^2 \left(x \gamma+\sigma\right) \text{Log}\left[1+\frac{x \gamma}{\sigma}\right] \left(-4 x \gamma (2+\gamma) \left(639+2 \gamma \left(-415+\gamma \left(223-62 \gamma+8 \gamma^2\right)\right)\right)-(-2+\gamma)^2 \left(-639+\gamma \left(452+\gamma \left(-683+2 \gamma \left(243-58 \gamma+4 \gamma^2\right)\right)\right)\right) \left(x \gamma+\sigma\right) \text{Log}\left[1+\frac{x \gamma}{\sigma}\right]\right)\right)/\left(3 \left(3-2 \gamma\right)^3 (-6+\gamma)^2 (-3+\gamma)^2 \gamma^4 \left(x \gamma+\sigma\right)^2\right)$$

## Comparison to the MLE

```
In[8]:= Expectation[(D[Log[f[x]], \gamma])^2, x \approx gpd]
```

$$\text{Out}[8]= \frac{2}{1+3 \gamma+2 \gamma^2}$$

```
In[9]:= Expectation[(D[Log[f[x]], \gamma])*(D[Log[f[x]], \sigma]), x \approx gpd]
```

$$\text{Out}[9]= \frac{1}{(1+3 \gamma+2 \gamma^2) \sigma}$$

```
In[10]:= Expectation[(D[Log[f[x]], \sigma])^2, x \approx gpd]
```

$$\text{Out}[10]= \frac{1}{(1+2 \gamma) \sigma^2}$$

```
In[35]:= Fish = {{2/(1 + 3 γ + 2 γ^2), 1/(1 + 3 γ + 2 γ^2) σ}, {1/(1 + 3 γ + 2 γ^2) σ, 1/(1 + 2 γ) σ^2}};
```

```
In[58]:= MatrixForm[FullSimplify[Inverse[Fish]]]
Out[58]//MatrixForm=
```

$$\begin{pmatrix} (1+\gamma)^2 & -(1+\gamma)\sigma \\ -(1+\gamma)\sigma & 2(1+\gamma)\sigma^2 \end{pmatrix}$$

```
In[59]:= First[
  FullSimplify[{D[F[x], γ], D[F[x], σ]}.FullSimplify[Inverse[Fish]].{{D[F[x], γ]}, {D[F[x], σ]}]]]
Out[59]=
```

$$\frac{1}{\gamma^4 (x \gamma + \sigma)^2} (1 + \gamma) \left(1 + \frac{x \gamma}{\sigma}\right)^{-2/\gamma} \left(x^2 \gamma^2 (1 + \gamma) (1 + 2 \gamma) + (x \gamma + \sigma) \text{Log}\left[1 + \frac{x \gamma}{\sigma}\right] \left(-2 x \gamma (1 + 2 \gamma) + (1 + \gamma) (x \gamma + \sigma) \text{Log}\left[1 + \frac{x \gamma}{\sigma}\right]\right)\right)$$

```
In[38]:= sigmaMLE[x_, γ_, σ_] :=
```

$$\frac{1}{\gamma^4 (x \gamma + \sigma)^2} (1 + \gamma) \left(1 + \frac{x \gamma}{\sigma}\right)^{-2/\gamma} \left(x^2 \gamma^2 (1 + \gamma) (1 + 2 \gamma) + (x \gamma + \sigma) \text{Log}\left[1 + \frac{x \gamma}{\sigma}\right] \left(-2 x \gamma (1 + 2 \gamma) + (1 + \gamma) (x \gamma + \sigma) \text{Log}\left[1 + \frac{x \gamma}{\sigma}\right]\right)\right)$$

```
In[41]:= N[sigmaMLE[1, 5, 1]]
Out[41]= 0.151181
```

```
In[47]:= γ=.
```

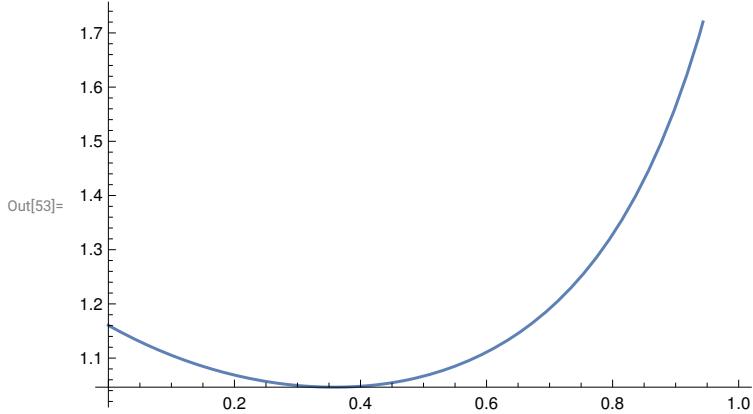
```
In[48]:= σ=.
```

```
In[45]:= N[Inverse[Fish]]
Out[45]= {{6.76, -5.98}, {-5.98, 27.508}}
```

```
In[51]:= Limit[sigma[x, γ, σ]/sigmaMLE[x, γ, σ], x → 0, Assumptions → {γ > 0, γ < 1, σ > 0}]
Out[51]=
```

$$\frac{2 (-423 + 944 \gamma - 843 \gamma^2 + 374 \gamma^3 - 84 \gamma^4 + 8 \gamma^5)}{3 (-3 + \gamma)^2 (1 + \gamma) (-3 + 2 \gamma)^3}$$

$$\text{In[53]:= Plot}\left[\frac{2 \left(-423+944 \gamma -843 \gamma^2+374 \gamma^3-84 \gamma^4+8 \gamma^5\right)}{3 (-3+\gamma)^2 (1+\gamma) (-3+2 \gamma)^3}, \{\gamma, 0, 1\}\right]$$



$$\text{In[54]:= Limit}\left[\sigma[x, \gamma, \sigma] / \sigma_{MLE}[x, \gamma, \sigma], x \rightarrow \text{Infinity}, \text{Assumptions} \rightarrow \{\gamma > 0, \gamma < 1, \sigma > 0\}\right]$$

$$\text{Out[54]= } -\frac{4 (-2+\gamma)^4 \left(-639+\gamma \left(452+\gamma \left(-683+2 \gamma \left(243-58 \gamma +4 \gamma^2\right)\right)\right)\right)}{3 \left(3-2 \gamma \right)^3 (-6+\gamma )^2 (-3+\gamma )^2 (1+\gamma )^2}$$

$$\text{In[56]:= Plot}\left[-\frac{4 (-2+\gamma)^4 \left(-639+\gamma \left(452+\gamma \left(-683+2 \gamma \left(243-58 \gamma +4 \gamma^2\right)\right)\right)\right)}{3 \left(3-2 \gamma \right)^3 (-6+\gamma )^2 (-3+\gamma )^2 (1+\gamma )^2}, \{\gamma, 0, 1\}\right]$$

