

mev package tutorial

Léo Belzile

2019-06-28

Contents

Preliminary remarks	5
1 Univariate peaks-over-threshold analysis	7
2 Penultimate approximation	15
3 Threshold selection	19
3.1 Robust selection	20
3.2 Threshold stability plots	20
3.3 Wadsworth's diagnostics: white noise process and simultaneous threshold stability plots	20
3.4 Changepoint tests based on penultimate approximation	21
3.5 Extended generalized Pareto	22
4 Simulation	23

Preliminary remarks

The `mev` package bundles routine for - likelihood-based inference for univariate extremes (including distributions, densities, score, information matrix and higher order asymptotics). The four basic likelihood (generalized extreme value, generalized Pareto, non-homogeneous Poisson process and r -largest) are implemented. For the first two distributions, additional higher order inference tools are available, including profile likelihoods in most parametrizations of interest. Many recent threshold selection diagnostics that have appeared in the literature are implemented.

The second includes multivariate goodness-of-fit diagnostics and tests (max-stability test, information matrix test, coefficient of tail dependence, extremal coefficient, estimators of the extremal index, extremogram, nonparametric estimation of the angular measure, etc.) Pairwise composite likelihood for general models will be included in a future update.

Last, but not least, the package includes functions to simulate from multivariate max-stable vectors in dimension $d \geq 2$, including extensions of most of the families in the `evd` package. The package also includes simulation algorithms for R -Pareto processes using accept-reject and composition sampling, and accept-reject for generalized R -Pareto processes.

You can install the latest version of the package (v . 1 . 12) directly from CRAN via

```
install.packages("mev")
```


Chapter 1

Univariate peaks-over-threshold analysis

Let $\ell(\mathbf{y}; \boldsymbol{\theta})$ denotes the log-likelihood of an n sample with a p -dimensional parameter $\boldsymbol{\theta}$. The score vector is $U(\boldsymbol{\theta}) = \partial \ell / \partial \boldsymbol{\theta}$, while the Fisher information is $i(\boldsymbol{\theta}) = E\{U(\boldsymbol{\theta})U(\boldsymbol{\theta})^\top\}$. Under regularity conditions, we also have $i(\boldsymbol{\theta}) = -E(\partial^2 \ell / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top)$. The observed information is the Hessian $-\partial^2 \ell / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top$, evaluated at the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$.

By definition, the maximum likelihood estimator solves the score equation, i.e. $U(\hat{\boldsymbol{\theta}}) = \mathbf{0}_p$. If the maximum likelihood estimator is not available in closed-form, this can be used to verify that the optimization routine has converged. It can also be used for Newton methods.

There are four basic likelihoods for univariate extremes: the likelihood of the generalized extreme value (GEV) distribution for block maxima, the likelihood for the generalized Pareto distribution and that of the non-homogeneous Poisson process (NHPP) for exceedances above a threshold u , the likelihood of the r -largest observations, which can be used to model the r -largest observations per block or threshold exceedances where the threshold is the r th order statistic. The NHPP likelihood includes a contribution for the fraction of points that exceeds the threshold, whereas the generalized Pareto is a conditional distribution, whose normalizing constant $\zeta_u = \Pr(Y > u)$ is the third parameter.

The generalized extreme value (GEV) distribution with location parameter $\mu \in \mathbb{R}$, scale parameter $\sigma \in \mathbb{R}_+$ and shape parameter $\xi \in \mathbb{R}$ is

$$G(x) = \begin{cases} \exp \left\{ - \left(1 + \xi \frac{x - \mu}{\sigma} \right)^{-1/\xi} \right\}, & \xi \neq 0, \\ \exp \left\{ - \exp \left(- \frac{x - \mu}{\sigma} \right) \right\}, & \xi = 0, \end{cases}$$

defined on $\{x \in \mathbb{R} : \xi(x - \mu)/\sigma > -1\}$ where $x_+ = \max\{0, x\}$. The case $\xi = 0$ is

commonly known as the Gumbel distribution. We denote the distribution by $\text{GEV}(\mu, \sigma, \xi)$.

The generalized Pareto (GP) distribution with scale $\sigma \in \mathbb{R}_+$ and shape $\xi \in \mathbb{R}$ is

$$G(x) = \begin{cases} 1 - \left(1 + \xi \frac{x}{\sigma}\right)_+^{-1/\xi}, & \xi \neq 0, \\ 1 - \exp\left(-\frac{x}{\sigma}\right), & \xi = 0. \end{cases}$$

The range of the generalized Pareto distribution is $[0, -\sigma/\xi]$ if $\xi < 0$ and is \mathbb{R}_+ otherwise. We denote the distribution by $\text{GP}(\sigma, \xi)$.

Let $Y_{(1)} \geq \dots \geq Y_{(r)}$ denote the r largest observations from a sample. The likelihood of the limiting distribution of the point process for the r -largest observations is

$$\ell(\mu, \sigma, \xi; \mathbf{y}) \equiv -r \log(\sigma) - \left(1 + \frac{1}{\xi}\right) \sum_{j=1}^r \log\left(1 + \xi \frac{Y_{(j)} - \mu}{\sigma}\right)_+ - \left(1 + \xi \frac{Y_{(r)} - \mu}{\sigma}\right)_+^{-1/\xi}, \quad \mu, \xi \in \mathbb{R}, \sigma > 0.$$

Consider a sample of N observations, of which n_u exceed u and which we denote by y_1, \dots, y_{n_u} . The likelihood associated to the limiting distribution of threshold exceedances is

$$L(\mu, \sigma, \xi; \mathbf{y}) = \exp\left[-c \left\{1 + \xi \left(\frac{u - \mu}{\sigma}\right)\right\}_+^{-1/\xi}\right] (c\sigma)^{-n_u} \prod_{i=1}^{n_u} \left\{1 + \xi \left(\frac{y_i - \mu}{\sigma}\right)\right\}_+^{-1/\xi - 1}, \quad \mu, \xi \in \mathbb{R}, \sigma > 0, \quad (1.1)$$

where $(\cdot)_+ = \max\{0, \cdot\}$. The quantity c is a tuning parameter whose role is described in §7.5 of Coles (2001). If we take $c = N/m$, the parameters of the point process likelihood correspond to those of the generalized extreme value distribution fitted to blocks of size m .

1.0.1 Profile likelihood

The profile likelihood ℓ_p , a function of $\boldsymbol{\psi}$ alone, is obtained by maximizing the likelihood pointwise at each fixed value $\boldsymbol{\psi} = \boldsymbol{\psi}_0$ over the nuisance vector $\boldsymbol{\lambda}_{\boldsymbol{\psi}_0}$,

$$\ell_p(\boldsymbol{\psi}) = \max_{\boldsymbol{\lambda}} \ell(\boldsymbol{\psi}, \boldsymbol{\lambda}) = \ell(\boldsymbol{\psi}, \hat{\boldsymbol{\lambda}}_{\boldsymbol{\psi}}).$$

The observed profile information function is

$$j_p(\boldsymbol{\psi}) = -\frac{\partial \ell_p(\boldsymbol{\psi})}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}^\top} = \{j^{\boldsymbol{\psi} \boldsymbol{\psi}}(\boldsymbol{\psi}, \hat{\boldsymbol{\lambda}}_{\boldsymbol{\psi}})\}^{-1}.$$

The profile likelihood is not a genuine likelihood in the sense that it is not based on the density of a random variable.

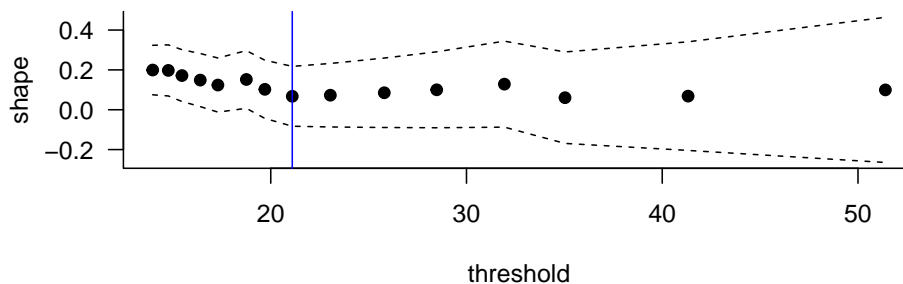
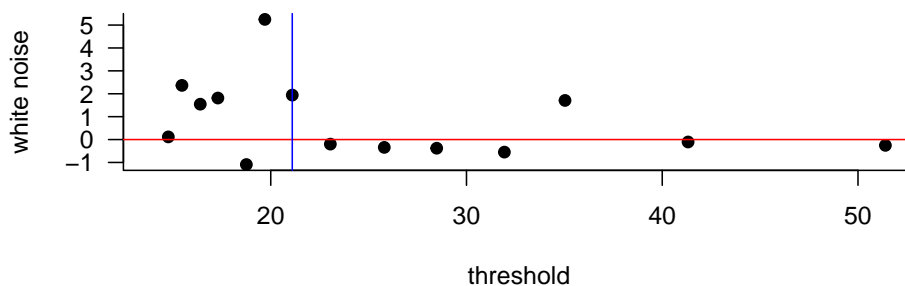
This example illustrates some of the functions used in peaks-over-threshold analysis based on fitting a generalized Pareto distribution to threshold exceedances. We use the Venezuelan rainfall data, a time series of daily rainfall precipitations at Maiquetia airport in Venezuela, for the purpose of illustration.


```
library(mev)
library(lubridate, quietly = TRUE, warn.conflicts = FALSE)

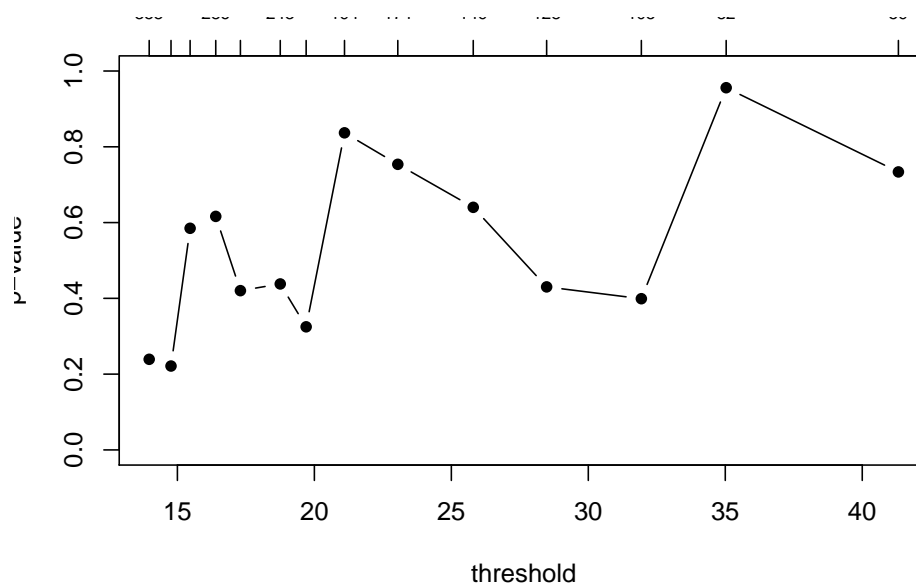
data("maiquetia")
day <- seq.Date(from = as.Date("1961-01-01"),
               to = as.Date("1999-12-31"), by = "day")
# Keep non-zero rainfall, exclude 1999 observations
nzrain <- maiquetia[year(day) < 1999 & maiquetia > 0]
```

We will ignore temporal dependence and stationarity, but these should be considered. The first step in our analysis is to choose a threshold. We select candidates based on quantiles and test whether they lead to stable inference.

```
ths <- quantile(nzrain, seq(0.9, 0.99, length.out = 15))
# Threshold selection diagnostics
W.diag(xdat = nzrain, model = "nhpp", plots = c("WN", "PS"), u = ths)
```



```
NC.diag(nzrain, u = ths)
```



Threshold selection diagnostics indicate that the shape is more or less constant over the range of threshold considered. Wadsworth's white noise sequence yields a very large residual around $u = 20$, but the Northrop–Coleman score test yields P -value path above 0.2 for every threshold under consideration.

```
# Fit using maximum likelihood estimation
gpdf <- fit.gpd(nzrain, threshold = 20, show = TRUE)
```

```
## Method: Grimshaw
## Log-likelihood: -832.629
##
## Threshold: 20
## Number Above: 216
## Proportion Above: 0.0604
##
## Estimates
##   scale   shape
## 15.5800  0.1088
##
## Standard Errors
##   scale   shape
## 1.60673  0.07785
##
## Optimization Information
##   Convergence: successful
```

The default optimization routine for the generalized Pareto distribution is

Grimshaw's method, which profiles out the likelihood. The method is almost guaranteed to converge. Because of non-regularity, the maximum likelihood estimator for $\xi < -1$ does not solve the score equation and leads to infinite log-likelihood, hence the maximum returned lies on the boundary. The standard errors are based on the inverse observed information matrix and provided only if $\xi > -1/2$. We can verify that our maximum likelihood estimate is indeed a maximum by checking if it solves the score equation if $\hat{\xi} > -1$.

```
isTRUE(all.equal(
  gpd.score(gpdf$estimate, dat = gpdf$exceedances),
  c(0,0), tolerance = 1e-5))
```

```
## [1] TRUE
```

If the sample is small, maximum likelihood estimators are biased for the generalized Pareto distribution (the shape parameter is negatively biased, regardless of the true value for ξ). Bias correction methods includes the modified score of Firth, but the default method is the implicit correction (subtract), which solves the implicit equation

$$\tilde{\theta} = \hat{\theta} - b(\tilde{\theta}). \quad (1.2)$$

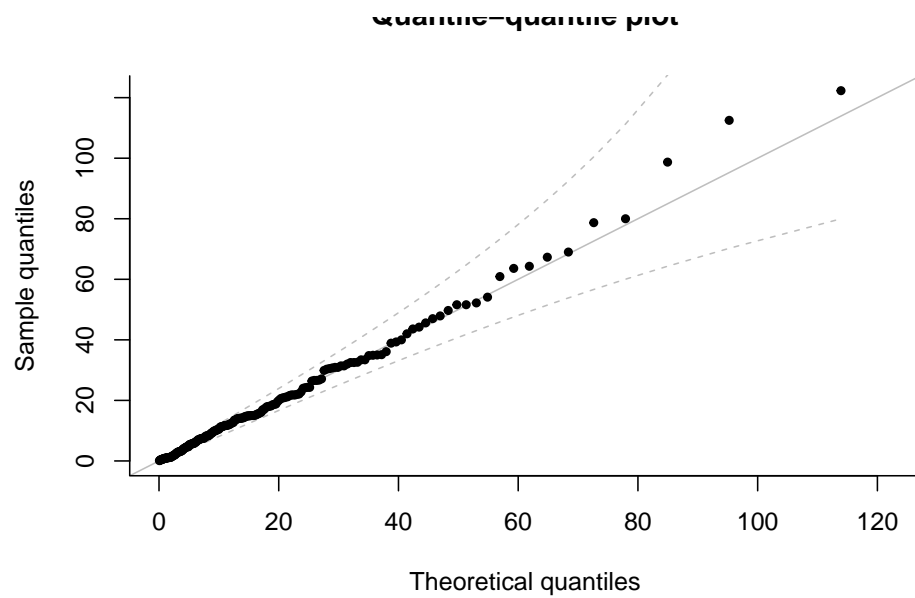
The point estimate $\tilde{\theta}$ is obtained numerically as the root of this nonlinear system of equations. In the present case, the sample size is large and hence the first-order correction, derived through asymptotic arguments from the generalized Pareto distribution likelihood, is small. Note that the bias correction requires $\xi > -1/3$, since it is based on third-order cumulants of the distribution.

```
gpdbcor <- gpd.bcor(dat = gpdf$exceedances, par = gpdf$estimate)
#print the differences between MLE and bias-corrected estimates
gpdf$estimate - gpdbcor
```

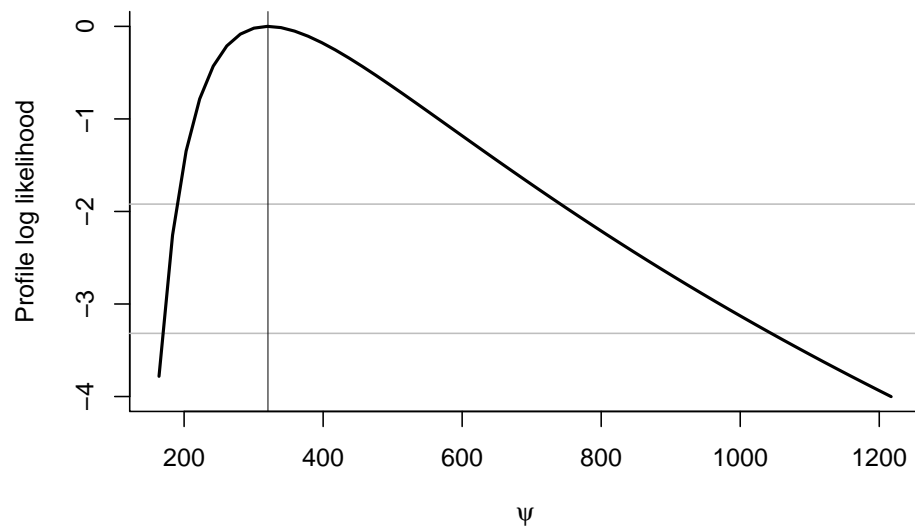
```
##      scale      shape
## 0.1915284 -0.0118876
```

The package includes some default diagnostic plots (probability-probability plots and quantile-quantile plots), which include approximate confidence intervals based on order statistics. We can also get profile likelihood and profile-based confidence intervals for most quantities of interest (parameters of the generalized Pareto distribution, excepted shortfall, return levels, N -observation maxima mean and quantiles). The example below gives the estimated profile for the median of the centennial maximum distribution conditional on exceeding 15, along with 95% confidence intervals.

```
plot(gpdf, which = 2) #Q-Q plot
```



```
# Profile of median of maxima of 100 years
profile <- gpd.pll(param = "Nquant", dat = nzrain, threshold = 15,
  N = length(nzrain)/(1998-1961)*100, q = 0.5)
```



```
# 95% confidence intervals
conf <- confint(profile, print = TRUE)
```

```
## Point estimate for the parameter of interest psi:
```

```
## Maximum likelihood      : 320.622
##
## Confidence intervals, levels : 0.025 0.975
## Wald intervals          : 118.72 522.524
## Profile likelihood       : 177.214 741.869
```


Chapter 2

Penultimate approximation

The generalized extreme value (GEV) distribution arises as the non-degenerate limiting distribution of maxima. A key property of the generalized extreme value distribution is max-stability, meaning that the distribution of maxima of $T \in \mathbb{N}$ is the same up to a location-scale transform: if $F \sim \text{GEV}$, then there exist location and scale constants $b \in \mathbb{R}$ and $a > 0$ such that $F^T(ax + b) = F(x)$. The parameters of the new distribution are easily derived: if $X_i \stackrel{\text{iid}}{\sim} \text{GEV}(\mu, \sigma, \xi)$, then $\max\{X_1, \dots, X_T\} \sim \text{GEV}(\mu_T, \sigma_T, \xi)$ with $\mu_T = \mu - \sigma(1 - T^\xi)/\xi$ and $\sigma_T = \sigma T^\xi$ for $\xi \neq 0$, or $\mu_T = \mu + \sigma \log(T)$ and $\sigma_T = \sigma$ if $\xi = 0$.

In practice, the GEV is fitted to data that are partitioned into k blocks of (potentially random) sizes m_i . We shall assume that the block sizes are all the same, so that $m_i = m$ for $i = 1, \dots, k$ and $n = km$. The quality of the generalized extreme value approximation increases with the block size m . For fixed block sizes, the estimated shape parameter generally differs from its asymptotic counterpart and this discrepancy usually introduces bias if one extrapolates far beyond the range of the observed data.

Let $F(x)$ denote a thrice-differentiable distribution function with endpoint x^* and density $f(x)$ and define $s(x) = -F(x) \log\{F(x)\}/f(x)$. The existence of the limit $\xi_\infty = \lim_{n \rightarrow \infty} s'\{b_n\}$ is necessary and sufficient for convergence to an extreme value distribution G ,

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = \exp\left\{-(1 + \xi_\infty x)^{-1/\xi_\infty}\right\} = G(x).$$

Smith (1987) shows that, for any $x \in \{y : 1 + \xi_\infty y > 0\}$, there exists z such that

$$\frac{-\log[F\{v + xs(v)\}]}{-\log[F(v)]} = \{1 + xs'(z)\}^{-1/s'(z)}, \quad v < z < v + xs(v).$$

For each $n \geq 1$, setting $v = b_n$ and $a_n = s(b_n)$ yields

$$F^n(a_n x + b_n) = \exp\left[-\{1 + s'(z)x\}^{-1/s'(z)}\right] + O(n^{-1})$$

for $z \in [\min(a_n x + b_n, b_n), \max(a_n x + b_n, b_n)]$, depending on the support of F . The ultimate approximation replaces $s'(z)$ by $s'(x^*) = \xi_\infty$, whereas Smith (1987) suggests instead suggests taking $s'(b_n)$, which is closer to $s'(a_n x + b_n)$ than is $s'(x^*)$.

The maximum likelihood estimate of the shape should change as the threshold or the block size increases. Consider for simplicity yearly maxima arising from blocking $m = n_y$ observations per year. We are interested in the distribution of the maximum of N years and thus the target has roughly shape ξ_{Nn_y} , but our estimates will instead give approximately ξ_{n_y} ; an extrapolation error arises from this mismatch between the shape values, which would be constant if the observations were truly max-stable. The curvature of s' determines how stable the estimates of ξ_t are when the extrapolation window increases.

We illustrate the previous discussion by looking at block maximum from a log-normal distribution.

```
library(mev)
set.seed(123)
x <- seq(1, 30, length=200)
fitted <- t(replicate(n = 1000,
                      fit.gev(apply(matrix(rlnorm(30*100), ncol= 30), 1, max),
                                    method = "BFGS")$est))
penult.bm.lnorm <- mev::smith.penult(model = "bm", m = (m <- 30), family = "lnorm")

par(mfrow = c(1,2), mar = c(5,5,1,1))
hist(fitted[,3], probability = TRUE, breaks = 20,
     xlab = "Estimated shape", main = "")
segments(y0 = -0.5, y1 = 0, x0 = penult.bm.lnorm$shape, col = "red", lwd = 2)
segments(y0 = -0.5, y1 = 0, x0 = 0, col = "black", lwd = 2)

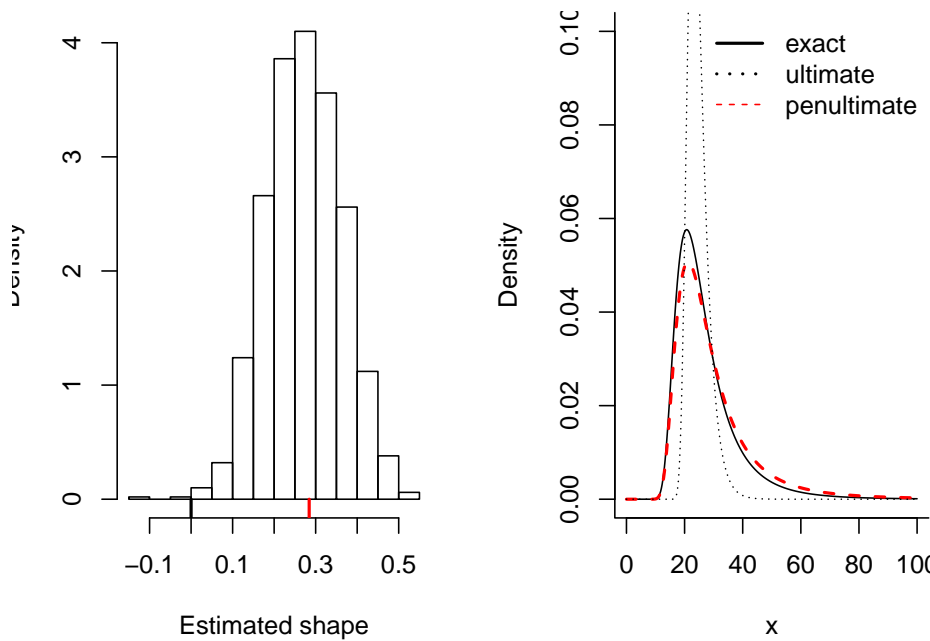
p30 <- penult.bm.lnorm
x = seq(0,100, length = 400)
N <- 1000; N30 = N/30
plot(x, N*exp((N-1)*plnorm(x, log.p=TRUE))*
     dlnorm(x),
     type="l", bty = "l", ylim = c(0,0.1),
     xlab="x", ylab="Density")
# Get parameters of maximum of N GEV
maxstabp <- function(loc, scale, shape, N){
  if(!isTRUE(all.equal(shape, 0, check.attributes = FALSE))){
    mut <- loc - scale*(1-N^shape)/shape
    sigmat = scale*N^shape
    return(c(loc = mut, scale = sigmat, shape = shape))
  } else{
    mut <- loc + scale*log(N)
    return(c(loc = mut, scale = scale, shape = shape))
  }
}
```



```

    }
  }
  p30e <- maxstabp(loc = p30$loc, scale = p30$scale, shape = p30$shape, N = N30)
  lines(x, evd::dgev(x, loc = p30e['loc'],
                    scale = p30e['scale'],
                    shape = p30e['shape']),
        col = "red", lwd = 2, lty = 2)
  p30l <- maxstabp(loc = p30$loc, scale = p30$scale, shape = 0, N = N30)
  lines(x, evd::dgev(x,
                    loc = p30e['loc'],
                    scale = p30l['scale'],
                    shape = 0), col = 1, lty=3, lwd=1)
  legend(x = "topright",
        legend = c("exact", "ultimate", "penultimate"),
        col = c(1,1,2),
        lwd = c(2,2,1), bty = "n",
        lty = c(1,3,2))

```



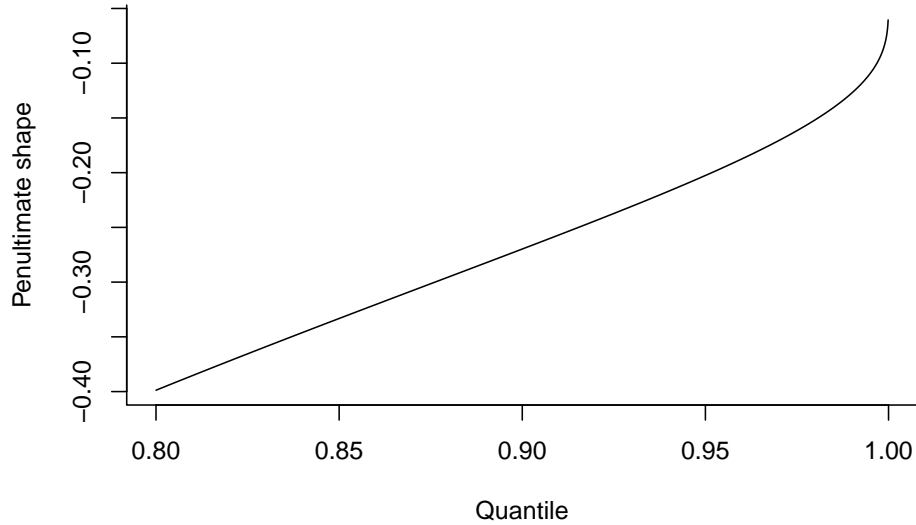
The left panel illustrates how maximum likelihood estimates of the parameters for repeated samples from a log-normal distribution are closer to the penultimate approximation than to the limiting tail index $\xi_\infty = 0$. Using max-stability, we could get an estimate of the distribution of the maximum of Nn_y observations. In this case, the penultimate shape for $m = 1000$ is approximately $\xi_{1000} = 0.215$, compared to $\xi_{30} \approx 0.284$, which is far from the limiting value $\xi_\infty = 0$. The extrapolated den-

sity estimate is based on $F_1^{1000/30}$, the generalized extreme value density associated to the distribution function of the penultimate approximation $F_1^{1000/30}$, with $F_1 \sim \text{GEV}(a_{30}, b_{30}, \xi_{30})$. The penultimate approximation F_2 is more accurate than the ultimate approximation $\text{GEV}(a_{30}, b_{30}, \xi_{\infty})$, which is too short tailed.

The function `smith.penult` computes the penultimate approximation for parameter models in **R** for the distribution assuming a model family with arguments `dfamily`, `qfamily` and `pfamily` exist. It returns the penultimate parameters for a given block size or quantile in case of threshold models. These functions are particularly useful for simulation studies in which one wants to obtain an approximation for F^N for large N , to get an approximation to the asymptotic distribution of a test statistic for finite N ; the GEV penultimate approximation is more numerically stable for certain models and its moments and cumulants are easily derived.

The following plot illustrates the penultimate shape parameter for the normal distribution based on threshold exceedances. In finite samples, the upper tail is bounded, but the quality of the approximation by the Gumbel distribution increases as $u \rightarrow \infty$, albeit very slowly.

```
penult <- smith.penult(family = "norm", method = "pot",
                      u = qnorm(u <- seq(0.8, 0.9999, by = 0.0001)))
plot(u, penult$shape, bty = "l", type='l', xlab='Quantile', ylab='Penultimate shape')
```



Chapter 3

Threshold selection

The generalized Pareto (GP) distribution with scale $\sigma \in \mathbb{R}_+$ and shape $\xi \in \mathbb{R}$ is

$$G(x) = \begin{cases} 1 - \{1 + \xi (\frac{x}{\sigma})\}_+^{-1/\xi}, & \xi \neq 0, \\ 1 - \exp(-x\sigma), & \xi = 0. \end{cases}$$

The range of the generalized Pareto distribution is $[0, -\sigma/\xi)$ if $\xi < 0$ and \mathbb{R}_+ otherwise.

Why is the generalized Pareto distribution so central to peaks-over-threshold analysis? The conditional distribution of exceedances over a threshold $u < x^*$ converges to a generalized Pareto distribution as u converges to the endpoint of the distribution,

$$\lim_{u \rightarrow x^*} \frac{1 - F(xa_u + u)}{1 - F(u)} = 1 - H(x),$$

where $H(x)$ is the distribution function of $\text{GP}(1, \xi)$.

If $X \sim \text{GP}(\sigma, \xi)$, straightforward calculations show that $X - u \mid X > u \sim \text{GP}(\sigma + \xi u, \xi)$ for any $u \in \mathbb{R}$ such that $\sigma + \xi u > 0$, meaning that conditional exceedances above a threshold u also follow a generalized Pareto distribution. This property is termed threshold-stability.

The limiting distribution of threshold exceedances is generalized Pareto as $u \rightarrow x^*$ but, in practice, we must choose a finite threshold in order to have enough exceedances to draw inference. Since the scaling constant a_u is unknown, we have $X \mid X > u \dot{\sim} \text{GP}(\sigma_u, \xi)$. The term $1 - F(u)$ in the denominator is the fraction of points exceeding the threshold, which has a binomial distribution.

Threshold selection is subtle and it is common to select a high percentile of the data, say the 95% value, as the threshold, even if this is asymptotically incorrect, as in this case $k/n \rightarrow 0$ as $n \rightarrow \infty$. Most approaches for threshold selection rely on properties of the generalized Pareto distribution (moments, threshold-stability) to determine a

region within which the asymptotic distribution fits the data well and the parameter estimates are stable.

Below, we focus on recent graphical selection tools for likelihood based inference; mixture models are reviewed in Scarrott & MacDonald (2012) and are available in the package `evmix`.

3.1 Robust selection

The extreme value distributions have unbounded influence functions and outliers can strongly affect the estimate of the shape. Dupuis (1999) proposes an optimal B -robust estimator of the generalized Pareto parameters. Points that are outlying or for which the fit is poor are downweighted; if the weights for the largest observations are very low, this suggests that the threshold is too low. While there is no guarantee that observations that were simulated from a generalized Pareto distributed would not be downweighted, systematic downweighting of the largest exceedances may be indicative of poor fit.

3.2 Threshold stability plots

Consider a sequence of ordered candidate thresholds $u_1 < \dots < u_k$; one of the most widely used tools for threshold selection is the threshold-stability plots of Davison & Smith (1990). These show the point estimates of the shape ξ and the modified scale $\sigma_{u_i} - \xi u_i$, which should be constant for any threshold $u_j > u_i$ assuming that the generalized Pareto above u_i holds exactly. In addition to the point estimates, the asymptotic pointwise 95% Wald confidence intervals are displayed; the standard errors are obtained from the observed information matrix. While these are displayed by many packages, notably `extRemes`, `ismev` and `evd`, `mev` allows you to use profile likelihood based confidence intervals, which typically offer better coverage and capture some of the asymmetry of the distribution. The `mev` package functions for producing threshold stability plots are `tstab.gpd` and `tstab.pp` for respectively the generalized Pareto and Poisson process likelihoods.

Parameter stability plots can be difficult to interpret because the confidence intervals are pointwise rather than simultaneous (each fit likewise uses an overlapping portion of the data). The plots also ignore changes in the estimated parameters due to the penultimate approximation.

3.3 Wadsworth's diagnostics: white noise process and simultaneous threshold stability plots

The problem with the threshold stability plots lies in the point-wise nature of the estimate. Assuming a superposition of k Poisson processes, Wadsworth (2016) derives the limiting distribution of the maximum likelihood estimators from the Poisson

process for overlapping windows as the number of windows $k \rightarrow \infty$ and $n_k \rightarrow \infty$. The joint asymptotic Gaussian distribution allows Wadsworth (2016) to propose two additional diagnostics: a white noise sequence of differences in estimates of the shape, standardized to have unit variance. The variables $\xi_i^* = (\hat{\xi}_{u_{i+1}} - \hat{\xi}_{u_i}) / \{(I_{u_{i+1}}^{-1} - I_{u_i}^{-1})_{\xi, \xi}^{1/2}\}$, where I_{u_i} is the Fisher information of the Poisson process likelihood for exceedances above u_i , should form a white-noise sequence of independent variables centered around the origin; systematic deviations are indicative of inadequacy. To formally test the hypothesis, a likelihood ratio test can be used assuming a simple alternative, namely a single change point at threshold u_j . The null hypothesis is $H_0 : \xi_i^* \stackrel{\text{iid}}{\sim} \text{No}(0, 1)$ for $i = 1, \dots, k-1$ against the alternative $H_a : \xi_i^* \sim \text{No}(\beta, \sigma)$ ($i = 1, \dots, j-1$) and $\xi_i^* \sim \text{No}(0, 1)$ for $j, \dots, k-1$. This alternative is motivated by results on model misspecification (White, 1982), which suggest that the asymptotic distribution may still be Gaussian, but with a different mean and variance. This can be used to automate threshold selection, by picking the smallest threshold for which the P -value is above the level α . The function `W.diag` returns diagnostics plots (for the likelihood ratio statistic path, the white noise process and threshold stability along with asymptotic simultaneous confidence intervals) for non-homogeneous Poisson process model and for the bivariate exponential and the over a sequence of thresholds, specified using `q1`, `q2` and `k`. The argument `M` is a tuning parameter that can be chosen in a way such that the parameters of the non-homogeneous Poisson process likelihood coincide with those of the generalized extreme value distribution for blocks of size m ; see Coles (2001) to this effect.

A main criticism of the proposals of Wadsworth (2016) is their lack of robustness. For the asymptotic result to be approximately valid, the number of thresholds must be large, which implicitly requires large samples for each superposed point process. Moreover, the estimated difference in Fisher information matrices often fails to be positive definite in practice. The procedure is highly sensitive to the choice of k . Changing the set of thresholds \mathbf{u} under consideration leads to potentially completely different parameter estimates being chosen by the automated procedure.

3.4 Changepoint tests based on penultimate approximation

Let $F(x)$ denote a thrice-differentiable distribution function with endpoint x^* and density $f(x)$. Define the reciprocal hazard function $r(x) = \{1 - F(x)\} / f(x)$. The existence of the limit $\xi_\infty = \lim_{n \rightarrow \infty} s'\{b_n\}$ is necessary and sufficient for convergence to an extreme value distribution and Smith (1987) shows that there exists y such that

$$\frac{1 - F\{u + xr(u)\}}{1 - F(u)} = \{1 + xr'(y)\}_+^{-1/r'(y)}, \quad u < y < u + xr(u),$$

unless $r'(x)$ is constant. The penultimate shape parameter for the generalized Pareto distribution is $r'(u)$, but the true shape parameter lies between $r'(u)$ and ξ_∞ .

When we fit the limiting parametric models to finite samples, maximum likelihood

estimates of the shape parameter will be closer to their penultimate counterparts than to the limiting value and we can expect them to vary as we increase the threshold.

Northrop & Coleman (2014) adapt the idea of Wadsworth & Tawn (2012) and fit a generalized Pareto model with piecewise constant shape to k different thresholds; continuity constraints at the thresholds impose $k - 1$ restrictions on scale parameters, so the model only has $k + 1$ parameters. A score test can be used to test the hypothesis of equal shape and it only requires evaluation of the model under the null hypothesis that a generalized Pareto distribution is valid above all thresholds. A diagnostic plot is obtained by plotting P -values against threshold. One can then choose to take, e.g., (a) the lowest threshold at which the P -value is non-significant, or (b) the lowest threshold at which the P -values for all higher thresholds are non-significant: under the null hypothesis, there is an $\alpha\%$ probability of rejection at any given threshold. The function `NC.diag` computes the P -value of the score test as a function of the threshold.

3.5 Extended generalized Pareto

Papastathopoulos & Tawn (2013) propose three extended generalized Pareto distributions: for example, the third extended generalized Pareto model has distribution function $\{1 - (1 + \xi x/\sigma)_+^{-1/\xi}\}^\kappa$ for $x > 0$ and $\kappa > 0$. Each family reduces to the generalized Pareto when the additional parameter $\kappa = 1$ and share the same tail index ξ , the extended generalized Pareto provide more flexibility for modelling departures from the limiting form. Standard parameter stability plots can be used to find a region in which $\kappa \approx 1$ and the shape parameter stabilizes. The additional parameter gives flexibility for modelling departures from the limiting distribution and the hope is one can fit to exceedances over lower threshold and increase the number of points to which the distribution is fitted. The stability plots, obtained through `tstab.egp`, suffer from the same caveats as classical diagnostics.

Chapter 4

Simulation

The original goal of the `mev` package was to implement the algorithm of Dombry et al. (2016) in order to perform exact simulation from max-stable vectors. Since then, other algorithms for simulating from standard R -Pareto processes and generalized R -Pareto processes have been implemented. These are based on accept-reject method or composition sampling.

Let's start by simulating from a simple max-stable random vector from the negative logistic model. The margins are unit Frechet, and we can verify this by pooling the observations and fitting a generalized extreme value distribution to the sample.

```
library(mev)
set.seed(1234)
samp <- rmev(n = 1000, d = 5, param = 0.1, model = "neglog")
fit.gev(c(samp), show = FALSE)$estimate
```

```
##      loc      scale      shape
## 0.994592 1.001284 1.035500
```

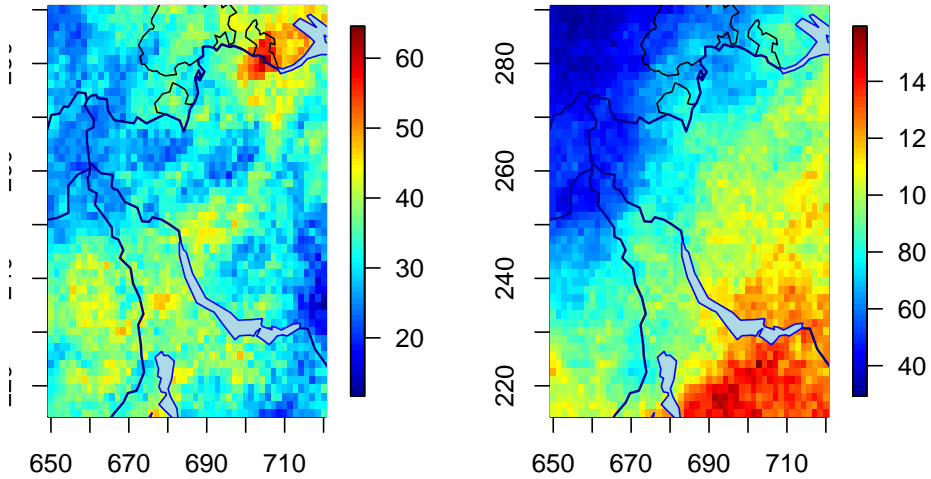
We simulate a spatial generalized R -Pareto process.

```
lon <- seq(650, 720, length = 50)
lat <- seq(215, 290, length = 50)
# Create a grid
grid <- expand.grid(lon, lat)
coord <- as.matrix(grid)
dianiso <- distg(coord, 1.5, 0.5)
sgrid <- scale(grid, scale = FALSE)
# Specify marginal parameters `loc` and `scale` over grid
eta <- 26 + 0.05*sgrid[,1] - 0.16*sgrid[,2]
```

```

tau <- 9 + 0.05*sgrid[,1] - 0.04*sgrid[,2]
# Parameter matrix of Huesler--Reiss
# associated to power variogram
Lambda <- ((dianiso/30)^0.7)/4
# Simulate generalized max-Pareto field above u=50
set.seed(345)
simu1 <- rgparp(n = 1, thresh = 50, shape = 0.1, riskf = "max",
               scale = tau, loc = eta, sigma = Lambda, model = "hr")
# The same, but conditional on an exceedance at a site
simu2 <- rgparp(n = 1, thresh = 50, shape = 0.1, riskf = "site",
               siteindex = 1225, scale = tau, loc = eta,
               sigma = Lambda, model = "hr")
#Plot the generalized max-Pareto field
par(mfrow = c(1,2))
fields::quilt.plot(grid[,1], grid[,2], simu1, nx = 50, ny = 50)
SpatialExtremes::swiss(add = TRUE)
fields::quilt.plot(grid[,1], grid[,2], simu2, nx = 50, ny = 50)
SpatialExtremes::swiss(add = TRUE)

```



```

# Value at conditioning coordinate should be greater than 50
simu2[1225]

```

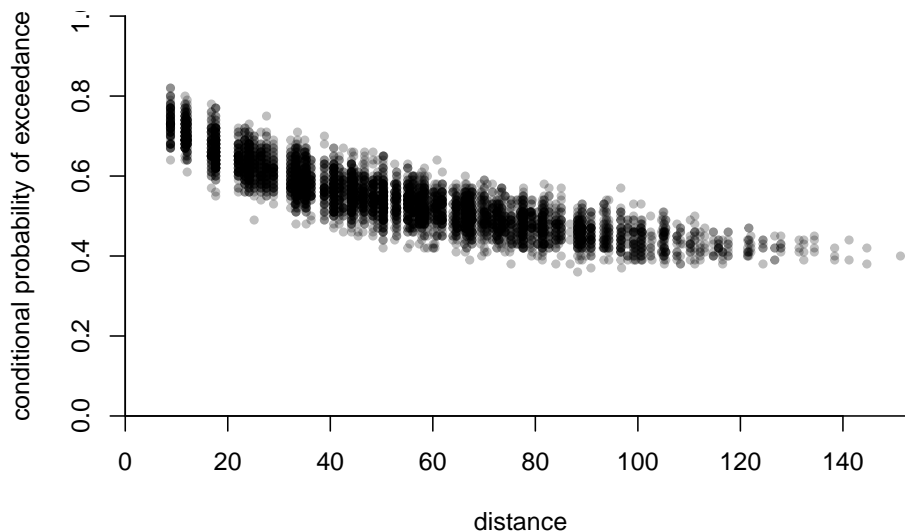
```
## [1] 84.81661
```

The code snippet below fits a Brown–Resnick model with power variogram to simulated data from the same model (based on more than one replicate). The parameters are estimated by minimizing the squared distance between empirical cloud of pairwise conditional probability of exceedance and the theoretical one. We include geometric anisotropy in the analysis.


```

lon <- seq(650, 720, length = 10)
lat <- seq(215, 290, length = 10)
# Create a grid
grid <- expand.grid(lon,lat)
coord <- as.matrix(grid)
dianiso <- distg(coord, 1.5, 0.5)
sgrid <- scale(grid, scale = FALSE)
# Specify marginal parameters `loc` and `scale` over grid
eta <- 26 + 0.05*sgrid[,1] - 0.16*sgrid[,2]
tau <- 9 + 0.05*sgrid[,1] - 0.04*sgrid[,2]
# Parameter matrix of Huesler--Reiss
# associated to power variogram
Lambda <- ((dianiso/30)^0.7)/4
# Simulate generalized max-Pareto field above u=50
set.seed(345)
simul <- rgparg(n = 1000, thresh = 50, shape = 0.1, riskf = "max",
               scale = tau, loc = eta, sigma = Lambda, model = "hr")
extdat <- extremo(dat = simul, margp = 0.9, coord = coord,
                 scale = 1.5, rho = 0.5, plot = TRUE)

```



```

# Constrained optimization
# Minimize distance between extremal coefficient from fitted variogram
mindistpvario <- function(par, emp, coord){
alpha <- par[1]; if(!isTRUE(all(alpha > 0, alpha < 2))){return(1e10)}
scale <- par[2]; if(scale <= 0){return(1e10)}
a <- par[3]; if(a<1){return(1e10)}

```

```

rho <- par[4]; if(abs(rho) >= pi/2){return(1e10)}
semivariomat <- mgp::power.vario(distg(coord, a, rho), alpha = alpha, scale = scale)
  sum((2*(1-pnorm(sqrt(semivariomat[lower.tri(semivariomat)]/2))) - emp)^2)
}
# constrained optimization for the parameters
hin <- function(par, ...){
  c(1.99-par[1], -1e-5 + par[1],
    -1e-5 + par[2],
    par[3]-1,
    pi/2 - par[4],
    par[4]+pi/2)
}
opt <- alabama::auglag(par = c(0.5, 30, 1.5, 0.5),
  hin = hin, control.optim = list(parscale = c(0.5, 30, 1.5, 0.5),
  fn = function(par){
    mindistpvario(par, emp = extdat[, 'prob'], coord = coord)})

```

```

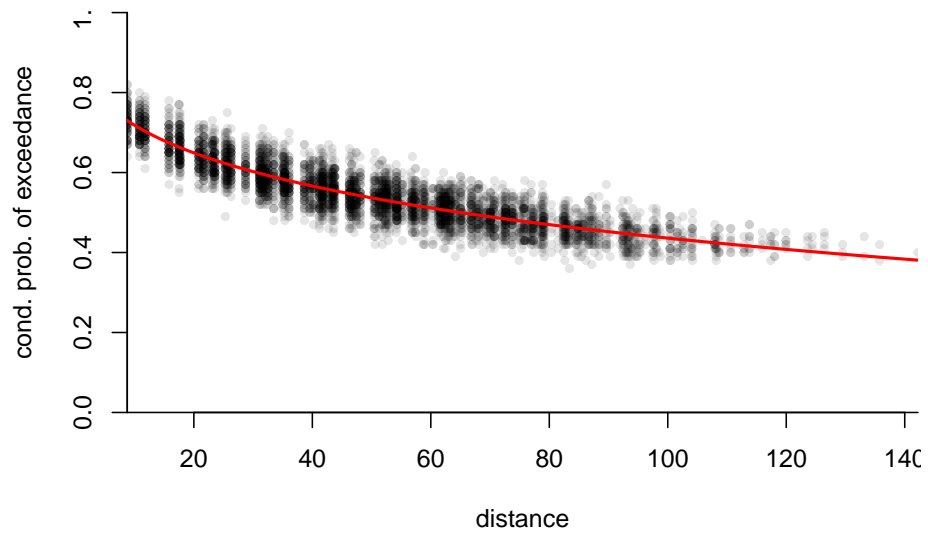
## Min(hin): 0.49999
## Outer iteration: 1
## Min(hin): 0.49999
## par: 0.5 30 1.5 0.5
## fval = 77.37
##
## Outer iteration: 2
## Min(hin): 0.3992261
## par: 0.668296 74.8862 1.39923 0.556808
## fval = 6.52
##
## Outer iteration: 3
## Min(hin): 0.3992502
## par: 0.668285 74.8855 1.39925 0.556807
## fval = 6.52
##

```

```

stopifnot(opt$kk1, opt$kk2)
# Plotting the extremogram in the deformed space
distfa <- distg(loc = coord, opt$par[3], opt$par[4])
plot(c(distfa[lower.tri(distfa)]), extdat[,2], pch = 20,
  col = scales::alpha(1,0.1), yaxs = "i", xaxs = "i", bty = "l",
  xlab = "distance", ylab = "cond. prob. of exceedance", ylim = c(0,1))
lines(x = (distvec <- seq(0,200, length = 1000)), col = 2, lwd = 2,
  2*(1-pnorm(sqrt(power.vario(distvec, alpha = opt$par[1], scale = opt$par[2])/2)))

```



Bibliography

- Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. London: Springer-Verlag.
- Davison, A. C. & Smith, R. L. (1990). Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society. Series B. (Methodological)*, 52(3), 393–442.
- Dombry, C., Engelke, S., & Oesting, M. (2016). Exact simulation of max-stable processes. *Biometrika*, 103(2), 303–317.
- Dupuis, D. (1999). Exceedances over high thresholds: A guide to threshold selection. *Extremes*, 1(3), 251–261.
- Northrop, P. J. & Coleman, C. L. (2014). Improved threshold diagnostic plots for extreme value analyses. *Extremes*, 17(2), 289–303.
- Papastathopoulos, I. & Tawn, J. A. (2013). Extended generalised Pareto models for tail estimation. *Journal of Statistical Planning and Inference*, 143(1), 131–143.
- Scarrott, C. & MacDonald, A. (2012). A review of extreme-value threshold estimation and uncertainty quantification. *REVSTAT – Statistical Journal*, 10(1), 33–60.
- Smith, R. L. (1987). Approximations in extreme value theory. Technical Report 205.
- Wadsworth, J. L. (2016). Exploiting structure of maximum likelihood estimators for extreme value threshold selection. *Technometrics*, 58(1), 116–126.
- Wadsworth, J. L. & Tawn, J. A. (2012). Likelihood-based procedures for threshold diagnostics and uncertainty in extreme value modelling. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(3), 543–567.
- White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1), 1–25.