A modeler’s guide to extreme value software

Léo R. Belzile, Christophe Dutang, Paul J. Northrop, and Thomas Opitz

Abstract

This review paper surveys recent development in software implementations for extreme value analyses since the publication of Stephenson and Gilleland (2006) and Gilleland et al. (2013). We provide a comparative review by topic and highlight differences in existing numerical routines, along with listing areas where software development is lacking. The online supplement contains two vignettes comparing implementations of frequentist and Bayesian estimation of univariate extreme value models.

This version of the article has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature’s AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: https://doi.org/10.1007/s10687-023-00475-9

1. Introduction

Extreme value analysis has seen strong development over the years. While software development typically lags behind methodological developments due in part to lack of recognition of the effort needed to provide reliable software, reproducibility requirements and individual efforts have led to a growth in the coverage of statistical methods. Many procedures developed in the last decades are now available, but the diversity of numerical implementations complicates somewhat the choice of routine to adopt.

Our intention, rather than to solely provide a catalog of existing software, is to discuss and compare existing implementations of statistical methods and to highlight numerical issues that are of practical importance yet are not typically discussed in theoretical or methodological papers. Our work also provides an update to the reviews of Stephenson and Gilleland (2006); Gilleland et al. (2013); Gilleland (2016) by including the most recent software development.

Given its ongoing popularity, we focus on implementations using the R programming language, unless stated otherwise. The Comprehensive R Archive Network (CRAN) Task View on Extreme Value Analysis (Dutang, 2023) provides an extensive list of package functionalities organized by topics; we follow this approach and broadly separate implementations into univariate, multivariate and functional extremes rather than present functionalities package by package. Using the RWsearch package (Kiener, 2022), we automated the process of searching for extreme-related packages on the CRAN and inspected all of the packages that have “extreme value” or “peak over threshold” as keywords in the package description. Additional searches were done for unpublished packages.

As the software landscape evolves quickly, our review is but a snapshot in time. Indeed, maintenance of R packages on the CRAN requires dedicated efforts given the increased number of checks and the relatively short time granted to correct inconsistencies signaled by these checks in order to avoid removal.

2. Univariate extremes

2.1. Asymptotic theory for univariate extremes

The starting point for univariate extreme value analysis is the extremal types theorem: let \( Y_i, i = 1, 2, \ldots \) be independent and identically distributed random variables with distribution function \( F \). If

---

*HEC Montréal, Department of Decision Sciences, HEC Montréal (leo.belzile@hec.ca)
†CEREMADE, CNRS and Université Paris Dauphine - PSL, Paris, France
‡Department of Statistical Science, University College London, United Kingdom
§Biostatistics and Spatial Processes, INRAE, Avignon, France
there exist normalizing sequences \( \{a_n, b_n\}_{n \in \mathbb{N}} \) satisfying \( a_n > 0 \) and \( b_n \in \mathbb{R} \) such that, as \( n \) goes to infinity, the limit distribution of the rescaled sample maximum is non-degenerate, then

\[
\lim_{n \to \infty} \Pr\left( \frac{\max_{i=1}^{n} Y_i - b_n}{a_n} \leq x \right) = \begin{cases} 
\exp \left\{ \left(1 + \frac{x - \mu}{\sigma} \right)^{-1/\xi} \right\}, & \xi \neq 0, \\
\exp \left\{ - \frac{x - \mu}{\sigma} \right\}, & \xi = 0,
\end{cases}
\]

(1)

where \( x_* = \max(x, 0) \). The right-hand side of Equation (1) is the distribution function of the generalized extreme value (GEV) distribution with location parameter \( \mu \in \mathbb{R} \), scale parameter \( \sigma \in \mathbb{R}_+ \), and shape parameter \( \xi \in \mathbb{R} \), with support \( \{x \in \mathbb{R} : \xi(x - \mu)/\sigma > -1\} \). For historical reasons, the distribution is categorized based on the sign of \( \xi \) in so-called “domains of attraction”. If \( \xi < 0 \), the distribution has a bounded upper tail, \( \xi = 0 \) leads to an exponential “light” tail and \( \xi > 0 \) to a “heavy tail” with polynomial decay and with finite moments only of order \( r < 1/\xi \).

If the extremal types theorem holds for a distribution \( F \), then we can equivalently consider conditional exceedances of \( Y \sim F \) above a threshold \( u \), as there exists \( a_u > 0 \) such that

\[
\lim_{a \to -x^*} \frac{\Pr(a^{-1}Y > x + u)}{\Pr(a^{-1}Y > u)} = G(x),
\]

(2)

where \( x^* = \sup\{x : F(x) < 1\} \) is the upper endpoint of \( F \) and

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

(3)

with \( \sigma_u = \sigma + \xi(u - \mu) \). The right-hand side of Equation (3) is the survival function of the generalized Pareto distribution with scale \( \sigma_u \) and shape \( \xi \in \mathbb{R} \). The unconditional distribution of \( F \) above \( u \) is \( \Pr(Y > x + u) \approx G(x) \Pr(Y > u) \). The probability of exceedance above the threshold is typically estimated empirically based on a binomial distribution. The threshold may be either a fixed value or an observation. An equivalent statement of the extremal types theorem is in terms of a point process representation, from which different likelihoods can be derived; see Coles (2001, Chapter 7) for more details.

2.2. Maximum likelihood estimation

Let \( \theta \) denote the \( p \)-vector of parameters of the extreme value model under consideration, e.g., \( \theta = (\mu, \sigma, \xi)^T \) for the generalized extreme value distribution. We can approximate the log likelihood \( \ell(\theta) \) by taking the limiting relations of, e.g., Equations (1) and (2), as exact for the maximum of a finite block of \( m \) observations or for exceedances above a large quantile \( u \); the unknown normalizing constants \( a_u, b_u \), etc., are absorbed by the location and scale parameters. If users have access to the full data (as opposed to say only threshold exceedances), they could choose to model extremes using either block maxima or peaks over threshold: even in the independent and identically distributed scenario, either method may be more suitable (Bücher and Zhou, 2021). Readers wishing to learn more about likelihood-based methods in the context of extremes are referred to Coles (2001).

Optimization: Likelihood-based inference for extreme value distributions is in principle straightforward, even if there is no closed-form solution for the maximum likelihood estimators (MLE). Properties of maximum likelihood estimators imply that the gradient of the log likelihood \( \partial \ell(\theta)/\partial \theta \) must be zero when evaluated at the MLE unless \( \xi = -1 \). Constrained gradient-based optimization algorithms are logical choices for finding the MLE, as the support translates into nonlinear inequality constraints: for example, when fitting a generalized extreme value distribution to a sample of block maxima \( z_1, \ldots, z_n \), one must impose \( \mu, \sigma, \xi : \sigma \xi (z_i - \mu) \geq 0 \), which depends on the maximum observation if \( \xi < 0 \) and on the minimum if \( \xi > 0 \). Many numerical implementations of the log likelihood simply return very large finite values for parameter combinations outside of the support, which can impact the convergence of gradient-based optimization routines: the user is invited to check convergence of whichever software is employed. Even then, the solution returned may not be a global maximum. For example, Figure 1 shows the conditional log likelihood surface for an inhomogeneous Poisson process model, obtained by fixing the scale. The feasible region is defined by a hyperbola and features two local maxima; depending on the starting value, a gradient algorithm would converge to different values.
Numerical implementation: Particular attention must be paid to numerical overflow when implementing the likelihood, score and information matrix of the generalized extreme value distribution, especially for terms of the form \( \log(1 + \xi x) \) when \( \xi \to 0 \) for the information and cumulants. For example, the entries of the expected information matrix for the shape, \( \mathcal{I}_\xi = f(\xi)/\xi^4 \) (Prescott and Walden, 1980), and the limit as \( \xi \to 0 \) is well-defined, but this expression is numerically unstable when \( \xi = 0 \). High precision functions such as \( \log(1+p) \) can be used to alleviate this somewhat, but interpolation of the cumulants based on Taylor series expansions around \( \xi \approx 0 \) is nevertheless recommended.

Dimension reduction: We can sometimes deploy dimension reduction strategies to facilitate numerical optimization. For the generalized Pareto distribution, Grimshaw (1993) uses a profile likelihood to reduce the problem to a one-dimensional optimization. This is arguably one of the safest maximum likelihood optimizations for the generalized Pareto, the inhomogeneous Poisson point process of exceedances and the \( \xi \)-largest order statistics. Thus, when \( \xi \leq -1 \) and check that the solution does not lie on the boundary of the parameter space: for the generalized extreme value distribution, the conditional maximum likelihood estimator when \( \xi = -1 \) is \( \hat{\mu}(x_1, \ldots, x_n) = \bar{x} \), the sample mean, and \( \hat{\sigma}_x(x_1, \ldots, x_n) = \max_i x_i - \bar{x} \). Similarly, for the generalized Pareto distribution, \( \hat{\sigma}_{\xi=1}(x_1, \ldots, x_n) = \max_i x_i \). For the likelihood of the \( r \)-largest order statistics

\[ c \left\{ 1 + \xi \left( \frac{\mu - \mu}{\sigma} \right) \right\}^{-1/\xi} \approx n_u. \]

We can thus fit a generalized Pareto distribution to threshold exceedances, whose maximum likelihood estimates we denote \( (\hat{\sigma}_u, \hat{\xi}) \), and then use as starting values for the point-process optimization routine

\[ \mu_0 = u - \sigma_0(n_u/c)^{-\xi} - 1/\xi, \quad \sigma_0 = \hat{\sigma}_u(n_u/c)^{\xi}, \quad \xi_0 = \hat{\xi}. \]

Regularity conditions and implications: Moments of some of the \( k \)th order derivatives of the log likelihood of extreme value distributions exist only if the shape \( \xi > -1/k \). Thus, when \( \xi \leq -1 \), the MLE does not solve the score equation. The likelihood functions for the generalized extreme value and the generalized Pareto, the inhomogeneous Poisson point process of exceedances and the \( r \)-largest observations are unbounded if \( \xi < -1 \), as there exists a combination of parameters that lead to infinite log likelihood values. This means one should restrict the parameter space \( \mathcal{S} \) to \( \mathcal{S} \cap \{ \xi : \xi \geq -1 \} \) and check that the solution does not lie on the boundary of the parameter space: for the generalized extreme value distribution, the conditional maximum likelihood estimator when \( \xi = -1 \) is \( \hat{\mu} = \bar{x} \), the sample mean, and \( \hat{\sigma} = \max_i x_i - \bar{x} \). Similarly, for the generalized Pareto distribution, \( \hat{\sigma}_{\xi=-1}(x_1, \ldots, x_n) = \max_i x_i \).
fitted to vectors of size \(r\),
\[
\hat{\xi}_{-1} = \frac{1}{r} \sum_{i=1}^{r} (x_{i}-\bar{x}_{(n-r+1)})/r, \quad \hat{\mu}_{-1} = \max_{i} x_{i} - \hat{\xi}_{-1},
\]
where \(\bar{x}_{(n-r+1)}\) is the mean of the \(r\)-largest observations.

The (lack of) existence of cumulants also impacts the calculation of standard errors, as elements of the Fisher information matrix are defined only if \(\xi > -1/2\). Most software implementations compute standard errors based on the numerically observed inverse Hessian matrix obtained via finite differences, but these are misleading if \(\xi \in (-1, -1/2]\) (Smith, 1985).

2.2.1 Case study

There is a plethora of implementations for univariate extremes, so we performed some sanity checks for various implementations of maximum likelihood estimation routines and parametric models. Specifically, we verified that density functions are non-negative and evaluate to zero outside of the domain of the distribution, and that distribution functions are non-decreasing and map to the unit interval. Certain packages have or had incorrect implementations of density and distribution functions; since authors were notified and the corresponding packages may get updated soon, we do not list such implementations here but only report them in the online supplementary material.

To assess the quality of the optimization routines for extreme value distributions, we simulated exceedances and block maxima from parametric distributions with varying tail behaviors. We compared the maximum likelihood estimates returned by default estimation procedures for different packages for simulated data, checking that the log likelihood value returned is a global optimum by comparing with other implementations and the gradient evaluated at the value is approximately zero whenever \(\xi > -1\). The purpose of the exercise was to check the reliability of the numerical routines for a range of sample sizes. When systematic differences in maximum log likelihood values and/or parameter estimates arose compared to other packages, they are often attributable to poor starting values, incorrect implementation of the density function, lack of handling of boundary constraints or to problems with optimization algorithms.

As an illustration, we generated 1000 samples of size \(n = 500\) from a gamma distribution with shape 3 and scale 2 and considered exceedances above the theoretical 0.95 quantile of this distribution, leading to an average of 25 exceedances. We then fitted by maximum likelihood the parameters of the generalized Pareto distribution. The dot plots in the left panel of Figure 2 show that the sampling distribution of the shape parameter is quite dispersed. The astute reader may notice some oddities: the QRM package has unexpected small spread and a positive bias for estimation of \(\xi\), different
Figure 2: Left: sampling distribution (dot plots) of generalized Pareto shape parameter estimates according to different packages. Right: absolute value of log gradient $\frac{\partial \ell}{\partial \xi}$ evaluated at the maximum likelihood estimator $(\hat{\sigma}, \hat{\xi})$ on the log-scale with base 10. Results for samples for which the numerical routines failed to converge are omitted.

From other packages because it fails more often when $\xi$ is negative due to poor starting values. Likewise, both `ercv` and `extRemes` (Gilleland and Katz, 2016) fits have noticeable point masses at $\xi = 0$, suggesting something is amiss as this value should be returned with probability zero. We can also see this by inspecting differences between the returned log likelihood values and the actual maximum log likelihood: `ercv` returns 9% of the time values that are more than 0.05 units away, `extRemes` 3.5% and 1.5% for `tea` and `eva` from the maximum. The maximum likelihood estimator of the shape cannot be less than $-1$, but only `SpatialExtremes` (Ribatet, 2022) and `mev` correctly return $-1$ by default and `Renext` when `shapeMin = -1.0`.

The right panel of Figure 2 shows the distribution of the gradient of the log likelihood of the generalized Pareto distribution evaluated at the maximum likelihood estimate over all replicates for the shape parameter, omitting non-zero gradients attributable to boundary cases $\xi < -1$: non-zero gradients are in most cases due to differences in numerical tolerance, as the differences in log likelihood relative to the maximum are negligible. It also suggests that convergence for most routines is based on log likelihood differences being small rather than gradients being zero.

The optimization routines for the generalized extreme value distribution yielded similar behavior and nearly all packages gave identical results. However, we noticed that some packages fare poorly when location or scale parameters are orders of magnitude larger than scaled components: since the generalized extreme value distribution is a location-scale family, scaling the data before passing them to the routine and back-transforming the MLE after the optimization may solve such issues.

### 2.3. Regression modelling

Most data encountered display various forms of nonstationarity, including trends, seasonality and covariate effects, which the extreme value distributions cannot capture without modification. One can thus consider regression models in which the parameters of the extreme value distributions are functions of covariates or vary smoothly in space or time. These parameters may be suitably transformed via a link function to ensure that the functions satisfy the usual range or positivity constraints. If we assume independent observations, then maximum likelihood estimates, standard errors, etc. are obtained as before by maximizing the log likelihood function, which is now a function of the regression coefficients and of other parameters arising in the nonstationary formulation of the extreme value distribution. In models with a relatively large number of parameters, it becomes useful to include an additive penalty term in the log likelihood: for example, generalized additive models for the parameters include smooth functions (`smooths` in short) via basis function representations (e.g.,

---

**Author's final, peer-reviewed manuscript**

*post-print, version of July 25, 2023*
parameters through maximum likelihood. The evgam fitting the penalty strength and shape — these are estimated simultaneously with all of the other parameters of the hyperparameters (e.g., variance and autocorrelation of regression coefficients) control-egional distributions to marginalize out the regression coefficients using Laplace’s method to obtain esti-

Northrop et al., 2016). Chavez-Demoulin and Davison (2005) use an orthogonal reparametrization (Eastoe and Tawn, 2009). It may be tempting to use directly the likelihood of eq. (4) instead (see covariates to ensure that the threshold stability property, which is used for extrapolation, is not lost recommends to pay special attention to the parametrization of the scale and shape functions with

In the Bayesian paradigm, the likelihood of a random sample $Y$ is combined with prior distributions for the model parameters $\theta = (\theta_1, \ldots, \theta_m)\top \in \Theta$, with prior density $p(\theta)$; we use the generic notation $p(\cdots)$ for various conditional and unconditional densities and mass functions. The distribution of the data given the parameter vector, $p(Y | \theta)$ is encoded by the likelihood function $\exp(\ell(\theta; Y))$, while the
posterior distribution,

$$p(\theta \mid Y) = \frac{p(Y \mid \theta)p(\theta)}{\int p(Y \mid \theta)p(\theta)d\theta}, \tag{5}$$

is proportional, as a function of \( \theta \), to the product of the likelihood and the priors in the numerator. The posterior density \( p(\theta \mid Y) \) usually does not correspond to any well-known distribution family and the integral appearing in the denominator of Equation (5) is therefore intractable in general. Posterior inferences about the components of \( \theta \) further involve marginalizing out the other components. For instance, to obtain the posterior density \( p(\theta_1 \mid Y) \) of the first parameter in \( \theta \), we have to evaluate the \((m - 1)\)-dimensional integral \( \int p(\theta_1 \mid Y)d(\theta_2, \ldots, \theta_m) \). Most of the field of Bayesian statistics revolves around the creation of algorithms that circumvent the calculation of the normalizing constant (or else provide accurate numerical approximation of the latter) or that allow for marginalizing out all parameters except for one.

Rather than a point estimator of the parameter vector, the target of Bayesian inference is the whole posterior distribution. The majority of estimation algorithms are simulation-based, and their typical output is an (approximate) sample drawn from the posterior distribution \( p(\theta \mid Y) \), from which any functional of interest can be estimated by Monte Carlo methods. Of particular interest is the posterior predictive distribution, which is obtained by simulating new observations from the response model by forward-sampling from \( p(Y \mid \theta^{(b)}) \) one new observation for each draw of \( \theta^{(b)} \) from the posterior.

In simple problems, exact sampling algorithms can provide independent and identical samples from the posterior, but this is the exception rather than the norm. Most of the time, users resort to Markov chain Monte Carlo (MCMC) algorithms for more complex settings: these algorithms admit the posterior distribution as the stationary distribution of a Markov chain with appropriately designed transition probabilities and provide auto-correlated samples from it. Another popular solution is through Laplace approximation for regression models when multivariate Gaussian priors are put on the vector of regression coefficients arising in the latent layer of the model, from which observations are conditionally independent; see the discussion in Section 2.3. In this setting, Laplace approximations give fast deterministic approximation of high-dimensional integrals, which avoids resorting to simulation-based estimation. Laplace approximations are particularly accurate when they are applied twice in a certain nested way, which is known as the integrated nested Laplace approximation (INLA, Rue et al., 2009), implemented in the general INLA package (Martins et al., 2013) offering extreme value functionality for generalized Pareto and generalized extreme value distributions.

Despite the computational overhead associated, the Bayesian paradigm has many benefits, including the capacity to incorporate physical constraints and expert opinion through the prior distributions (Coles and Tawn, 1996). It is easier and more natural to define hierarchical structures for

| package | functions | type | link | par. | model |
|---------|-----------|------|------|------|-------|
| eva     | gevrFit, gpFit | linear | custom | all | GEVR, GP |
| evd     | fgev     | linear | identity | μ | GEV |
| evgam   | evgam   | GAM   | logistic, probit, cloglog | all | GEV, GP, * |
| extRemes | fevd    | linear | identity, log | all | — |
| GEVcdn  | gevcdn.fit | NN   | all | — | GEV |
| ismnev  | gpd.fit, gev.fit | linear | custom | all | GEV, GP |
| ismnev  | gamGPDfit | GAM   | identity, log | σ, ξ | GP |
| texmex  | evm     | linear | identity, log | all | — |
| VGAM    | gev, gp | GAM   | identity, log, power | all | GEV, GP, * |

Table 1: Functionalities for modelling parameters of extreme value distributions using generalized linear models, generalized additive models (GAM) or neural network (NN). Model families supported include generalized extreme value distribution (GEV), generalized Pareto (GP), \( r \)-largest extremes (GEVR) and more general families or special cases of extreme value distributions (*). The column par. denotes the set of parameters which can vary, either all, location (μ), scale (σ) or shape (ξ) parameters.
parameters to pool information. For multivariate and functional extremes, priors can be used for regularization purposes to pool information, for instance across time and space.

2.4.2 Specificity of extremes

Readers wishing to learn more about Bayesian modelling for extreme values are referred to the extensive overview in Stephenson (2016). While Bayesian inference for extreme value models does not differ much from that of general models, additional care is required with prior specification. For example, in order to get a well-defined posterior distribution, improper reference priors such as the maximal data information (MDI) and Jeffreys priors for $\xi$ may need to be truncated (Northrop and Attaides, 2016) to result in proper (i.e., integrable) posterior distributions or else do not yield proper posteriors regardless of the sample size. Martins and Stedinger (2000) proposed using a shifted Beta distribution for $\xi$ to constrain the support of the latter to $[-0.5, 0.5]$. Other popular choices are vague normal priors for location, log-scale and shape parameters, or else penalized complexity priors (Simpson et al., 2017; Opitz et al., 2018). To avoid issues related to the finite and parameter-dependent lower endpoint in the generalized extreme value distribution for $\xi > 0$, the INLA package implements so-called *blended generalized extreme value distribution* that replaces the bounded lower distribution tail with the unbounded one of a Gumbel distribution through a mixture representation (Castro-Camilo et al., 2022).

Table 2 lists packages for Bayesian univariate models, where the ‘covariates’ column lists the parameters which are allowed to depend on covariates ($\text{Loc}$ refers to the location parameter of the generalized extreme value distribution, while threshold refers to the threshold parameter of the generalized Pareto distribution). Three packages, `evdbayes`, `extRemes` and `MCMC4Extremes`, provide MCMC algorithms for extreme value distributions, which implement so-called random walk Metropolis–Hastings steps. The underlying implementation of the MCMC algorithm for the function `posterior` in `evdbayes`, detailed in the user guide, allows for a linear trend in the location parameter. Gamma priors for quantile differences, used for expert prior elicitation, are also provided. Contrary to most implementations, `evdbayes` returns a list of posterior samples and relies on methods implemented in `coda` (Plummer et al., 2006) for diagnostic, summary and plots. The `extRemes` package also has functionalities for computing posterior summaries for univariate extremes through the `fevd` function, which allows users to specify their own priors and proposal distributions, but the sampling is notably slower than in other packages and more cumbersome to set up, as the default values are not adequate in most cases. Linear modelling of the parameters with covariates is also possible, and Bayes factors for comparisons between models are also supported even if the methods used to compute them are not recommended. For all relevant purposes, `MCMC4Extremes` (do Nascimento and Moura e Silva, 2016) is superseded by competitors as the latter have default tuning of proposal standard deviations and more flexible choices of priors. Package `texmex` also includes maximum a posteriori estimation and simulation from the posterior for extreme value distributions (with linear modelling of covariates) via the function `evm`, but only with normal priors. Behind the scenes, the `texmex` implementation uses an independent Metropolis–Hastings step with multivariate Cauchy or normal proposals with location vector and scale matrix based on a normal approximation to the posterior, using maximum a posteriori estimates. This translates into smaller autocorrelation (and thus larger effective sample size) than other package implementations, and it is the fastest of all MCMC implementations.

The data-driven prior proposed by Zhang and Stephens (2009), reputed to give better results than maximum likelihood, is implemented in `mev` and is the default method for Pareto-smoothed importance sampling (Vehtari et al., 2017) from the `loo` package (Vehtari et al., 2020). However, because it uses the data to construct the prior, performance benchmarks alleging superior performances are misleading because of double dipping.

The current state-of-the-art method for sampling from the posterior of univariate models in simple analyses without covariates is the `revdbayes` package, which relies on the ratio-of-uniforms method to generate independent samples from the posterior distribution of the models. Use of advanced techniques such as mode relocation, marginal Box–Cox transformations and rotation can drastically improve the efficiency of this accept-reject scheme and make it very competitive. The ratio-of-uniforms method generates independent draws, thus avoiding the need to monitor convergence to the stationary distribution of the Markov chain and removing tuning parameters. The sampling is also an order of magnitude faster than other implementations.
Table 2: Comparison of R packages for Bayesian univariate extreme value modelling. Families: generalized extreme value distribution (1), generalized Pareto distribution (2), inhomogeneous Poisson process (3), order statistics/\textit{r}-largest (4) or custom/other (*). Sampling: random walk Metropolis–Hastings (RWMH), exact sampling ratio-of-uniforms (RU), independent Metropolis–Hastings (IMH); the INLA package uses deterministic Laplace approximations. "PC“ priors refer to penalized complexity priors. All packages, except \texttt{evdbayes}, also provide S3 methods (notably \texttt{plot} and \texttt{summary}). All packages return a matrix of posterior draws.

| package            | function | models | covariates | sampling   | prior choice |
|--------------------|----------|--------|------------|------------|--------------|
| \texttt{evdbayes}  | posterior| 1–4    | loc./thresh | RWMH       | multiple     |
| \texttt{extRemes}  | fevd     | 1–4,*  | all        | RWMH       | custom       |
| \texttt{INLA}      | inla     | 1–2,*  | loc./thresh | –          | PC           |
| \texttt{MCMC4Extremes} | ggev, \ldots | 1–2,* | no         | RWMH       | fixed        |
| \texttt{revdbayes} | rpost    | 1–4    | no         | RU         | custom       |
| \texttt{texmex}    | evm      | 1–2,*  | all        | IMH        | Gaussian     |

While the aforementioned packages are dedicated to extreme value distributions, other popular programming languages could be used even if they would require users to implement likelihood functions themselves. Notably, the Stan programming language (Stan Development Team, 2023) uses Hamiltonian Monte Carlo, a state-of-the-art MCMC method, for simulating samples from the posterior distribution. The latter can easily be combined with multilevel models, but requires implementation of bespoke code for likelihood and priors that are specific to extreme value analysis; sample code is provided online. The Hamiltonian Monte Carlo sampling algorithm leads to rejection due to boundary constraints and leads to incorrect posterior draws for, e.g., the generalized extreme value distribution when \( \xi \approx 0 \); this can be corrected by using a Taylor series approximation. The Matlab package NEVA uses a differential evolution Markov chain algorithm for estimating univariate non-stationary models (Cheng et al., 2014).

Some splicing models, which combine a distribution for the bulk of the data with a generalized Pareto tail, can also be fitted using Markov chain Monte Carlo methods; example includes \texttt{extremix} for the Gamma mixture model of do Nascimento et al. (2012).

2.5. Semiparametric inference for univariate extremes

In the semiparametric approach to extremes, some components of the probability structure are handled through a relatively general (and nonparametric) asymptotic structure, which can be extrapolated towards higher yet unobserved quantile levels, for instance for the purpose of extreme-quantile estimation. The parametric form includes the shape parameter \( \xi \) and potentially second-order regular variation indices, \( \rho \). Caeiro and Gomes (2016) provides a review of many estimators discussed next with an emphasis on the choice of the number of order statistics to keep for inference, which has close ties to threshold selection methods discussed in Section 2.6.

Consider a sample of independent and identically distributed variables \( Y_1, \ldots, Y_n \sim F \) with quantile function \( Q \) and order statistics \( Y_{(1)} \leq \cdots \leq Y_{(n)} \). Assuming that the extremal types theorem holds for \( F \) with positive limiting shape parameter \( \xi > 0 \), we can write the survival function as \( S(x) = x^{-\xi}L_F(x) \) and the quantile function as \( Q(1 - 1/x) = x^\xi L_U(x) \), with \( L_F \) and \( L_U \) slowly varying functions, meaning \( \lim_{t \to \infty} L(tx)/L(x) = 1 \) for any \( t > 0 \) (e.g., Ledford and Tawn, 1996, § 5). Nonparametric estimators of the extreme value index are widespread, most of them variants of the Hill (1975) estimator for positive shape parameters. The Hill estimator is the mean excess value of log-transformed data of the \( k \) largest values,

\[
H_{k,n} = \frac{1}{k} \sum_{j=1}^{k} \log \left( \frac{Y_{(n-j+1)}}{Y_{(n-k)}} \right). \tag{6}
\]

Hill’s estimator is generally computed for a wide range of values of \( k \), which leads to so-called Hill plots \( (k, H_{k,n}) \), \( k = 1, \ldots, n \). A large number of R packages provides functions to estimate (6) and
to make Hill plots such as evir (Pfaff and McNeil, 2018), evmix (Hu and Scarrott, 2018), extremefit (Durrieu et al., 2019), ExtremeRisks (Padoan and Stupfler, 2020), ptsuite (Munasinghe et al., 2019), QRM (Pfaff and McNeil, 2020), Relns (Reynkens and Verbelen, 2023) and tea (Ossberger, 2020).

The performance of the Hill estimator strongly depends on the number of observations kept to estimate the tail index: $H_{k,n}$ has a large variance if $k$ is too small, whereas the Pareto-type tail behavior might not be verified for the selected $k$ largest values if $k$ is too large. The choice of $k$ is typically based either on an empirical rule to find the area where $H_{k,n}$ is “stable” or by minimizing the asymptotic mean squared error (AMSE). A large number of those algorithms to minimize the latter are provided in tea along with the bootstrap methods of Hall and Welsh (1985), Hall (1990), Danielsson et al. (2001), Caeiro and Gomes (2014) and Caeiro and Gomes (2016).

Since the Hill estimator has nondifferentiable sample paths with respect to the threshold value, the choice of threshold is notoriously difficult. Resnick and Stărică (1997) proposed a smoothed version of the Hill estimator based on averaging consecutive estimates via a moving window; these plots are provided in evmix and tea. The random block maximum estimator (Wäger, 2014) in rbm, constructed as a $U$ statistic, has infinitely differentiable sample paths and is thus much less sensitive to the choice of $k$ than most Hill-type estimators. Packages evt0 (Manjunath and Caeiro, 2013) and Relns implement the generalized Hill estimator based on a uniform kernel estimation (Beirlant et al., 1996). evt0 also provides functions for the location-scale invariant version of the Hill estimator introduced by Santos et al. (2006) and the biased-reduced version of Figueiredo et al. (2012), as well as a mixed moment estimator and location invariant alternative. The package extremefit implements the kernel-weighted version of the Hill estimator of Grama et al. (2008); the authors provide an automatic selection procedure for the threshold $u$, with functions to handle these weighted estimators either for user-supplied weights or for weights automatically selected using an adaptive selection.

2.5.1 Moment estimators and other alternative estimators

While maximum likelihood estimation and Hill-type estimators are most commonly used for the shape parameter, other estimators are available and may be more robust in small samples. One such was proposed by Dekkers et al. (1989) and evt0 provides a generalization of the latter by Santos et al. (2006). Since moments of extreme value distributions may not exist if $\xi > 0$, we can consider instead a bijection between the parameter vector $\theta$ and probability weighted moments of the form $E[Y^pF(Y)^q(1 - F(Y))^r]$ for integers $p, q, r$ (Hosking and Wallis, 1987). Another avenue is to match sample linear combinations of order statistics with their theoretical counterparts using (trimmed) $L$-moments (Hosking, 1990). A group of R packages, including lmom (Hosking, 2019), lmomco (Asquith, 2021), TLMoments (Lilienthal, 2022) implement these approaches for a variety of common distributions (as does the Python package lmoments), but some also allow custom distribution functions. extRemes also implements $L$-moments, while RobExtremes (Ruckdeschel et al., 2019) provides robust estimators of the extreme value parameters and laeken (Alfons and Templ, 2013) proposes robust modelling of Pareto data. Package extremeStat (Boessenkool, 2017) includes functionalities to compute extreme quantiles based on $L$-moments estimator.

2.5.2 Quantile, expectile and extremiles

In the heavy-tailed setting, Weissman (1978) proposed estimating the tail quantile at level $1 - p$, $Q(1 - p)$, for small $p$, using the estimator

$$Q_{k,n}^W(1 - p) = Y_{(n-k)} \left\{ \frac{k + 1}{p(n + 1)} \right\}^{H_{k,n}},$$

where $H_{k,n}$ is the Hill estimator eq. (6) of the shape parameter and the threshold is $Y_{(n-k)}$, the $(n-k)$th order statistic. Relns implements the Weissman estimator either specified by the probability level $p$ or by the return period $1/p$. The Weissman-type estimator for the class of estimators proposed by Santos et al. (2006) are provided by evt0, whereas extremefit gives the quantile corresponding to weighted Hill estimator. Bias-corrected versions of the Weissman estimator also exist, yet are seemingly not implemented in software.

Quantiles can be formulated as the solution of an asymmetric piecewise linear loss function. Taking instead an asymmetric quadratic loss function yields expectiles (Newey and Powell, 1987), another risk measure gaining popularity in risk management (Bellini and Di Bernardino, 2017). Many recent work studies their extremal property: on the software side, ExtremeRisks implements the
methodology of Davison et al. (2022); Padoan and Stupfler (2022), including estimation of expectiles using Hill-type estimators, test of equality of tail expectiles and confidence regions for extreme expectiles. An alternative risk measure, the so-called extremile, has been developed recently (e.g., Daouia et al., 2022). An implementation of expectiles of common distributions and of estimators for the heavy-tailed setting is provided in \texttt{Expectrem}, which also allows for the possibility to use bias-reduced tail index estimators.

| package     | estimation   | function       | features |
|-------------|--------------|----------------|----------|
| \texttt{evir}    | —            | \texttt{hill}   | e, p     |
| \texttt{evmix}   | smoothing    | \texttt{hillplot} | p        |
| \texttt{evt0}    | location invariant | \texttt{gh, PORT.Hill} | p, q     |
| \texttt{extremefit} | weighted, time series | \texttt{hill,hill.adapt,hill.ts} | e, p, q, o |
| \texttt{ExtremeRisks} | time series, CI | \texttt{HTailIndex, EBTailIndex} | e, o     |
| \texttt{fExtremes} | —           | \texttt{hillPlot, shaparmHill} | e, p     |
| \texttt{ptsuite} | —           | \texttt{alpha_hills} | e        |
| \texttt{QRM}     | —           | \texttt{hill,hillPlot} | e, p     |
| \texttt{rbm}     | random block | \texttt{rbm,rbm.plot} | e, p     |
| \texttt{ReIns}   | conditional, censoring | \texttt{(c)Hill, (c)genHill, crHill,...} | e, p     |
| \texttt{tea}     | smoothing    | \texttt{althill, avhill} | p        |

Table 3: Main functionalities of R packages for nonparametric Hill-type estimators of the shape parameter, including functionalities for estimation of the shape or tail index (e), Hill threshold diagnostic plots (p), quantile estimates (q) and other methods (o).

2.6. Threshold selection

Many methods are driven by analyses of the most extreme observations. In the univariate case, these are the $k$ largest order statistics or, equivalently, observations that exceed a threshold $u$ as presented in the previous section. The underlying theory considers limiting behavior as the threshold increases. In practice, a suitably high threshold is set empirically, balancing the bias from using a low threshold that violates the theory with statistical imprecision from using a threshold that is unnecessarily high. For information about many of the following methods, see the review of Scarrott and MacDonald (2012). Methods for semiparametric estimators based on variants of Hill’s estimator for the shape were presented in Section 2.5.

2.6.1 Visual threshold selection diagnostics

In a threshold stability plot, point and interval estimates of parameters are plotted against a range of threshold values. A particular example is the Hill plot featured in Section 2.5 (see Table 3 for an overview of available implementations). In the univariate case, the focus is often on the shape parameter $\xi$: we choose the lowest threshold above which we judge the point estimates of $\xi$ to be approximately constant in threshold, bearing in mind statistical uncertainty quantified by the interval estimates. These inferences may be based on the generalized Pareto distribution (3) for threshold exceedances or the inhomogeneous Poisson process model, using a frequentist or Bayesian analysis. In the generalized Pareto case, the threshold-independent scale parameter $\sigma_0 = \sigma_u - \xi u$ is used. In the frequentist case, it is useful to have the option to calculate the intervals using profile likelihoods, because they tend to have better coverage properties than Wald intervals, especially for high thresholds.

If a generalized Pareto distribution with $\xi < 1$ applies at threshold $u$ then the mean excess $E(Y - v | Y > u)$ is a linear function of $v$ for all $v > u$. This motivates the mean residual life (MRL) plot, in which the sample mean of excesses of a range of thresholds are plotted against the threshold, with pointwise confidence intervals superimposed. We choose the lowest threshold above which the plot appears linear. Table 4 summarises the functionality of R packages in terms of these plots.

The \texttt{lmomplot} function in the \texttt{POT} (Ribatet and Dutang, 2022) package can help to identify for which thresholds the sample L-skewness and L-kurtosis of excesses are related as expected under
Table 4: Comparison of R packages for classical visual methods. Stability: function name for a threshold stability plot; models: either generalized Pareto (1), inhomogeneous Poisson process (2) or extended generalized Pareto model of Papastathopoulos and Tawn (2013) (3); profile: whether confidence intervals are computed using the profile likelihood or not; inference: method of inference, either maximum likelihood estimation (MLE) or Bayesian (B); MRL: mean residual life plot, if applicable.

| package   | stability       | models | profile | inference | MRL  |
|-----------|-----------------|--------|---------|-----------|------|
| eva       | gpdDiag         | 1      | yes     | MLE       | mrlPlot |
| evd       | tcplot          | 1,2    | no      | MLE/B     | mrlPlot |
| evir      | shape           | 1      | no      | MLE       | meplot |
| evmix     | tcplot          | 1      | no      | MLE       | mrlPlot |
| extRemes  | threshrange.plot | 1,2    | no      | MLE       | mrlPlot |
| fExtremes | gpdShapePlot,... | 1      | no      | MLE       | mrlPlot |
| ismev     | gpd.fitrange,pp.fitrange | 1,2    | no      | MLE       | mrlPlot |
| mev       | tstab.egp,tstab.gpd | 1,3    | yes     | MLE/B     | automrl |
| POT       | tcplot          | 1      | no      | MLE       | mrlPlot |
| QRM       | xiplot          | 1      | no      | MLE       | MEplot |
| Relns     | 1Dmle           | 1      | —       | MLE       | MeanExcess |
| texmex    | egp3RangeFit,gpdRangeFit | 1,3    | no      | MLE/B     | mrl |
| threshr   | stability       | 1      | yes     | MLE       | —     |

a generalized Pareto distribution. These plots require the use of subjective judgement to select a threshold. More formal methods seek to reduce subjectivity and perhaps introduce a greater degree of automation.

2.6.2 More formal methods

Penultimate models. Formal testing procedures compare the null hypothesis of having a generalized Pareto distribution above a threshold $u$ against an alternative model. Theoretically-justified alternative models can be derived from the penultimate approximation to extremes, either by selecting piecewise constant shape (Northrop and Coleman, 2014) or by using tilting function to provide more general models that should have faster convergence. The models proposed in Papastathopoulos and Tawn (2013) lead to a threshold stability plot for an additional parameter. These approaches are implemented in mev.

Goodness-of-fit diagnostics. One drawback of the threshold stability plot and tests is that they do not entirely indicate whether the tail model fits the data well. Goodness-of-fit diagnostics can thus complement other diagnostics. The eva package (Bader and Yan, 2020) provides multiple testing methods with the Cramér–von Mises and Anderson–Darling criteria and Moran’s tests, all with control for the false discovery rate (Bader et al., 2018). The benefit of this approach, compared to visual diagnostics, is that it does not require user input and is more readily implementable with large multivariate or spatial data sets. The approach of Dupuis (1999), based on examination of the weights attached to the largest observations from the sample and obtained using a robust fitting procedure, can be obtained via mev.

Sequential analysis and changepoints. Parameter estimates obtained by fitting a tail model at multiple consecutive thresholds are dependent because of the non-negligible sample overlap. The mev package provides the method of Wadsworth (2016), which exploits a technique from sequential analysis by fitting a point process over a range of thresholds and building an approximate white noise sequence from the differences between consecutive estimates using their asymptotic covariance matrix, suitably rescaled to be standard normal. The tea package provides the Pearson $\chi^2$ test of normality applied to sequences of differences of scale estimates, following Thompson et al. (2009), while threshold stability plots based on estimates of the coefficient of variation and sequential testing of del Castillo and Padilla (2016) are included in ercv (del Castillo et al., 2019).
Table 5: Overview of formal threshold selection methods and numerical implementations

| type              | methods                                      | package | function        |
|-------------------|----------------------------------------------|---------|-----------------|
| penultimate       | Northrop and Coleman (2014)                  | mev     | NC.diag         |
|                   | Papastathopoulos and Tawn (2013)             | mev     | tstab.egp       |
|                   | Gerstengarbe and Werner (1989)               | tea     | ggplot          |
|                   | Hosking and Wallis (1997)                    | POT     | lmomplot        |
|                   | Bader et al. (2018)                          | eva     | gpdSeqTests     |
| sequential        | Wadsworth (2016)                             | mev     | W.diag          |
|                   | Thompson et al. (2009)                       | tea     | TH              |
|                   | del Castillo and Padilla (2016)              | ercv    | cvplot, thrselect |
| predictive        | Northrop et al. (2017)                       | thresr  | ithresh         |
| mixture           | Hu and Scarrott (2018)                       | evmix   | —               |
|                   | Durrieu et al. (2015)                        | extremefit | paretomix     |
|                   | Naveau et al. (2016)                         | mev     | fit.extgpe      |

**Predictive performance.** The `thresr` package (Northrop et al., 2017) looks at the predictive performance of the generalized Pareto for a binomial-generalized Pareto model fitted using the Bayesian approach. The scheme uses a leave-one-out cross validation scheme for values at a fixed validation threshold \( v \) at or above the range of potential thresholds considered.

**Mixture models.** The generalized Pareto specifies a distribution only for exceedances above a threshold \( u \), but having a model below this threshold may be desirable, with some options enabling automatic threshold selection. The `evmix` package (Hu and Scarrott, 2018) provides implementations of most of the mixture models listed in Scarrott and MacDonald (2012): this includes parametric models for the bulk of the data (for which users can inform threshold selection by looking at the profile likelihood for \( u \)), nonparametric and kernel-based approaches for the data below the threshold. Many such models are discontinuous at the threshold and require choosing a fixed threshold. The `extremefit` package (Durrieu et al., 2019) provides a mixture model implementation with a kernel-based bulk model and adaptive selection rules for the bandwidth parameter. The `mev` package provides the extended generalized Pareto model of Naveau et al. (2016) for modelling rainfall. The extension proposed in Gamet and Jalbert (2022) comes with Julia code.

**Univariate extremes implementations in other programming languages**

While R is arguably the programming language boasting the most software implementations used for extreme value analyses, some basic routines are available elsewhere for estimation of univariate models using maximum likelihood or probability weighted moments: these include the Julia package Extremes, the Matlab EVIM package and the Python packages wafo, pyextremes and scikit-extremes.

3. Multivariate extremes

The lack of ordering of \( \mathbb{R}^D \) leads to multiple definitions of extremes (Barnett, 1976). We focus on componentwise maxima and concomitant exceedances, which lead to the multivariate analog of block maximum and peaks over threshold methods. Another option, structure variables, reduces the data to univariate summaries and can be dealt with using tools presented before.

3.1. Multivariate maxima

Consider an independent and identically distributed sequence of \( D \)-variate random vectors \( \{Y_i\}_{i \geq 1} \), where each vector \( Y_i \) has marginal distribution functions \( F_j (j = 1, \ldots, D) \). By analogy with the univariate case, we consider the random vector of componentwise maxima \( M_n = (M_{n,1}, \ldots, M_{n,D}) \), where \( M_{n,j} = \max(Y_{1,j}, \ldots, Y_{n,j}) \). If there exists sequences of location and scale vectors \( a_n \in \mathbb{R}^D \) and \( b_n \in \mathbb{R}^D \)
such that
\[
\lim_{n \to \infty} \Pr \{ a_n^{-1} (M_n - b_n) \} = G(y),
\]
with non-degenerate limit distribution \( G \), then \( G \) is a multivariate extreme value distribution, or equivalently a max-stable distribution with generalized extreme value distributed margins. Suppose without loss of generality that the normalizing constants are chosen so that the limiting location and scale marginal parameter vectors are \( \mu = 0_D \) and \( \sigma = 1_D \). With \( t(y) \) denoting a transformed vector whose \( j \)th component is \( (1 + \xi_j y_j)^{1/\xi_j} \), the limiting max-stable distribution is
\[
G(y) = \exp \left\{ -D \int_{S_D} \max \left\{ \frac{u}{t(y)} \right\} dH(u) \right\},
\]
where the so-called spectral measure \( H \) is a probability measure on the \( D \)-simplex \( S_D = \{ \omega \in \mathbb{R}^D : \| \omega \|_1 = 1 \} \). The distribution \( H \) must only satisfy the moment constraints \( E(S_j) = 1/D \) for \( S \sim H \): the set of probability measures satisfying these is infinite, unlike in the univariate case. The \texttt{copula} package includes three tests of the max-stability assumption; see Kojadinovic et al. (2011); Kojadinovic and Yan (2010); Ben Ghorbal et al. (2009), while the graphical diagnostic proposed by Gabda et al. (2012) is part of \texttt{mev}.

**Likelihood-based estimation.** The likelihood of a simple max-stable random vector \( Z \) with a parametric model for \( V \) is obtained by differentiating the distribution function \( \exp(-V(z)) \) with respect to each \( z_1, \ldots, z_D \). The number of terms in the likelihood is the \( D \)th Bell number, which is the total number of partitions of \( D \) elements into \( k \) (\( k = 1, \ldots, D \)) elements. Even in moderate dimensions, the number of distinct likelihood contributions is huge and the calculations become prohibitive. One way to circumvent this problem is to add the information about the partition if occurrence times are recorded (Stephenson and Tawn, 2005). The likelihood is biased unless \( n \gg D \) since the empirical partition also needs to converge to the limiting hitting scenario; for weakly dependent processes, use of the observed partition may induce bias (Wadsworth, 2015). Instead, Thibaud et al. (2016) propose to impute the partition using a Gibbs sampler, while Huser et al. (2019) use a stochastic expectation-maximisation algorithm; the \( E \)-step for the missing partition uses a Monte-Carlo estimator, where approximate draws are obtained from the Gibbs sampler of Dombry et al. (2013). None of these extensions have been implemented in publicly available software packages.

**Parametric models.** While max-stable models have been around for a while, there are few software implementations for estimating such models. The \texttt{evd} and \texttt{copula} packages provide functionalities that are restricted to the bivariate setting, while \texttt{ExtremalDep} (Beranger et al., 2023) includes composite likelihood estimation via it’s function \texttt{ExtDep} for a variety of models. The \texttt{SpatialExtremes} and \texttt{CompRandFld} packages have methods for fitting max-stable processes using pairwise composite likelihood for spatial models; see Section 4.

There are only handful of useful parametric models that generalize to dimension \( D > 2 \). The prime example is the logistic multivariate extreme value model, which is overly simplistic and lacks flexibility since the distribution is exchangeable. Many existing models are special cases of a Dirichlet family of distributions (Belzile and Nešlehová, 2017) and obtained through tilting (Coles and Tawn, 1991) to satisfy the moment constraint. These all have the drawback that the number of parameters is constant or grows linearly with the number of dimensions \( O(D) \) and this typically isn’t enough for characterizing complex data. Two models derived from elliptical distributions, the Hüsler–Reiss model (Hüsler and Reiss, 1989) and the extremal Student- \( t \) (Nikoloulopoulos et al., 2009), are more useful in large dimensions because their scale matrix can be used to parametrize the pairwise dependence individually with \( O(D^2) \) entries, and they can be more readily adapted to the functional setting, with extensions for skew-symmetric families (Beranger et al., 2017). The last parametric family, of which the most prominent example is the asymmetric logistic distribution, are max-mixtures (Stephenson and Tawn, 2005) that assign different weights to multiple simultaneous combinations of extremes. This allows for some degree of asymmetry and asymptotic independence, but such models are overparametrized with \( O(2^D) \) coefficients.

Joint estimation of all marginal and dependence parameters is complicated because of the potential high-dimensionality of the optimization problem, but also because of potential model misspecification that leads to unplausible parameter estimates. It is therefore common to use a two-stage
A modeler’s guide to extreme value software

Unconditional simulation algorithms. For a long time, exact unconditional simulation algorithms for max-stable processes were elusive outside of special cases (Schlather, 2002). Both mev and graphicalExtremes (Engelke et al., 2022) implement the algorithm of Dombry et al. (2016) for selected multivariate models (including for the latter extremal graphical models on trees) ensuring exact simulation, whereas evd uses dedicated algorithms for logistic and asymmetric logistic models in arbitrary dimensions (Stephenson, 2003). The copula (evCopula1a objects) (Yan, 2007) and SimCop packages (Tajvidi and Turlach, 2018) have functionalities for simulation of some bivariate extreme value distributions and the multivariate logistic model, or Gumbel copula, and the package ExtremalDep generates observations from a semiparametric dependence model in the bivariate setting by using its spectral measure (Marcon et al., 2017a) and from elliptical extreme-value models by using componentwise maxima of simulations of the underlying elliptical models. Packages mev and BMAmevt (Sabourin and Naveau, 2014) provide simulators for selected parametric angular density models.

3.2. Threshold models

Multivariate regular variation, which underlies the max-stable distribution of Equation (7) for the case where marginal distributions have been standardized such that \( \xi_j = \alpha > 0 \) for \( j = 1, \ldots, D \), can also be used for threshold exceedances by considering the associated Poisson point process of extremes with intensity measure \( \Lambda \) on a risk region \( \mathcal{R} \subset \mathbb{R}_+^D \setminus \{0_D\} \), i.e., the positive orthant excluding the origin (Resnick, 1987). Assuming the intensity measure is absolutely continuous, the intensity function \( \lambda(x) = \partial^D \Lambda(x)/\partial x_1 \cdots \partial x_D \) exists and we can define a density over \( \mathcal{R} \) by renormalizing \( \lambda(x) \) by the measure of the risk region, \( \Lambda(\mathcal{R}) \). The resulting likelihoods of the point process, multivariate generalized Pareto distributions and more general threshold models are much simpler than their max-stable counterpart, but there are typically two numerical bottlenecks associated to fitting these models. The first arises from the calculation of the measure of the risk region, which is often intractable and must thus be estimated using Monte Carlo methods. There are closed-form expressions for few risk regions, notably \( \mathcal{R} = \{ x \in \mathbb{R}_+^D : x_j > u \} \); if \( \xi = 1_D \), then \( \mathcal{R} = \{ x \in \mathbb{R}_+^D : \|x\| > u \} \) has risk measure \( \Lambda(\mathcal{R}) = Du^{-1} \) irrespective of the model for \( \Lambda \). The second bottleneck is due to censoring: not all components of a random vector may be extreme and the limiting model may be a poor approximation at finite levels for weakly dependent vectors (Ledford and Tawn, 1996). To reduce the bias arising from consideration of the asymptotic distribution, it is customary to left-censor observations falling below marginal thresholds. Most multivariate peaks over threshold models are based on the multivariate generalized Pareto (Rootzén and Tajvidi, 2006), defined over \( \mathcal{R} = \{ x \in \mathbb{R}_+^D : \max_{j=1}^D x_j > u \} \). Alternative constructions of multivariate generalized Pareto are described in Rootzén et al. (2018). Kiriliouk et al. (2019) provide expressions for the likelihood of many parametric models with strate-
gies for diagnostics; these are not currently implemented in software. The point process likelihood can also be used in place of the multivariate generalized Pareto: the evd package proposes it for the bivariate case (Smith et al., 1997), but the censored likelihood implemented therein actually uses the max-stable copula (Ledford and Tawn, 1996).

Most implementations are restricted to the bivariate setting or are reserved for spatial data. The graphicalExtremes package (Engelke and Hitz, 2020) is a notable exception: it implements the multivariate Hüsler–Reiss generalized Pareto distribution for graphical models. Exploiting the relation between the model and conditional extremal dependence, the parameters of the Hüsler–Reiss or Brown–Resnick process are directly related to the variogram matrix, whose entries are estimated empirically using pairwise empirical estimators of $\chi$. The full likelihood can be used (including censoring), but the factorization of the likelihood over cliques allows for higher-dimensional models to be fitted through maximum likelihood at reasonable cost, since each component is low dimensional. The bvt.cp1.ot function in the evd package provides threshold stability plots in the bivariate case based on the spectral measure. mev provides composition sampling algorithms for threshold models for various risk functionals $\mathcal{R}$ in the multivariate setting (Ho and Dombry, 2019).

Rather than condition on the maximum component exceeding a threshold, we can focus instead on exceedances of the $j$th component, i.e., consider a limiting model for $Y_{-j \mid Y_j > u}$. Heffernan and Tawn (2004) showed that a particular choice of normalizing sequences allows for the existence of non-degenerate limiting measure, including for asymptotically independent models. Inference for the conditional extremes model is usually performed in two stages. In the first, the marginal distributions are estimated semiparametrically and data are transformed to Laplace margins (Keef et al., 2013). In the second step, the dependence parameter vectors are estimated using a nonlinear regression model under the assumption of Gaussian residuals. Inference for the conditional extremes model as implemented in the texmex package relies on simulation: the probability of extreme events is obtained by calculating the fraction of simulated points falling in the risk region and uncertainty quantification is done using the bootstrap scheme described in Heffernan and Tawn (2004).

The multivariate regular variation representation provides another modelling approach for peaks over threshold using radial exceedances. For this, random vectors are first transformed so that their marginal distributions are standardized with $\xi = 1$, say $Y \sim Y^*$, and then mapped to radius and pseudo-angles $(R, \Omega)$, with, e.g., $R = \|Y^*\|$ and $\Omega = Y^*/R$. Since $R$ and $\Omega$ become stochastically independent as $R$ tends to infinity, one can focus on modelling the spectral measure $H(\omega)$ appearing in eq. (7). ExtremalDep, through $\mathcal{f}_{\text{ExtDep}}$, supports composite likelihood maximum estimation with pseudo-angles for $D$-dimensional distributions, with composite likelihood information criteria to compare models, density functions, plots, etc., for multiple parametric models. Nonparametric estimation of the spectral measure only requires the user to impose mean contraints. Starting from a sample of pseudo-angles, these can be enforced through empirical likelihood method (Einmahl and Segers, 2009) or Euclidean likelihood (de Carvalho et al., 2013). The extremis package (de Carvalho et al., 2020) implements these functionalities in the bivariate setting, and mev in higher dimensions. The unpublished EVcopula package implements the bivariate model of Wadsworth (2016) along with likelihood-based estimation methods and can be used to estimate probabilities of large bivariate quantities for both asymptotic (in)dependence scenarios. The BMAmevt package is dedicated to the implementation of a Bayesian model averaging based on semiparametric models for pseudo-angles in moderate dimensions (Sabourin et al., 2013). The Matlab package ECSADES performs penalised piecewise-constant marginal generalised Pareto and conditional extremes regression modelling (Ross et al., 2020).

### 3.3. Coefficients of tail dependence and structural variables

In multivariate settings, knowing the speed of decay of the dependence between pairs of random variable is useful for risk assessment. This also helps validate empirically if asymptotic multivariate extreme value models are warranted or not. The tail correlation coefficient is $\chi = \lim_{\nu \to 1} \chi_{\nu}$ (Coles et al., 1999), where

$$\chi_{\nu} = \frac{\Pr[\min_i\{F_i(Y_i) > \nu\}]}{1 - \nu}. \quad (8)$$

The latter is used to assess whether extremes are asymptotically independent ($\chi = 0$) or dependent ($\chi > 0$). Equation (8) suggests replacing the unknown distribution functions by their empirical coun-
terpart to estimate the coefficient. In the bivariate case, the estimator is often rather defined as
\(2 - \log[\Pr(F_1(Y_1) < v, F_2(Y_2) < v)]/\log(v)\) for \(v \approx 1\).

A related coefficient measuring dependence is the coefficient of tail dependence, often denoted \(\eta\),
which can be used to characterize the speed of decay for asymptotically independent variables. With random vectors transformed to unit Pareto margins, say \(Y_p\), the structural variable \(T = \min_{j=1}^D Y_j^p\) is such that, for large \(u\) (Ledford and Tawn, 1996, eq. 5.6),

\[
\Pr(T > u + t \mid T > u) \approx \frac{L(u + t)}{L(u)} (1 + t/u)^{-1/\eta},
\]

with \(L(x)\) a slowly varying function. The coefficient of tail dependence can be estimated by fitting
a generalized Pareto distribution with shape \(\eta\) and scale \(\eta u\) to exceedances of \(T\) above \(u\). If data
are transformed to the exponential scale instead, the scale parameter of the structural variable is \(\eta\)
and the maximum likelihood estimator of the latter coincides with Hill’s estimator (Section 2.5).

The coefficient of tail dependence takes values in \((0, D^{-1})\) if the variables are negatively associated,
\(\eta = D^{-1}\) for independent variables, and \(\eta \in (D^{-1}, 1)\) if the variables exhibit positive association.
In the bivariate setting, the coefficients \(\eta_C\) for subsets \(C < \{1, \ldots, D\}\) satisfy ordering constraints (de Haan and Zhou, 2011, § 4.2).

In the bivariate setting, it is customary to consider \(\check{\eta} = 2\eta - 1\) instead of \(\eta\), which gives \(\check{\eta} \in (-1, 1]\)
(Coles et al., 1999). The \texttt{evd} package function \texttt{emp\_chi}\_plot provides plots of \(\check{\eta}\) and \(\eta\) based on the empirical
distribution of the minimum, with approximate pointwise standard errors through the delta-
method. The \texttt{mev} package provides various estimators of \(\eta\) and \(\check{\eta}\), while \texttt{graphicalExtremes} includes
empirical estimators \texttt{emp\_ch}\_1 that can be used to obtain empirical estimates of the dependence ma-
trix of the Hüsler–Reiss distribution.

Extensions that consider different tail decays have emerged in the last decade, leading to angular
dependence function. For example, Beirlant et al. (2011) and Dutang et al. (2014) consider projections
of the form \(Z_{\omega} = \min(Y_1^\omega, Y_2^\omega \omega/(1 - \omega))\) for \(\omega \in (0, 1)\) a fixed angle. Under a regular variation
assumption, the distribution of \(Z_{\omega}\) can be approximated by the so-called extended Pareto distribution.
The parameters of the latter can be estimated using the minimum density power divergence
(MDPD) criterion (Dutang et al., 2014), which includes the maximum likelihood estimator as a spe-
cial case. The \texttt{RTDE} package (Dutang, 2020) provides various functions to estimate the parameters
of this model, and the returned objects allow users to summarize/plot fitted outputs, to compute the
bivariate tail probability as well as to perform a simulation analysis. A similar approach is consid-
ered in Wadsworth and Tawn (2013) and implemented in \texttt{lambdadep} function of the \texttt{mev} package;
the authors look at different extrapolation paths by replacing the multivariate regular variation by
a collection of univariate regular variation assumptions. Mhalla et al. (2019) also use such ideas to
implement generalized additive regression for extremal dependence parameters. The drawback of
these approaches, termed structural variables since they use univariate projections, is that estimation
is carried independently for every angle \(\omega\), but alternative estimators based on limit sets (Nolde
and Wadsworth, 2022) are being proposed at the time of writing.

### 3.4 Time series and graphical models

Data on a single variable collected over time often exhibit short-term temporal dependence, which
can lead to extremes occurring in clusters. As a minimum, statistical methods for time series ex-
tremes need to account for dependence in the data and to estimate the extent to which extremes
cluster, either directly or using a dependence model. For reviews of this area see Chavez-Demoulin
and Davison (2012) and Reich and Shaby (2016).

#### 3.4.1 Extremal index estimation

For stationary processes satisfying the \(D(u_n)\) condition, which limits long-range dependence at ex-
treme levels, the strength of local serial extremal dependence is commonly measured by the extremal
index. The latter can be interpreted as the reciprocal of the limiting mean cluster size in a Poisson
cluster process of exceedances of increasingly high thresholds. Table 6 gives basic information about
the direct estimators of the extremal index that feature in this section, while Table 7 summarises im-
plementations of these estimators, including information about diagnostics for the choice of tuning
parameters. When a threshold is involved these diagnostics can be used for threshold selection. The
| estimator | reference | tuning parameter(s) |
|-----------|-----------|---------------------|
| runs      | Smith and Weissman (1994) | run length, threshold |
| blocks (blocks 1) | Smith and Weissman (1994) | block size, threshold |
| modified blocks (blocks 2) | Smith and Weissman (1994) | block size, threshold |
| intervals (FS) | Ferro and Segers (2003) | threshold |
| iterative least squares (ILS) | Süveges (2007) | threshold |
| K-gaps | Süveges and Davison (2010) | run length K, threshold |
| semiparametric maxima (SPM) | Northrop (2015) | block size |

Table 6: Overview of some direct estimators of the extremal index with associated references and tuning parameters.

diagnostics in the evd, evir, exdex, fExtremes (Wuertz et al., 2017) and texmex packages are threshold stability plots for the extremal index. The information matrix test of Süveges and Davison (2010), which is based on a model for truncated inter-exceedance times called K-gaps, is provided by the exdex and mev packages. The packages evd (function clusters), extRemes (deCluster), fExtremes (deCluster), POT (clust) and texmex (deCluster) use an estimate of the extremal index to decluster exceedances of a threshold to form a series of sample cluster maxima.

3.4.2 Marginal modelling

Suppose that interest is limited to marginal extremes. The limiting distributions of cluster maxima and a randomly chosen threshold exceedance are identical, so inferences can be made using a marginal generalized Pareto model for sample cluster maxima or for all exceedances. The texmex (Southworth et al., 2020) package is the most complete implementation of the analysis of cluster maxima: it uses a semi-parametric bootstrap procedure to account for uncertainty in declustering and in marginal inference and can also accommodate covariate effects. The declustering approach is wasteful of data and Fawcett and Walshaw (2012) show that the difficulty of identifying clusters reliably can lead to substantial bias. When using all exceedances appropriate adjustment must be made for dependence in the data and for the value of the extremal index (Fawcett and Walshaw, 2012): the lite package (Northrop, 2022) uses the methodology of Chandler and Bate (2007) to estimate a marginal log likelihood that has been adjusted for clustering using a sandwich estimator of the covariance matrix of the marginal parameters and combines this with a log likelihood for the extremal index under the K-gaps model. The extremefit package provides a semiparametric procedure for time series extremes, as described in Section 2.5. Table 8 gives summaries of these packages and the packages that enable the estimation of time series dependence.

3.4.3 Models for dependence

In some applications it is important to infer more about the behavior of an extreme event than the size of a cluster of extreme values. For example, the duration of an extreme event or an accumulation of the extreme values may be of interest. This requires the nature of serial extremal dependence to be modeled. The extremogram (Frolova and Cribben, 2016) package implements the extremogram (Davis and Mikosch, 2009; Davis et al., 2011, 2012) to inform modelling by exploring quantitatively serial extremal dependence within stationary time series and between different time series. In the univariate case, it gives estimates of the conditional probabilities that a variable exceeds a user-supplied high threshold at time \( t + l \) given that it exceeded this threshold at time \( t \). The stationary bootstrap is used to provide confidence intervals.

The fitmcgpd function in the POT package performs maximum likelihood inference using a first-order Markov chain model, in which one of several bivariate extreme value distributions is used as a model for successive threshold exceedances (Smith et al., 1997). The function simmc simulates from this type of model, as does the simmc function in the evd package. The tsxtreme package models time series dependence using the conditional extremes approach of Heffernan and Tawn (2004), which enables a greater range of dependence structures to be modeled. Inferences are performed using two-step maximum likelihood fitting and a Bayesian approach in which inferences are made about a
Table 7: Comparison of R packages for the direct estimation of the extremal index. Estimator(s): name(s) of the estimators available; estimation: function name(s) for estimation; uncertainty quantification (UQ): are methods for estimating uncertainty provided?; diagnostics: function names(s) for choosing tuning parameters.

| package     | estimator(s) | estimation | UQ | diagnostics                |
|-------------|--------------|------------|----|----------------------------|
| evd         | runs, FS     | exi        | no | exiplot                    |
| evir        | blocks 2     | exindex    | no | exindex                    |
| extRemes    | runs, FS     | extremalIndex | yes | —                          |
| exdex       | ILS          | iwls       | no | —                          |
|             | K-gaps       | kgsps      | yes| choose_uk                  |
|             | SPM          | spm        | yes| choose_b                   |
|             | runs         | runTheta   | no | exiplotPlot                |
|             | blocks 1     | clusterTheta | no | exindexesPlot              |
|             | blocks 2     | blocktheta | no |                            |
|             | intervals    | ferrosgegersTheta | no |                            |
| mev         | ILS, FS      | ext.index  | no | ext.index                  |
|             | K-gaps       | ext.index  | no | informat.test, ext.index   |
| POT         | runs         | fitexi     | no | exiplot                    |
| revdbayes   | K-gaps       | kgsps_post | yes| —                          |
| texmex      | FS           | extremalIndex | yes | extremalIndexRangeFit      |
| tsxtreme    | runs         | thetaruns  | yes| —                          |

more flexible model in which all inferences are performed simultaneously (Lugrin et al., 2016). The functions theta2fit (MLE) and thetafit (Bayesian) provide inferences for the sub-asymptotic extremal index of Ledford and Tawn (2003).

The ev.trawl package implements the modelling approach described in Noven et al. (2018), which is based on the representation of a generalized Pareto distribution as a mixture of exponential distributions in which the exponential rate has a gamma distribution. An exponential trawl process introduces time series dependence in a latent gamma process, while a marginal probability integral transform allows both negative and positive shape parameter values. The CTRE package deals with processes for which inter-exceedance times have a heavy-tailed distribution and therefore a Poisson cluster representation is not appropriate (Hees et al., 2021). Parameter stability plots are provided to guide the selection of a suitable threshold.

3.4.4 Graphical extremes

Under the first-order Markov chain model for time series extremes of Smith et al. (1997), the value of a variable at time \( t+1 \) is assumed to be conditionally independent of its value prior to time \( t \) given the value at time \( t \). This simple dependence structure could be represented as a graphical model in which nodes representing the value of the variable are only connected by an edge if they correspond to adjacent time points.

The packages graphicalExtremes (Engelke and Hitz, 2020) and gremes (Asenova et al., 2021) provide more general graphical modelling frameworks for extremes, based on a multivariate Hüsler–Reiss generalized Pareto model for peaks over thresholds; see also Section 3.2. A graph represents conditional independences between variables. If the graph is sparse then the joint distribution decomposes into the product of lower-dimensional distributions, which results in a more parsimonious and tractable model. If the graph is a tree, that is, there is exactly one path along edges between any pair of nodes, then this decomposition is particularly simple. The graphicalExtremes and gremes packages provide functions to fit a multivariate Hüsler–Reiss generalized Pareto model given a user-supplied graph and functions to simulate from this model. The specifics of the theory underlying
these packages differ but the resulting model structures coincide when based on a tree.

In some applications, such as the analysis of extreme river flows, there is a physical network from which the graph can be constructed. In other cases the graph is conceptual: graphicalExtremes also provides a means to infer the structure of a graph from data.

4. Functional extremes (including spatial extremes)

Functional extremes designates a relatively recent branch of extreme value analysis concerned with stochastic processes over infinite-dimensional spaces, especially spatial and spatio-temporal extremes in geographic space (Davison et al., 2012; Huser and Wadsworth, 2022). We here use the term space for $\mathbb{R}^d$ with $d \geq 1$, including the combination of geographic space and time ($d = 3$), and we explicitly refer to time only where necessary. In practice, we usually work with finite discretizations of the study domain, such that many multivariate results and techniques carry over to the functional setting, although usually in relatively high dimension.

Common exploratory tools for extremal dependence are coefficients for bivariate distributions assessed as a function of spatial distance or temporal lag (e.g., extremal coefficient function based on bivariate extremal coefficients $\theta_2$, tail correlation function based on the $\chi$ measure, $F$-madogram, concurrence probability for maxima).

The asymptotic mechanisms for functional maxima and threshold exceedances are similar to the multivariate setting. Available statistical implementations are summarized in Section 4.1. Marginal and dependence modelling is discussed in Section 4.3. Aspects that we consider as still underdeveloped in existing implementations are listed in Section 4.4.

We use $Y(s)$ for stochastic processes indexed by $s \in \mathcal{S} \subset \mathbb{R}^d$, representing the process of the original event data. Usually we have a random sample of observations $Y_i(s_j)$ for $j = 1, \ldots, D$ locations observed at $i = 1, \ldots, n$ time points and denote a realization in space by $Y = \{Y(s_1), \ldots, Y(s_D)\}$, by analogy with the multivariate case.

Max-stable processes are the natural class of models for locationwise maxima taken over temporal blocks of the same length, such as annual maxima observed at fixed spatial locations. A max-stable process possesses finite-dimensional max-stable distributions, and convergence to a max-stable process can be defined through the convergence of all finite-dimensional distributions, such that strong links arise with the univariate and multivariate setting. If there exist sequences of normalizing functions $a_n(s) > 0$ and $b_n(s)$ such that the law of the scaled maximum converges for all finite-dimensional distributions,

$$\lim_{n \to \infty} \Pr\{a_n(s)[M_n(s) - b_n(s)] \leq x\} = \Pr\{Z(s) \leq x\}, \quad s \in \mathcal{S},$$

with $Z(s)$ a nondegenerate limit process, then $Z(s)$ is max-stable.

The most widely used setting for functional peaks over threshold follows the multivariate setting by assuming that data have been standardized to $Y^*(s)$, i.e., marginally transformed with a transformation $g$ that is strictly monotonic (i.e., $g(x_2) > g(x_1)$ if $x_2 > x_1$), and that ensures positivity (i.e.,...
$g(x) \geq 0$ with standardized tails of transformed random variables for which $\lim_{x \to \infty} x \Pr[g(Y(s)) > x] = 1$. For example, we can choose $Y^* (s) = g^*_x[Y(s)] = 1/(1 - F^*_x[Y(s)])$, where $F^*_x$ is the marginal distribution of $Y(s)$. Risk-Pareto processes (Ferreira and de Haan, 2014; Thibaud and Opitz, 2015; Domby and Ribatet, 2015; de Fondeville and Davison, 2018; Engkel et al., 2019) arise asymptotically when a functional $r$ of the standardized process $Y^*(s)$ exceeds a threshold that tends towards the upper endpoint of the probability distribution of $r$.

Typically, summary functionals are homogeneous, meaning $r(tx) = tr(x)$ for $t > 0$; examples include the average $r(x) = |\mathcal{S}|^{-1} \int_{\mathcal{S}} x(s) ds$, the minimum $r(x) = \min_{s \in \mathcal{S}} x(s)$, the maximum $r(x) = \max_{s \in \mathcal{S}} x(s)$, or the median. Convergence is assumed in the space of continuous functions over compact $\mathcal{S}$, such that the distribution of the functional $r[g_x(Y(s))]$ is well defined. Functional convergence of maxima in (10) implies functional convergence to $r$-Pareto processes $Z_r(s)$:

$$
\lim_{u \to \infty} \Pr[u^{-1} Y^*(s) \leq x | r(Y^*(s)) \geq u] = \Pr[Z_r(s) \leq x], \quad s \in \mathcal{S}.
$$

(11)

Max-stable and generalized Pareto processes have different probabilistic structures, but there always is a one-to-one correspondence between their dependence structures. Estimation of the marginal distributions and of the dependence structure is often conducted in two separate steps. The space-time dependence between sites is normally captured by correlation functions or variograms, which leads to much fewer parameters to infer than in the unstructured multivariate setting.

These asymptotic models can accommodate either asymptotic dependence or full independence among the variables $Y(s_1)$ and $Y(s_2)$ at locations $s_1, s_2 \in \mathcal{S}$. However, many stochastic processes, for example non-degenerate Gaussian processes, exhibit dependence at finite levels even if they are asymptotically independent in the limit, so the above characterization is too restrictive for accurate modelling. The coefficient of tail dependence introduced in (9), if considered for $D$ sites $s_1, \ldots, s_D$, is therefore restricted to values $\eta \in \{1/D, 1\}$. More flexible dependence structures can be achieved within the conditional extremes framework with conditioning on a fixed location (Wadsworth and Tawn, 2022; Simpson et al., 2023). Finally, so-called subasymptotic models do not arise as classical extreme value limits but focus on flexibly capturing dependence remaining at subasymptotic levels, for instance with asymptotic independence where $1/D < \eta(s_1, s_2) < 1$ is possible; for example, the class of max-ininitely divisible processes (Huser et al., 2021), which is useful for flexible modelling of location-wise maxima. Most such proposals do not come with packaged and generic software implementations so far.

### 4.1. Max-stable processes for maxima data

Suppose that data consist of locationwise block maxima $M_i(s_j)$, where $i = 1, \ldots, m$ indexes the blocks, e.g., the observation year in case of annual block maxima. The SpatialExtremes package provides the most comprehensive collection of functions for exploration and statistical inference with max-stable processes for spatial maxima data in geographical space ($d = 2$). While standard full likelihoods are not tractable even for moderately many locations with the common models, pairwise likelihood has become the standard approach for fitting max-stable processes, with implementations in SpatialExtremes and CompRandFld. ExtremalDep offers estimation routines using the Stephenson–Tawn likelihood and composite likelihoods for (skewed) extremal-$t$ processes, including the Schlather model. The unpublished package BRdac accompanying Hector and Reich (2023) offers pairwise composite likelihood estimation via distributed learning using using a divide-and-conquer procedure for Brown–Resnick max-stable processes, offering a scalable estimation strategy through local likelihoods.

Global dependence measures such as concurrence maps (Dombry et al., 2018), available from ConcurrenceMap in SpatialExtremes, can be constructed from bivariate summaries. The intractability of the multivariate max-stable distribution function $G$, described in eq. (7), has led to pairwise likelihood becoming the standard estimation method for spatial max-stable processes. In SpatialExtremes, joint frequentist estimation of marginal and dependence parameters is possible, where auxiliary variables can be flexibly included in the three parameters of the marginal generalized extreme value distribution. Similar to generalized additive models, smoothness penalties can be imposed on nonlinear effects modeled through spline functions. In contrast to the aforementioned generalized additive model approach without dependence, the numerical optimization becomes more involved here, such that only a moderate number of marginal parameters can be reasonably estimated.
RandomFields (Schlather et al., 2015) provides a large variety of max-stable models and particularly of tail correlation functions, with a focus on implementing simulation from such models. Moreover, the package encapsulates vast functionality, especially simulation, for Gaussian random fields, which are often the building blocks for the more sophisticated extreme value models. The package provides multiple state-of-the-art algorithms for simulating Brown–Resnick max-stable processes. Exact unconditional simulation of max-stable processes is available in RandomFields, mev and SpatialExtremes, but only the latter offers conditional simulation of max-stable random fields (conditional on observed values at given locations) using Gibbs sampling (Dombry et al., 2013). CompRandFld’s simulation routine for max-stable processes uses an interface to RandomFields. For a particular hierarchically structured max-stable dependence model, known as the Reich–Shaby model (Reich and Shaby, 2012) that is constructed using spatial kernel functions and is derived from the spectral representation of a max-stable process based on a $l_p$-norm (Oesting, 2018), estimation tools are available in the hkevp package. It is difficult to fit because of the dual role of its nugget parameter $\alpha > 0$. The hkevp package provides a Metropolis-within-Gibbs algorithm for Bayesian estimation of the model and for simulation.

4.2. Peaks-over-threshold modelling

For functional peaks over threshold, mvPot provides parametric simulation and estimation tools for various $r$-Pareto processes using Brown–Resnick and extremal Student- $t$ dependence structures (de Fondeville and Davison, 2018, 2022). Parameter estimates are computed using optimization of either full likelihood or gradient score functions; the latter remains computationally tractable for settings where full likelihood does not. Estimation of the marginal transformation $T$ is not implemented and has to be performed prior to estimating the extremal dependence parameters using mvPot. A competitive estimation procedure is the gradient score estimating equation of de Fondeville and Davison (2018), which does not require calculation of the normalizing constant of the model and also replaces censoring with downweighting. While statistically less efficient than full likelihood estimation, the procedure is more robust and can be applied in very high-dimensional settings. For estimation, numerical implementations are currently restricted to the Brown–Resnick model. The package mvPot also offers tools for simulation and calculation of likelihoods for the extremal-$t$ dependence model. The mev package also proposes likelihood functions and unconditional simulation routines for generalized $r$-Pareto processes (de Fondeville and Davison, 2022).

Some other implementations allowing estimation of asymptotic dependence structures use original event data $Y_i(s_j)$ and can be viewed as working on the interface of max-stable and peaks over threshold models. For example, moment-based estimation of parametric models, based on contrasting empirical and parametric versions of a variant of the so-called tail dependence function, is implemented in the package tailDepfun (Einmahl et al., 2018).

4.3. Modeling spatially varying marginal distributions

In practice, marginal distributions $F_s$ in functional data are usually not stationary, such that variation of marginal extreme value parameters with respect to space and time, or with respect to other available auxiliary variables, has to be captured. In the locationwise maxima setting, we can use use the generalized extreme value distribution and consider its parameters as functions of space, i.e., $\xi(s), \mu(s), \sigma(s)$. Different options exist in the peaks over threshold setting. A common approach is to fix a high, potentially nonstationary threshold $u(s)$, and then estimate the threshold exceedance probability $p(s) = \Pr(Y(s) > u(s)) = 1 - F_s(u(s))$ and the generalized Pareto parameters $\xi(s), \sigma(s)$ based on observations of the exceedances $Y(s) - u(s) > 0$.

The regression framework discussed in Section 2.3 are relevant for modelling marginal extreme value parameters that vary with location in a first modelling step. Generalized additive modelling allows capturing complex nonlinear patterns of spatial nonstationarity using relatively large numbers of parameters. Some care may be required in tuning smoothing hyperparameters since in this step one usually assumes independence of observations $Y_i(s_j)$, so functional dependence across space or time is disregarded. Specifically, MCMC-based Bayesian estimation of marginal parameters (using Gaussian process priors) for generalized extreme value distributions for maxima is possible through SpatialExtremes, and hkevp (Sebille, 2016) includes a similar function. The SpatialGEV package (Chen et al., 2021) provides a template for fitting latent spatial models with marginal generalized extreme value distributions and Gaussian process priors on the parameters using quadratic approxima-
tions to the marginal posterior. The unpublished package SpatGEVBMA fits a latent model with generalized extreme value margins whose parameters follow Gaussian process priors with explanatory variables. Its defining functionalities are the use of Laplace approximations for automating proposals, and Bayesian model averaging of regression models to account for variable selection uncertainty (Dyrrdal et al., 2014).

An important alternative to Monte Carlo methods is to estimate complex integrals arising from Equation (5) through the integrated nested Laplace approximation (INLA). The INLA package proposes computationally convenient representations of the spatial Matérn covariance function through the stochastic partial differential equation approach of Lindgren et al. (2011) for spatial and spatio-temporal latent Gaussian modelling. As mentioned in Section 2.4, INLA provides implementation for generalized extreme value distributions (with covariates and random effects in the location parameter) and the generalized Pareto distribution (with covariates and random effects in a quantile at a probability level \( \alpha \in (0, 1) \) specified by the user; see Opitz et al. (2018) and Krainski et al. (2018, Chapter 6).

The package further allows joint estimation of several regression designs where some of the random effects can be in common (i.e., shared through a scaling coefficient) among these, which is beneficial to obtain cross-correlation in the posteriors of the predictors of several response types. For example, we could combine a logistic regression for the exceedance probability with a generalized Pareto regression for the excess above the threshold, and a shared random effect with a positive sharing coefficient would entail positive posterior correlation between the exceedance probability and the size of the excess.

4.4. Outlook for functional extremes

The coverage of max-stable processes, which remains an area of very active research, is much more comprehensive than others, with the notable exception of composite or full log likelihood inference for max-stable processes. Formulae exist for many partial derivatives of the exponent function \( V \) arising in the multivariate max-stable cumulative distribution functions and, in principle, the Stephenson–Tawn likelihood (or a bias-corrected version thereof) could be programmed for full likelihood inference beyond the bivariate case. Most of the models are also implemented with spatial applications in mind, even if temporal or spatio-temporal applications are possible. Max-infinitely divisible models are not covered in software yet, and Bayesian models with latent processes are often not provided with numerical implementation because of the complexity of implementation and also sometimes very long execution times of codes.

There are much fewer implementations for threshold models. Whereas their construction can be viewed as more flexible and intuitive than the one of the corresponding max-stable processes, they are conditional models with respect to threshold exceedance of the summary functional \( r \). In the finite-sample setting of statistical practice, this means that observations at some locations may not correspond to marginal exceedances and may therefore not be coherent with the asymptotic model. A common remedy is censoring, but this makes estimation more costly because the likelihood functions now include high-dimensional distribution functions which typically must be calculated via Monte-Carlo methods for each vector of observation. Generic full likelihood estimation procedures have been proposed, and are available (though computationally costly) for some models. However, available implementations do not yet come with a comprehensive set of models and methods for parameter inference, model validation and comparison. An obvious solution to facilitate such implementation, provided that parameters are identifiable from lower-dimensional summaries, would be to use composite likelihood. Likewise, Bayesian generalized Pareto models with latent Gaussian process priors could be easily coded in many probabilistic programming languages outside of \( R \), such as Stan, but no general-purpose routines exist so far.

Simulation algorithms for unconditional simulation from generalized \( r \)-Pareto processes with arbitrary risk functionals \( r \) are still elusive, as designing efficient accept-reject methods requires case-by-case analysis. Available conditional simulation code typically amounts to simulation of elliptical distributions (log-Gaussian or Student-\( t \)) with linear constraints.

Implementations with documented code are often available as supplementary material to methodological papers but have not been encapsulated in officially validated packages; see Huser and Wadsworth (2019); Bacro et al. (2020); Simpson et al. (2023) for recent examples. Huser and Wadsworth (2019) has companion code for frequentist estimation of a flexible subasymptotic spatial model in the unpub-
| methods | package | functions | scope |
|---------|---------|-----------|-------|
| Tajvidi and Turlach (2018) | copula | rCopula* | b, m |
| Stephenson (2003) | SimCop | — | b |
| Dombry et al. (2016) | evd | rbvevd, rmvevd | b, m |
| Engelke and Hitz (2020) | graphicalExtremes | rmstable | m, f |
| Beranger et al. (2017) | ExtremalDep | rExtDep, rExtDepSp | m, f |
| Padoan and Bevilacqua (2015) | CompRandFld | RFSim* | f |
| Dombry et al. (2015) | SpatialExtremes | condrmaxstab | f |
| Dombry et al. (2016) | SpatialExtremes | rmmaxstab | f |
| Schlather et al. (2015) | RandomFields | RFsimulate* | f |
| Reich and Shaby (2012) | hkevp | hkevp.rand | f |
| Ballani and Schlather (2011) | BMAmvt | rnestlog, rpairbeta | a |
| Ho and Dombry (2019) | mev | rparpcs | p |
| de Fondeville and Davison (2018) | mev | rparp | p |
| de Fondeville and Davison (2018) | mvPot | — | p |
| de Fondeville and Davison (2022) | mev | rgparp | p |

Table 9: Overview of simulation algorithms for bivariate (b) and multivariate (m) max-stable distributions and for max-stable processes (f), and for angular (a) and Pareto processes (p) with associated references. Some of the listed functions (*) are generic and include specific classes for max-stable models, but other models as well.

5. Specialized topics

While our review has ranged mostly over software providing implementation of relatively generic methods that can be useful in various application contexts, there also has been active development of software libraries targeting specific application fields, and we here cite some of them.

**Hydrology and climate**: Regional frequency analysis using L-moments is possible with the `lmomrfa` (Hosking, 2023) package. The `climextRemes` (Paciorek et al., 2018) package leverages `extRemes` for climate extremes and implements methods highly relevant for this field, such as local likelihood fitting; the package is also available in Python. `IDF` provides intensity-duration-frequency (IDF) curves (Ulrich et al., 2020). `jointPm` implements the method of Zheng et al. (2015) for evaluating bivariate probabilities of exceedance. An example of a highly specialized package is `futureheatwaves` (Anderson et al., 2016) and facilitates finding, characterizing and exploring heatwaves in climate projections, while the Python package `teca` is dedicated to tracking extremes of large scale climate models. `Renext` includes methods for peaks over threshold with a variety of distributions and the possibility to include historical maximum records, along with tests of exponentiality and goodness-of-fit.

**Financial and actuarial science**: Some packages provide implementation of various generic models and methods for extreme values, but make strong use the semantics of those fields in their documentation. The packages `QRM`, its successor `qrmtools` (Hofert et al., 2022) and `ReIns` implement various functions to accompany the books McNeil et al. (2015) and Albrecher et al. (2017), respectively.

The package `fExtremes` provides functions for financial analysis used by the `Rmetrics` project.
| reference | package | functions | dim | data |
|-----------|---------|-----------|-----|------|
| Coles and Tawn (1991) | evd | fbvevd | b | max |
| Coles and Tawn (1991) | copula | fitCopula | b | max |
| Einmahl et al. (2018) | tailDepFun | Estimation... | m, f | max |
| Pickands (1981) and Caperaa et al. (1997) | evd | abvnonpar | b | ang |
| Gadendorf and Segers (2012) | copula | An | b, m | ang |
| Einmahl and Segers (2009) and de Carvalho et al. (2013) | extremis | angcdf | b | ang |
| Marcon et al. (2017b) | ExtremeDep | madogram, beed | m | ang |
| Beranger et al. (2021a) | ExtremeDep | fExtDep.np | b | ang |
| Wadsworth (2016) | EVCopula | fit.EV.copula | b | ang |
| Sabourin and Naveau (2014) | BMAMEVT | posteriorMCMC | m | ang |
| Smith et al. (1997) | evd | fbvpo | b | pot |
| Engelke et al. (2019) | graphicalExtremes | fmpareto_graph_HR | m | pot |
| Heffernan and Tawn (2004) | texmex | m | m | pot |
| Davison et al. (2012) | SpatialExtremes | fitcopula, fitmaxstab | f | max |
| Padoan and Bevilacqua (2015) | CompRandFld | FitComposite | f | max |
| Reich and Shaby (2012) | hkevp | hkevp.fit | f | max |
| Reich and Shaby (2012) | extRemes | abba | f | max |
| Beranger et al. (2021b) | ExtremeDep | fExtDep, fExtDepSpat | m, s | max |

Table 10: Overview of multivariate and functional estimation procedures for extremes according to dimension, either bivariate (b), multivariate (m) or functional (f) and data type/paradigm, one of block maximum (max), pseudo-angles (ang) or threshold exceedances (pot). Packages which only include likelihood but no optimization wrapper are excluded.
The package **VaRES** (Nadarajah et al., 2023) provides two popular risk measures (value at risk and expected shortfall) for a large collection of probability distributions, including many heavy-tailed distributions. The **extremis** package proposes functionalities to cluster multivariate financial time series based on their frequency and magnitude of extreme events.

**Machine learning:** The interface between statistical machine learning and extreme values has been growing in recent years, with proposals encompassing the use of gradient boosting for extremes (Velthoen et al., 2021, `gbex`) and of extremal random forests (Gnecco et al., 2022, `erf`) for modelling high quantiles of a univariate distribution. Another area of active research is open-set classification, dealing with classification of observations in categories not observed in training data: the Python package **EVM** implements the extreme value machine of Rudd et al. (2018), whereas R package **evtdist** includes the algorithms described in Vignotto and Engelke (2020).

**Survival analysis:** Presence of censoring or truncation mechanisms, common in survival analysis, require dedicated software implementations because they affect the likelihood contribution of observations. The Matlab **LATools** (Rootzén and Zholud, 2017) proposes an interface for interval-truncated generalized Pareto observations, while the **longevity** package (Belzile et al., 2022) handles more general partial observation schemes.

### 6. Discussion and conclusion

We have covered in this review a wide range of available software implementations for extremes, and we sincerely hope without omissions that are considered as important by authors or users. The development of extreme value software is key to extreme value analysis in practice and has become an active area of research, but the availability of implementations tends to lag behind methodological innovations since these are often not accompanied by generic, easily reusable and validated codes. To encourage modelers in applied sciences and in operational services to make use of the most advanced methods and models, off-the-shelf implementations are desirable. However, generic code may be difficult to provide due to the high sophistication of approaches as, for instance, with functional extremes. Designing generic estimation procedures that are flexible enough to be useful while at the same time being robust requires particular care. Writing this review made us aware of how challenging it is for the extreme value community to develop tested and easily reusable software that keeps pace with methodological progress: most software was written more than a decade ago, there are only handful of active maintainers, and most models proposed in the literature are not put together with software.

Many methods proposed recently are still not available and this is a major impediment for their adoption. The most obvious gap is in software for fitting multivariate max-stable models (with composite likelihood) and multivariate generalized Pareto distributions with censoring in moderate dimensions for the parametric models with suitable tools. The conditional spatial extremes model, which extends the Heffernan–Tawn approach to the spatial setting, has been used in many recent papers but no software has been released.

More refined tools are also required for the nonstationary exploration and inference of extreme values. In many application fields, physical change processes (e.g., climate change, land-use change) require tools to explore, model and infer nonstationary behavior in extremes, for instance for climate-change detection and attribution. Currently, nonstationary modelling is implemented for marginal distributions through regression designs, but implementations providing dedicated methods for extreme value detection and attribution under climate change are scarce, and easily reusable codes for nonstationary extreme value dependence structures are yet missing.

### Acknowledgements

Funding in partial support of this work was provided by the Natural Sciences and Engineering Research Council (RGPIN-2022-05001, DGERC-2022-00461). The authors declare that they have no conflict of interest. They thank three anonymous referees, the associate editor and B. Béralanger for helpful comments.

### References

Albrecher H, Beirlant J, Teugels J (2017) Reinsurance: Actuarial and Statistical Aspects. Wiley
Alfons A, Templ M (2013) Estimation of social exclusion indicators from complex surveys: The R package laeken. Journal of Statistical Software 54(15):1–25, doi:10.18637/jss.v054.i15

Anderson B, Eason C, Barnes E (2016) futureheatwaves: Find, Characterize, and Explore Extreme Events in Climate Projections. R package version 1.0.3

Asenova S, Mazo G, Segers J (2021) Inference on extremal dependence in the domain of attraction of a structured Hüsler–Reiss distribution motivated by a Markov tree with latent variables. Extremes 23(4):461–500, doi:10.1007/s10687-021-00407-5

Asquith W (2021) lmomco — L-moments, censored L-moments, trimmed L-moments, L-comoments, and many distributions. R package version 2.4.7

Bacro JN, Gaetan C, Opitz T, Toulemonde G (2020) Hierarchical space-time modeling of asymptotically independent exceedances with an application to precipitation data. Journal of the American Statistical Association 115(530):555–569, doi:10.1080/01621459.2019.1617152

Bader B, Yan J (2020) eva: Extreme value analysis with goodness-of-fit testing. R package version 0.2.6

Bader B, Yan J, Zhang X (2018) Automated threshold selection for extreme value analysis via ordered goodness-of-fit tests with adjustment for false discovery rate. The Annals of Applied Statistics 12(1):310–329, doi:10.1214/17-AOAS1092

Bellini F, Di Bernardino E (2017) Risk management with expectiles. The European Journal of Finance 23(6):487–506, doi:10.1080/1351847X.2015.1052150

Belzile LR, Nešlehová JG (2017) Extremal attractors of Liouville copulas. Journal of Multivariate Analysis 160:68–92, doi:10.1016/j.jmva.2017.05.008

Belzile LR, Davison AC, Gampe J, Rootzén H, Zholud D (2022) Is there a cap on longevity? A statistical review. Annual Review of Statistics and its Application 9:22–45, doi:10.1146/annurev-statistics-040120-025426

Belzile LR, et al. (2023) mev: Modelling extreme values. R package version 1.15

Ben Ghorbal N, Genest C, Nešlehová J (2009) On the Ghoudi, Khoudraji, and Rivest test for extreme-value dependence. Canadian Journal of Statistics 37(4):534–552, doi:10.1002/cjs.10034

Beranger B, Padoan SA, Sisson SA (2017) Models for extremal dependence derived from skew-symmetric families. Scandinavian Journal of Statistics 44(1):21–45, doi:10.1111/sjos.12240

Beranger B, Padoan SA, Sisson SA (2021a) Estimation and uncertainty quantification for extreme quantile regions. Extremes 24(2):349–375, doi:10.1007/s10687-019-00364-0

Beranger B, Stephenson AG, Sisson SA (2021b) High-dimensional inference using the extremal skew-t process. Extremes 24(3):653–685, doi:10.1007/s10687-020-00376-1

Boessenkool B (2017) extremeStat: Extreme value statistics and quantile estimation. R package version 1.5.5

Boessenkool B, Zhou C (2021) A horse race between the block maxima method and the peak–over-threshold approach. Statistical Science 36(3):360 – 378, doi:10.1214/20-STS795

Bücher A, Zhou C (2021) Extreme Value Modeling and Risk Analysis: Methods and Applications, CRC Press, Boca Raton, FL, pp 69–86

Cannon AJ (2010) A flexible nonlinear modelling framework for nonstationary generalized extreme value analysis in hydroclimateology. Hydrological Processes 24(6):673–685, doi:10.1002/hyp.7506
Belzile, Dutang, Northrop and Opitz

Caperaa P, Fougères A, Genest C (1997) A non-parametric estimation procedure for bivariate extreme value copulas. Biometrika 84:567–577, doi:10.1093/biomet/84.3.567

Carrer NL, Gaetan C (2022) Distributional regression models for extended generalized Pareto distributions. arXiv:220904660

de Carvalho M, Oumow B, Segers J, Warchol M (2013) A Euclidean likelihood estimator for bivariate tail dependence. Communications in Statistics — Theory and Methods 42(7):1176–1192, doi:10.1080/03610926.2012.709905

del Castillo J, Padilla M (2016) Modelling extreme values by the residual coefficient of variation. Statistics and Operations Research Transactions 40(2):303–320, doi:10.2436/20.8080.02.45

del Castillo J, Serra I, Padilla M, Morin D (2019) Fitting tails by the empirical residual coefficient of variation: The ercv package. The R Journal 11:56–68, doi:10.32614/RJ-2019-044

Castro-Camilo D, Huser R, Rue H (2022) Practical strategies for generalized extreme value-based regression models for extremes. Environmetrics 33(6):e2742, doi:10.1002/env.2742

Chandler RE, Bate S (2007) Inference for clustered data using the independence loglikelihood. Biometrika 94(1):167–183, doi:10.1093/biomet/asm015

Chavez-Demoulin V, Davison AC (2005) Generalized additive modelling of sample extremes. Journal of the Royal Statistical Society: Series C (Applied Statistics) 54(1):207–222, doi:10.1111/j.1467-9876.2005.00479.x

Chavez-Demoulin V, Davison AC (2012) Modelling time series extremes. REVSTAT-Statistical Journal 10(1):109–133, doi:10.57805/revstat.v10i1.113

Chen M, Ramezan R, Lysy M (2021) Fast approximate inference for spatial extreme value models. arXiv:211007061

Cheng L, AghaKouchak A, Gilleland E, Katz RW (2014) Non-stationary extreme value analysis in a changing climate. Climatic Change 127(2):353–369, doi:10.1007/s10584-014-1254-5

Coles SG (2001) An Introduction to Statistical Modeling of Extreme Values. Springer–Verlag, London

Coles SG, Tawn JA (1991) Modelling extreme multivariate events. Journal of the Royal Statistical Society Series B (Methodological) 53(2):377–392, doi:10.1111/j.2517-6161.1991.tb01830.x

Coles SG, Tawn JA (1996) A Bayesian analysis of extreme rainfall data. Journal of the Royal Statistical Society Series C (Applied Statistics) 45(4):463–478, doi:10.2307/2986068

Coles SG, Heffernan JE, Tawn JA (1999) Dependence measures for extreme value analyses. Extremes 2(4):339–365, doi:10.1023/A:1009963131610

Danielsson J, de Haan L, Peng L, de Vries C (2001) Using a bootstrap method to choose the sample fraction in tail index estimation. Journal of Multivariate Analysis 76(2):226–248, doi:10.1006/jmva.2000.1903

Daouia A, Gijbels I, Stupfler G (2022) Extremile regression. Journal of the American Statistical Association doi:10.1080/01621459.2021.1875837

Davis RA, Mikosch T (2009) The extremogram: A correlogram for extreme events. Bernoulli 15(4):977–1009, doi:10.3150/09-BEJ213

Davis RA, Mikosch T, Cribben I (2011) Estimating extremal dependence in univariate and multivariate time series via the extremogram. arXiv:11075592

Davis RA, Mikosch T, Cribben I (2012) Towards estimating extremal serial dependence via the bootstrapped extremogram. Journal of Econometrics 170(1):142–152, doi:10.1016/j.jeconom.2012.04.003

Davison AC, Padoan SA, Ribatet M (2012) Statistical modeling of spatial extremes. Statistical Science 27(2):161–186, doi:10.1214/11-STS376

Davison AC, Padoan SA, Stupfler G (2022) Tail risk inference via expectiles in heavy-tailed time series. Journal of Business & Economic Statistics (in press):1–14, doi:10.1080/07350015.2022.2078332

de Carvalho M, Rubio R, Palacios V (2020) extRemes: Statistics of extremes. R package version 1.2.1

de Fondeville R, Belzile L (2021) mvPot: Multivariate peaks-over-threshold modelling for spatial extreme events. R package version 0.1.5

de Haan L, Zhou C (2011) Extreme residual dependence for random vectors and processes. Advances in Applied Probability 43(1):217–242, doi:10.1239/aap/1300198520

Dekkers ALM, Einmahl JHJ, de Haan L (1989) A moment estimator for the index of an extreme-value distribution. The Annals of Statistics 17(4):1833–1855, doi:10.1214/aos/1176347397

Deville Y, IRSN (2022) Renext: Renewal method for extreme values extrapolation. R package version 3.1-3

---

post-print, version of July 25, 2023 Author' s final, peer-reviewed manuscript
do Nascimento FF, Moura e Silva WV (2016) **MCMC4Extremes**: Posterior distribution of extreme value models in R. R package version 1.1

Domby C, Ribatet M (2015) Functional regular variations, Pareto processes and peaks over threshold. Statistics and its Interface 8(1):9–17, doi:10.4310/SII.2015.v8.n1.a2

Domby C, Éyi Minko F, Ribatet M (2013) Conditional simulation of max-stable processes. Biometrika 100(1):111–124, doi:10.1093/biomet/ass067

Domby C, Engellke S, Oesting M (2016) Exact simulation of max-stable processes. Biometrika 103(2):303–317, doi:10.1093/biomet/asw008

Domby C, Ribatet M, Stoey S (2018) Probabilities of concurrent extremes. Journal of the American Statistical Association 113(524):1565–1582, doi:10.1080/01621459.2017.1556318

Dupuis D (1999) Exceedances over high thresholds: A guide to threshold selection. Extremes 1(3):251–261, doi:10.1023/A:1009914915709

Durrieu G, Grama I, Pham QK, Tricot JM (2015) Nonparametric adaptive estimation of conditional probabilities of rare events and extreme quantiles. Extremes 18(3):437–478, doi:10.1007/s10687-015-0219-z

Durrieu G, Grama I, Jaunatre K, Pham QK, Tricot JM (2019) **extremefit**: A package for extreme quantiles. Journal of Statistical Software 87(12):1–20, doi:10.18637/jss.v087.i12

Dutang C (2018) RTDE: Robust tail dependence estimation. R package version 0.2-1

Dutang C (2023) CRAN Task View: Extreme Value Analysis. URL https://CRAN.R-project.org/view=ExtremeValue, version 2023-03-30

Dutang C, Goegebeur Y, Guillou A (2014) Robust and bias-corrected estimation of the coefficient of tail dependence. Insurance: Mathematics and Economics 57

Dyrrdal AV, Lenkoski A, Thorarinsdottir TL, Stordal F (2014) Bayesian hierarchical modeling of extreme hourly precipitation in Norway. Environmetrics 26(2):89–106, doi:10.1002/env.2301

Eastoe EF, Tawn JA (2009) Modelling non-stationary extremes with application to surface level ozone. Journal of the Royal Statistical Society: Series C (Applied Statistics) 58(1):25–45, doi:10.1111/j.1467-9876.2008.00638.x

Einmahl JHH, Segers J (2009) Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution. The Annals of Statistics 37(5B):2953–2989, doi:10.1214/08-AOS677

Einmahl JHH, Kiriliouk A, Segers J (2018) A continuous updating weighted least squares estimator of tail dependence in high dimensions. Extremes 21(2):205–233, doi:10.1007/s10687-017-0303-7

Engelke S, Hitz AS (2020) Graphical models for extremes (with discussion). Journal of the Royal Statistical Society: Series B (Statistical Methodology) 82(4):87–93, doi:10.1111/rssb.12355

Engelke S, de Fondeville R, Oesting M (2019) Extremal behaviour of aggregated data with an application to downscaling. Biometrika 106(1):127–144, doi:10.1093/biomet/asy052

Engelke S, Hitz AS, Gnecco N, Hentschel M (2022) **graphicalExtremes**: Statistical Methodology for Graphical Extreme Value Models. R package version 0.2.0

Falk M, Hüsler J, Reiss RD (2011) Laws of Small Numbers: Extremes and Rare Events, extended edn. Birkhäuser, doi:10.1007/978-3-0348-0009-9

Fawcett L, Walshaw D (2012) Estimating return levels from serially dependent extremes. Environmetrics 23(3):272–283, doi:10.1002/env.2133

Ferreira A, de Haan L (2014) The generalized Pareto process; with a view towards application and simulation. Bernoulli 20(4):1717–1737, doi:10.3150/13-BEJ538

Ferro CAT, Segers J (2003) Inference for clusters of extreme values. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 65(2):545–556, doi:10.1111/1467-9868.00401

Figueiredo F, Gomes MI, Henriques-Rodrigues L, Miranda MC (2012) A computational study of a quasi-PORT methodology for VaR based on second-order reduced-bias estimation. Journal of Statistical Computation and Simulation 82(4):587–602, doi:10.1080/00949655.2010.547196

de Fondeville R, Davison AC (2018) High-dimensional peaks-over-threshold inference. Biometrika 105(3):575–592, doi:10.1093/biomet/asy026

de Fondeville R, Davison AC (2022) Functional peaks-over-threshold analysis. Journal of the Royal Statistical Society Series B: Statistical Methodology 84(4):1392–1422, doi:10.1111/rssb.12498

Frolova N, Cribben I (2016) **extremogram**: Estimation of Extreme Value Dependence for Time Series Data. R package version 1.0.2
Gabda D, Towe R, Wadsworth J, Tawn J (2012) Discussion of “Statistical modeling of spatial extremes” by A. C. Davison, S. A. Padoan and M. Ribatet. Statistical Science 27(2):189–192, doi:10.1214/12-STS376B

Gamet P, Jalbert J (2022) A flexible extended generalized Pareto distribution for tail estimation. Environmetrics 33(6):e2744, doi:10.1002/env.2744

Gerstengarbe FW, Werner PC (1989) A method for the statistical definition of extreme-value regions and their application to meteorological time series. Zeitschrift fuer Meteorologie 39(4)

Gilleland E (2016) Computing software. In: Extreme Value Modeling and Risk Analysis, Chapman and Hall/CRC, pp 525–536

Gilleland E, Katz R (2016) extRemes 2.0: An extreme value analysis package in R. Journal of Statistical Software 72(8):1–39, doi:10.18637/jss.v072.i08

Gilleland E, Ribatet M, Stephenson AG (2013) A software review for extreme value analysis. Extremes 16(1):103–119, doi:10.1007/s10687-012-0155-0

Gnecco N, Merga Terefe E, Engelke S (2022) Extremal random forests. arXiv:220112865

Grimshaw SD (1993) Computing maximum likelihood estimates for the generalized Pareto distribution. Technometrics 35(2):185–191, doi:10.1080/00401706.1993.10485040

Gudendorf G, Segers J (2012) Nonparametric estimation of multivariate extreme-value copulas. Journal of Statistical Planning and Inference 142(12):3073–3085, doi:10.1016/j.jspi.2012.05.007

Hall P (1985) Adaptive estimates of parameters of regular variation. Journal of the Annals of Statistics 13(1):331–341, doi:10.1214/aos/1176346596

Hector EC, Reich BJ (2023) Distributed inference for spatial extremes modeling in high dimensions. Journal of the American Statistical Association (in press), doi:10.1080/01621459.2023.2186886

Hees K, Nayak S, Straka P (2021) Statistical inference for inter-arrival times of extreme events in bursty time series. Computational Statistics & Data Analysis 155:107096, doi:10.1016/j.csda.2020.107096

Hosking JRM (2019) L-moments. R package, version 2.9

Hosking JRM (2023) Regional Frequency Analysis Using L-Moments. R package version 3.5

Husler R, Wadsworth JL (2019) Modeling spatial processes with unknown extremal dependence class. Journal of the American Statistical Association 114(525):434–444, doi:10.1080/01621459.2017.1411813

Hu Y, Scarrott C (2018) evmix: An R package for extreme value mixture modeling, threshold estimation and boundary corrected kernel density estimation. Journal of Statistical Software 84(5):1–27, doi:10.18637/jss.v084.i05

Hu You, Scarrott C (2018) evmix: An R package for extreme value mixture modeling, threshold estimation and boundary corrected kernel density estimation. Journal of Statistical Software 84(5):1–27, doi:10.18637/jss.v084.i05
Huser R, Wadsworth JL (2022) Advances in statistical modeling of spatial extremes. Wiley Interdisciplinary Reviews: Computational Statistics 14(e1537), doi:10.1002/wics.1537
Huser R, Domby C, Ribatet M, Genton MG (2019) Full likelihood inference for max-stable data. Stat 8(1):1–14, doi:10.1002/sta4.218
Huser R, Opitz T, Thibaud E (2021) Max-infinity divisible models and inference for spatial extremes. Scandinavian Journal of Statistics 48(1):321–348, doi:10.1111/sjos.12491
Hüsler J, Reiss RD (1989) Maxima of normal random vectors: between independence and complete dependence. Statistics & Probability Letters 7(4):283–286, doi:10.1016/0167-7152(89)90106-5
Keef C, Tawn JA, Lamb R (2013) Estimating the probability of widespread flood events. Environmetrics 24(1):13–21, doi:10.1002/env.2190
Kiener P (2022) R RWsearch: Lazy search in R packages, task views, CRAN, the web. all-in-one download. R package version 5.0.5
Kiriliouk A, Rootzén H, Segers J, Wadsworth JL (2019) Peaks over thresholds modelling with multivariate generalized Pareto distributions. Technometrics 61(1):123–135, doi:10.1080/00401706.2018.1462738
Kojadinovic I, Yan J (2010) Nonparametric rank-based tests of bivariate extreme-value dependence. Journal of Multivariate Analysis 101(9):2234–2249, doi:10.1016/j.jmva.2010.05.004
Kojadinovic I, Segers J, Yan J (2011) Large-sample tests of extreme-value dependence for multivariate copulas. Canadian Journal of Statistics 39(4):703–720, doi:10.1002/cjs.10110
Krainski E, Gómez-Rubio V, Bakka H, Lenzi A, Castro-Camilo D, Simpson D, Lindgren F, Rue H (2018) Advanced spatial modeling with stochastic partial differential equations using R and INLA. Chapman and Hall/CRC
Ledford AW, Tawn JA (1996) Statistics for near independence in multivariate extreme values. Biometrika 83(1):169–187, doi:10.1093/biomet/83.1.169
Ledford AW, Tawn JA (2003) Diagnostics for dependence within time series extremes. Journal of the Royal Statistical Society Series B (Statistical Methodology) 65(2):521–543, doi:10.1111/1467-9868.00400
Lilienthal J (2022) TLMoments: Calculate TL-moments and convert them to distribution parameters. R package version 0.7.5.3
Lindgren F, Rue H, Lindström J (2011) An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73(4):423–498, doi:10.1111/j.1467-9868.2011.00777.x
Lugrin T, Davison AC, Tawn JA (2016) Bayesian uncertainty management in temporal dependence of extremes. Extremes 19(3), doi:10.1007/s10687-016-0258-0
Manjunath BG, Caeiro F (2013) ext0: Mean of order p, peaks over random threshold Hill and high quantile estimates. R package version 1.1-3
Marcon G, Naveau P, Padoan S (2017a) A semi-parametric stochastic generator for bivariate extreme events. Stat 6(1):184–201, doi:10.1002/sta4.145
Marcon G, Padoan S, Naveau P, Muliere P, Segers J (2017b) Multivariate nonparametric estimation of the Pickands dependence function using Bernstein polynomials. Journal of Statistical Planning and Inference 183:1–17, doi:10.1016/j.jspi.2016.10.004
Martins ES, Stedinger JR (2000) Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. Water Resources Research 36(3):737–744, doi:10.1029/1999WR900330
Martins TG, Simpson D, Lindgren F, Rue H (2013) Bayesian computing with INLA: New features. Computational Statistics and Data Analysis 67:68–83
McNeil A, Frey R, Embrechts P (2015) Quantitative Risk Management: Concepts, Techniques and Tools - Revised Edition. Princeton Series in Finance, Princeton University Press
Mhalla L, Opitz T, Chavez-Demoulin V (2019) Exceedance-based nonlinear regression of tail dependence. Extremes 22(3):523–552, doi:10.1007/s10687-019-00342-6
Moinz T, Arbel J, Girard S, Dutfoy A (2023) Reparameterization of extreme value framework for improved Bayesian workflow. Computational Statistics and Data Analysis (to appear), doi:10.1016/j.csda.2023.107807
Munasinghe R, Kossina P, Jayasinghe D, Wijeratne D (2019) ptauite: Tail index estimation for power law distributions. R package version 1.0.0
Nadarajah S, Chan S, Afuecheta E (2023) VaRES: Computes value at risk and expected shortfall for over 100 parametric distributions. R package version 1.0.2

do Nascimento FF, Gamerman D, Lopes HF (2012) A semiparametric Bayesian approach to extreme value estimation. Statistics and Computing 22(2):661–675, doi:10.1007/s11222-011-9270-z

Naveau P, Huser R, Ribereau P, Hannart A (2016) Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. Water Resources Research 52(4):2753–2769, doi:10.1002/2015WR018552

Newey WK, Powell JL (1987) Asymmetric least squares estimation and testing. Econometrica 55(4):819–847, doi:10.2307/1911031

Nikoloulopoulos AK, Joe H, Li H (2009) Extreme value properties of multivariate t copulas. Extremes 12(2):129–148, doi:10.1007/s10687-008-0072-4

Nolde N, Wadsworth JL (2022) Linking representations for multivariate extremes via a limit set. Advances in Applied Probability 54(3):688–717, doi:10.1017/apr.2021.51

Northrop PJ (2015) An efficient semiparametric maxima estimator of the extremal index. Extremes 18(4):585–603, doi:10.1007/s10687-015-0221-5

Northrop PJ (2022) lite: Likelihood-based inference for time series extremes. R package version 1.1.0

Northrop PJ (2023) revdbayes: Ratio-of-uniforms sampling for Bayesian extreme value analysis. R package version 1.5.1

Northrop PJ, Attalides N (2016) Posterior propriety in Bayesian extreme value analyses using reference priors. Statistica Sinica 26(2):721–743, doi:10.5705/ss.2014.034

Northrop PJ, Attalides N (2020) threshr: Threshold selection and uncertainty for extreme value analysis. R package version 1.0.3

Northrop PJ, Christodoulides C (2022) exdex: Estimation of the extremal index. R package version 1.2.1

Northrop PJ, Coleman CL (2014) Improved threshold diagnostic plots for extreme value analyses. Extremes 17(2):289–303, doi:10.1007/s10687-014-0183-z

Northrop PJ, Yin C (2021) lax: Loglikelihood adjustment for extreme value models. R package version 1.2.0

Northrop PJ, Jonathan P, Randell D (2016) Threshold modeling of nonstationary extremes. In: Dey DK, Yan J (eds) Extreme Value Modeling and Risk Analysis: Methods and Applications, CRC Press, Boca Raton, FL, pp 87–108

Northrop PJ, Attalides N, Jonathan P (2017) Cross-validatory extreme value threshold selection and uncertainty with application to ocean storm severity. Journal of the Royal Statistical Society: Series C (Applied Statistics) 66(1):93–120, doi:10.1111/rssc.12159

Noven R, Veraart A, Gandy A (2018) A latent trawl process model for extreme values. Journal of Energy Markets 11(3):1–24, doi:10.21314/JEM.2018.179

Oesting M (2018) Equivalent representations of max-stable processes via ℓp-norms. Journal of Applied Probability 55(1):54–68, doi:10.1017/jpr.2018.5

Opitz T, Huser R, Bakka H, Rue H (2018) INLA goes extreme: Bayesian tail regression for the estimation of high spatio-temporal quantiles. Extremes 21(3):441–462, doi:10.1007/s10687-018-0324-x

Ossberger J (2020) teq: Threshold estimation approaches. R package version 1.1

Paciorek C, Stone D, Wehner M (2018) Quantifying statistical uncertainty in the attribution of human influence on severe weather. Weather and Climate Extremes 20:69–80, doi:10.1016/j.wace.2018.01.002

Padoan S, Stupfler G (2020) ExtremeRisks: Extreme risk measures. R package version 0.0.4

Padoan SA (2015) Analysis of random fields using CompRandFld. Journal of Statistical Software 63(9):1–27, doi:10.18637/jss.v063.i09

Padoan SA, Stupfler G (2022) Joint inference on extreme expectiles for multivariate heavy-tailed distributions. Bernoulli 28(2):1021 – 1048, doi:10.3150/21-BEJ1375

Papastathopoulos I, Tawn JA (2013) Extended generalised Pareto models for tail estimation. Journal of Statistical Planning and Inference 143(1):131–143, doi:10.1016/j.jspi.2012.07.001

Pfaff B, McNeil A (2018) evir: Extreme values in R. R package version 1.7-4

Pfaff B, McNeil A (2020) QRM: Provides R-language code to examine quantitative risk management concepts. R package version 0.4-31
Pickands J (1981) Multivariate extreme value distributions. In: Proceedings 43rd, Session of International Statistical Institution

Plummer M, Best N, Cowles K, Vines K (2006) CODA: Convergence diagnosis and output analysis for MCMC. R News 6(1):7–11

Prescott P, Walden AT (1980) Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. Biometrika 67(3):723–724, doi:10.1093/biomet/67.3.723

Reich BJ, Shaby BA (2012) A hierarchical max-stable spatial model for extreme precipitation. The Annals of Applied Statistics 6(4):1430–1451, doi:10.1214/12-AOAS591

Reich BJ, Shaby BA (2016) Bayesian inference for extreme value modelling. In: Dipak K Dey IV (ed) Extreme Value Modeling and Risk Analysis: Methods and Applications, CRC Press, Boca Raton, FL, pp 131–151

Resnick S, Stáríča C (1997) Smoothing the Hill estimator. Advances in Applied Probability pp 271–293, doi:10.2307/1428780

Resnick SI (1987) Extreme values, regular variation, and point processes, Applied Probability. A Series of the Applied Probability Trust, vol 4. Springer–Verlag, New York

Reynkens T, Verbelen R (2023) ReIns: Functions from “Reinsurance: Actuarial and Statistical Aspects”. R package version 1.0.12

Ribatet M (2022) SpatialExtremes: Modelling spatial extremes. R package version 2.1-0

Ribatet M, Dutang C (2022) POT: Generalized Pareto Distribution and Peaks Over Threshold. R package version 1.1-10

Rooßen H, Tajvidi N (2006) Multivariate generalized Pareto distributions. Bernoulli 12(5):917–930, doi:10.3150/bj/1161614952

Rooßen H, Zhoulud D (2017) Human life is unlimited — but short (with discussion). Extremes 20(4):713–728

Rooßen H, Segers J, Wadsworth JL (2018) Multivariate peaks over thresholds models. Extremes 21(1):115–145, doi:10.1007/s10687-017-0294-4

Ross E, Astrup OC, Bütter-Gregersen E, Bunn N, Feld G, Gouldby B, Huseby A, Liu Y, Randell D, Vanem E, Jonathan P (2020) On environmental contours for marine and coastal design. Ocean Engineering 195:106194, doi:10.1016/j.oceaneng.2019.106194

Ruckdeschel P, Kohl M, Horbenko N (2019) RobExtremes: Optimally robust estimation for extreme value distributions. R package version 1.2.0

Rudd EM, Jain LR, Scheier WJ, Boulton TE (2018) The extreme value machine. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 40(3):762–768, doi:10.1109/TPAMI.2017.2707495

Rue H, Martino S, Chopin N (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). Journal of the Royal Statistical Society: Series B (Statistical Methodology) 71(2):319–392, doi:10.1111/j.1467-9868.2008.00700.x

Sabourin A, Naveau P (2014) Bayesian Dirichlet mixture model for multivariate extremes: A re-parametrization. Computational Statistics & Data Analysis 71:542–567, doi:10.1016/j.csda.2013.04.021

Sabourin A, Naveau P, Fougeres AL (2013) Bayesian model averaging for multivariate extremes. Extremes 16(3):325–350, doi:10.1007/s10687-012-0163-0

Santos PA, Alves MF, Gomes M (2006) Peaks over random threshold methodology for tail index and quantile estimation. REVSTAT-Statistical Journal 4(3):227–247, doi:10.57805/revstat.v4i3.37

Scarrott C, MacDonald A (2012) A review of extreme value threshold estimation and uncertainty quantification. REVSTAT-Statistical Journal 10(1):33–60, doi:10.57805/revstat.v10i1.110

Schlather M (2002) Models for stationary max-stable random fields. Extremes 5(1):33–44, doi:10.1023/A:1020977924878

Schlather M, Malinowski A, Menck PJ, Oesting M, Strokorb K (2015) Analysis, simulation and prediction of multivariate random fields with package RandomFields. Journal of Statistical Software 63(8):1–25, doi:10.18637/jss.v063.i08

Sébille Q (2016) hexp: Spatial extreme value analysis with the hierarchical model of Reich and Shaby (2012). R package version 1.1.5

Shooter R, Ross E, Ribal A, Young IR, Jonathan P (2021) Spatial dependence of extreme seas in the North East Atlantic from satellite altimeter measurements. Environmetrics 32(4):e2674, doi:10.1002/env.2674
Shooter R, Ross E, Ribal A, Young IR, Jonathan P (2022) Multivariate spatial conditional extremes for extreme ocean environments. Ocean Engineering 247:110647, doi:10.1016/j.oceaneng.2022.110647

Simpson D, Rue H, Riebler A, Martins TG, Sorbye SH (2017) Penalising model component complexity: A principled, practical approach to constructing priors. Statistical Science 32(1):1–28, doi:10.1214/16-STSS576

Simpson ES, Opitz T, Wadsworth JL (2023) High-dimensional modeling of spatial and spatio-temporal conditional extremes using INLA and the SPDE approach. Extremes

Smith RL (1985) Maximum likelihood estimation in a class of nonregular cases. Biometrika 72(1):67–90, doi:10.1093/biomet/72.1.67

Smith RL, Weissman I (1994) Estimating the extremal index. Journal of the Royal Statistical Society Series B (Methodological) 56(3):515–528, doi:10.1111/j.2517-6161.1994.tb01997.x

Smith RL, Tawn JA, Coles SG (1997) Markov chain models for threshold exceedances. Biometrika 84(2):249–268, doi:10.1093/biomet/84.2.249

Southworth H, Heffernan JE, Metcalfe PD (2020) texmex: Statistical modelling of extreme values. R package version 2.4.8

Stan Development Team (2023) RStan: the R interface to Stan. R package version 2.21.8

Stephenson AG (2002) evd: Extreme value distributions. R News 2(2)

Stephenson AG (2003) Simulating multivariate extreme value distributions of logistic type. Extremes 6(1):49–59, doi:10.1023/A:1026277229992

Stephenson AG (2016) Bayesian inference for extreme value modelling. In: Dey DK, Yan J (eds) Extreme Value Modeling and Risk Analysis: Methods and Applications, CRC Press, Boca Raton, FL, pp 257–280

Stephenson AG, Gilleland E (2006) Software for the analysis of extreme events: The current state and future directions. Extremes 8:87–109, doi:10.1007/s10687-006-7962-0

Stephenson AG, Tawn JA (2005) Exploiting occurrence times in likelihood inference for componentwise maxima. Biometrika 92(1):213–227, doi:10.1093/biomet/92.1.213

Süveges M (2007) Likelihood estimation of the extremal index. Extremes 10(1):41–55, doi:10.1007/s10687-007-0034-2

Süveges M, Davison AC (2010) Model misspecification in peaks over threshold analysis. The Annals of Applied Statistics 4(1):203–221, doi:10.1214/09-AOAS292

Tajvidi N, Turlach B (2018) A general approach to generate random variates for multivariate copulae. Australian & New Zealand Journal of Statistics 60(1):140–155, doi:10.1111/anzs.12209

Thibaud E, Opitz T (2015) Efficient inference and simulation for elliptical Pareto processes. Biometrika 102(4):855–870, doi:10.1093/biomet/avv045

Thibaud E, Aalto J, Cooley DS, Davison AC, Heikkinen J (2016) Bayesian inference for the Brown–Resnick process, with an application to extreme low temperatures. The Annals of Applied Statistics 10(4):2303–2324, doi:10.1214/16-AOAS980

Thompson P, Cai Y, Reeve D, Stander J (2009) Automated threshold selection methods for extreme wave analysis. Coastal Engineering 56(10):1013–1021, doi:10.1016/j.coastaleng.2009.06.003

Ulrich J, Jurado OE, Peter M, Scheibl M, Rust HW (2020) Estimating IDF curves consistently over durations with spatial covariates. Water 12(11:3119), doi:10.3390/w12113119

Vehtari A, Gelman A, Gabry J (2017) Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and Computing 27:1413–1432, doi:10.1007/s11222-016-9696-4

Vehtari A, Gabry J, Magnusson M, Yao Y, Bürkner PC, Paananen T, Gelman A (2020) loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models. R package version 2.4.1

Velthoen J, Dombrzy C, Cai JJ, Engelke S (2021) Gradient boosting for extreme quantile regression. arXiv:210300808

Vignotto E, Engeld E (2020) Extreme value theory for anomaly detection – the GPD classifier. Extremes 23(4):501–520, doi:10.1007/s10687-020-00393-0

Wadsworth J, Tawn J (2022) Higher-dimensional spatial extremes via single-site conditioning. Spatial Statistics 51:100677, doi:10.1016/j.spasta.2022.100677

Wadsworth JL (2015) On the occurrence times of componentwise maxima and bias in likelihood inference for multivariate max-stable distributions. Biometrika 102(3):705–711, doi:10.1093/biomet/asv029
Wadsworth JL (2016) Exploiting structure of maximum likelihood estimators for extreme value threshold selection. Technometrics 58(1):116–126, doi:10.1080/00401706.2014.998345

Wadsworth JL, Tawn JA (2013) A new representation for multivariate tail probabilities. Bernoulli 19(5B):2689–2714, doi:10.3150/12-BEJ471

Wager S (2014) Subsampling extremes: From block maxima to smooth tail estimation. Journal of Multivariate Analysis 130:335–353, doi:10.1016/j.jmva.2014.06.010

Weissman I (1978) Estimation of parameters and large quantiles based on the k largest observations. Journal of the American Statistical Association 73:812–815, doi:10.1080/01621459.1978.10480104

Wood S (2017) Generalized Additive Models: An Introduction with R, 2nd edn. Chapman and Hall/CRC

Wood SN, Pya N, Säfken B (2016) Smoothing parameter and model selection for general smooth models. Journal of the American Statistical Association 111(516):1548–1563, doi:10.1080/01621459.2016.1180986

Wertz D, Setz T, Chalabi Y (2017) fExtremes: Rmetrics - Modelling extreme events in finance. R package version 3042.82

Wertz D, Setz T, Chalabi Y, Smith P (2023) fCopulae: Rmetrics — Bivariate Dependence Structures with Copulae. R package version 4022.85

Yan J (2007) Enjoy the joy of copulas: With a package copula. Journal of Statistical Software 21(4):1–21, doi:10.18637/jss.v021.i04

Yee TW, Stephenson AG (2007) Vector generalized linear and additive extreme value models. Extremes 10:1–19, doi:10.1007/s10687-007-0032-4

Youngman BD (2022) evgam: An R package for generalized additive extreme value models. Journal of Statistical Software 103(1):1–26, doi:10.18637/jss.v103.i03

Zhang J, Stephens MA (2009) A new and efficient estimation method for the generalized Pareto distribution. Technometrics 51(3):316–325, doi:10.1198/tech.2009.08017

Zheng F, Leonard M, Westra S (2015) Efficient joint probability analysis of flood risk. Journal of Hydroinformatics 17(4):584–597, doi:10.2166/hydro.2015.052
Table 11: Evaluation of generalized Pareto model density and distribution functions.

| package        | location | density            | distribution function |
|----------------|----------|--------------------|-----------------------|
| eva            | yes      | correct            | correct               |
| evd            | yes      | incorrect for $x = u$ | correct               |
| evir           | yes      | incorrect for $x < u$ | incorrect outside support |
| extraDistr     | yes      | incorrect for $x = u$ | correct               |
| extRemes       | yes      | incorrect for $x = u$ | correct               |
| fExtremes      | yes      | incorrect for $x = u$ | correct               |
| lmom           | yes      | incorrect          | incorrect outside support |
| lmomco         | yes      | correct            | incorrect outside support |
| mev            | yes      | correct            | correct               |
| POT            | yes      | incorrect for $x = u$ | correct               |
| QRM            | no       | incorrect for $x = u$ | correct               |
| qrmtools       | no       | correct            | correct               |
| ReIns          | yes      | correct            | correct               |
| Renext         | yes      | incorrect for $x = u$ | correct               |
| revdbayes      | yes      | correct            | correct               |
| SpatialExtremes| yes      | incorrect for $x = u$ | correct               |
| tea            | yes      | correct            | correct               |
| texmex         | yes      | correct            | correct               |
| TLMoments      | yes      | correct            | correct               |

A. Likelihood inference for univariate extremes

A.1. Density and distribution function checks

We performed some sanity checks for various maximum likelihood estimation routines and parametric model implementations. Specifically, we verified that density functions are non-negative and evaluate to zero outside of the domain of the distribution, and that distribution functions are non-decreasing and map to the unit interval.

The generalized Pareto distribution has lower bound at the location parameter $u$ and is bounded above at $u - \sigma / \xi$ whenever $\xi < 0$. Many software implementations forgo the location parameter, since for modelling large quantiles of a random variable $Y$ above threshold $u$, it suffices to look at threshold exceedances $Y - u > 0$. No threshold exceedance should be exactly equal to zero so the value of the density at that point is immaterial, even if it should be set to zero in practice.

Certain packages, listed in Table 11 and Table 12, have incorrect implementations of density and distribution functions.

A.2. Optimization routines

We compared the maximum likelihood estimates returned by default estimation procedures for different packages for simulated data, checking that the value returned is a global optimum and the gradient is approximately zero whenever $\xi > -1$.

A.2.1 Generalized Pareto distribution

For threshold exceedances, we simulated 50 exceedances from a generalized Pareto distribution $\text{GP}(\sigma = 1000, \xi = -0.5)$ and from an exponential distribution with $\sigma = 1000$. The large scale value is intended to check the robustness of gradient-based algorithms; from an optimization perspective, it is wise to ensure that the gradient of each component, scale and shape, are not magnitudes apart. The data can easily be scaled prior to the optimization in case this is problematic.

Figure 3 shows the distribution of the score vector, i.e., the gradient of the log likelihood. The latter should vanish when evaluated at the maximum likelihood estimator $(\hat{\theta}, \hat{\xi})$ provided $\hat{\xi} > -1$. Most instances of non-zero gradient are attributable to boundary cases with $\xi = -1$ not accounted for. Other
Figure 3: Magnitude of the shape component of the score vector at the value returned by the optimization routine. The density plots are based on 1000 samples simulated from a generalized Pareto distribution with shape $\xi = -0.5$ and scale $\sigma = 1000$, split by simulations yielding a boundary case ($\xi = -1$, gray) and regular case ($\xi > -1$, black); the y-axis scale for each package is different to ease visualization. Results for samples for which the numerical routines failed to converge or the gradient is unevaluated are not shown.
Table 12: Evaluation of generalized extreme value density and distribution functions.

| package            | density | distribution function |
|--------------------|---------|-----------------------|
| EnvStats           | correct | correct               |
| evd                | correct | correct               |
| evir               | correct | correct               |
| extraDistr         | correct | correct               |
| ExtremalDep        | correct | correct               |
| extRemes           | correct | correct               |
| fExtremes          | correct | correct               |
| lmomco             | incorrect for $x < \mu$ | incorrect for $x < \mu$ |
| mev                | correct | correct               |
| QRM                | correct | correct               |
| qrmtools           | correct | correct               |
| revdbayes          | correct | correct               |
| SpatialExtremes    | correct | correct               |
| texmex             | correct | correct               |
| TLMoments          | correct | correct               |

discrepancies are due to numerical tolerance for convergence, but the differences in log likelihood relative to the maximum over all routines are negligible in most non-boundary cases investigated. Some routines, based on Nelder–Mead simplex algorithm, do not check the gradient but this is immaterial if the value of the function is nearly identical to that at the maximum likelihood estimate.

Figure 4 shows these differences through survival function plots, highlighting instances where the package fails to return correct values. Most packages do fine, except for a handful: evd, extRemes and POT (which uses routines from evd) stand out of the lot.

We can figure out the source of some of these oddities by plotting the distribution of the shape parameter estimates over all 1000 replications (see Figure 5). Both POT and evd return sampling distributions that are underdispersed relative to other implementations, while ercV and extRemes both have a large number of runs that return exactly zero for the shape parameter. The QRM package has unexpectedly small spread and a positive bias for estimation of $\xi$, different from other packages because it fails more often when $\xi$ is negative. Both ercv and extRemes routines return zero shape estimates, leading to noticeable point masses. Only SpatialExtremes and mev correctly return $\xi = -1$, while rMext returns a hard-coded lower bound which can also be set to $\xi = -1$.

Some packages have routines that fail to converge often when the shape is negative; the most likely culprit for this is poor starting values. The routines in ercv and fExtremes (same as evir) fail often in small samples: for $n = 20$ exceedances, the function returned an error in 225 simulations. For the latter, the error is due to poor implementation of the log-likelihood that leads to infinite finite differences between estimates. For QRM, the choice of starting values, which cannot be modified by the user, is not adequate with strong negative shapes: it failed in more than 50 ($n = 20$), 122 ($n = 50$), 169 ($n = 100$) and 253 ($n = 1000$) times for negative shapes, indicating that the issue is not sample size. The qrmtools package, which supersedes QRM, has no such problems.

A.2.2 Generalized extreme value distribution

The optimization routines for the generalized extreme value distribution with scale $\sigma = 1000$ and shape parameters $\xi \in \{-0.5, 0, 0.5\}$ are better behaved and nearly all packages give identical results: only evd and texmex failed to converge and returned abnormally high shape values in a handful of instances out of 1000 simulations.

Unsurprisingly, the portrait (see Figures 6 and 7) is the same for the generalized extreme value distribution when it comes to boundary constraints: for example, climextRemes does not return shapes less than or equal to $-1$. extRemes has odd behaviour with a visible point mass at $\xi = 0$ in the simulations, even when this value has measure zero. Only mev and SpatialExtremes handle the
Figure 4: Differences between the likelihood evaluated at the parameters returned by the routines and the maximum likelihood over all routines for generalized Pareto samples with negative shape ($\xi = -0.5$, left) and exponential samples (right), both with large scale parameter $\sigma = 1000$. Results for samples for which the numerical routines failed to converge are not shown. Only packages with 90\% percentile giving a discrepancy larger than 10\(^{-4}\) are shown.

Figure 5: Dot plots of shape parameter estimates returned by optimization routine for generalized Pareto samples with negative shape ($\xi = -0.5$, left) and exponential samples (right). Results for samples for which the numerical routines failed to converge are not shown.
boundary constraints. Figure 6 shows the difference in maximum likelihood returned by the packages, excluding cases with $\xi = -1$ for which the log likelihood becomes unbounded for combinations of $\sigma$ and $\xi < -1$. Some packages, such as evd, also sometimes return a local optimum (perhaps due to use of the BFGS routine) and this in turn leads to erroneous comparisons of nested models.

Table 14 gives a breakdown of the number of instances for which the maximisation routine failed: two packages, climextRemes and EnvStats, stand out for negative shapes and the percentage of failures increases with the sample size.
Figure 7: Dot plots of shape parameter estimates returned by optimization routines for generalized extreme value samples with negative shape ($\xi = -0.5$, left) and Gumbel samples (right). Results for samples for which the numerical routines failed to converge are not shown.

Table 13: Number of failures for the optimization routine for maximum likelihood-based estimation of the generalized Pareto model (out of 1000 simulations).

(a) bounded tail ($\xi = -0.5$)

|       | 20  | 50  | 100 | 1000 |
|-------|-----|-----|-----|------|
| evir  | 225 | 15  | 0   | 0    |
| fExtremes | 225 | 15  | 0   | 0    |
| QRM   | 50  | 122 | 169 | 253  |

(b) exponential ($\xi = 0$)

|       | 20  | 50  | 100 | 1000 |
|-------|-----|-----|-----|------|
| evir  | 37  | 0   | 0   | 0    |
| fExtremes | 37  | 0   | 0   | 0    |
| QRM   | 4   | 12  | 7   | 0    |

(c) heavy tail ($\xi = 0.5$)

|       | 20  | 50  | 100 | 1000 |
|-------|-----|-----|-----|------|
| evir  | 7   | 0   | 0   | 0    |
| fExtremes | 7   | 0   | 0   | 0    |
| QRM   | 1   | 0   | 0   | 0    |
Table 14: Number of failures for the optimization routine for maximum likelihood-based estimation of the generalized extreme value model (out of 1000 simulations).

(a) bounded tail ($\xi = -0.5$)

|         | 100  | 1000 | 20   | 50   |
|---------|------|------|------|------|
| climextRemes | 151  | 200  | 172  | 127  |
| EnvStats   | 156  | 215  | 100  | 131  |
| evir       | 0    | 0    | 27   | 0    |
| fExtremes  | 23   | 0    | 92   | 28   |
| ismev      | 0    | 0    | 4    | 0    |
| mev        | 0    | 0    | 11   | 0    |
| texmex     | 0    | 0    | 3    | 0    |

(b) light tail ($\xi = 0$)

|         | 100  | 1000 | 20   | 50   |
|---------|------|------|------|------|
| climextRemes | 0    | 0    | 4    | 0    |
| EnvStats   | 0    | 0    | 4    | 0    |
| fExtremes  | 0    | 0    | 1    | 0    |

(c) heavy tail ($\xi = 0.5$)

|         | 100  | 1000 | 20   | 50   |
|---------|------|------|------|------|
| climextRemes | 2    | 0    | 1    | 1    |
| EnvStats   | 9    | 1    | 10   | 10   |
| evir       | 0    | 0    | 4    | 0    |
| fExtremes  | 2    | 0    | 5    | 3    |
B. Bayesian univariate inference for extremes

Creating a benchmark for Bayesian univariate analysis of extremes is complicated because approximate posterior samples returned by Markov chain Monte Carlo are autocorrelated so we cannot rely only on speed of execution or correctness. The effective sample size, which measures the equivalent number of independent draws from the posterior, is a better unit than the number of draws returned. We also factor in the amount of time it takes for the algorithm to proceed, but note that higher initialization costs may make such comparisons unfair if the cost of setup is larger than that of sampling. The different packages use different sampling algorithms: those that are implemented in low-level programming languages like C are inherently faster. Most packages that use random walk Metropolis–Hastings steps discard initial draws during the so-called burn-in period (sometimes to tune proposal standard deviations, mostly to let the chain reach the posterior distribution and reduce impact of starting values). Other considerations include flexibility of methods, the choice of likelihood or the possibility to include covariates.

- In the `texmex` package, users can run multiple chains with burn-in and thinning, but iterations are preserved (which results in a heavier memory footprint). The choice of prior is restricted relative to most other packages. Rather than random walk Metropolis steps, proposals are drawn independently from a distribution which is centered at the maximum a posteriori, with a scale matrix matching the Hessian at the mode. This allows for good mixing, at the expense of a preliminary optimization (and tentatively terrible results should the latter fail to converge to the maximum a posteriori distribution).

- The `evdbayes` package has a comprehensive documentation, but some of its features are unconventional: the generalized Pareto model includes a location parameter that is modeled along with the threshold, but this is typically fixed. This leads to many proposals for the random walk Metropolis–Hastings ratio that lead to negative infinity, so we discard this altogether from the comparison. It can lead to adaption of the proposal.

- While very flexible, `extRemes` is noticeably slower than other packages and particularly inefficient with the default options (not setting `proposalParams` leads to effective sample sizes that are insufficient for any analysis to be reliable in our examples). It can be somewhat customized (and includes more flexible prior specification), but there is limited documentation on how to complete this in the package itself (but see the accompanying Journal of Statistical Software paper). The current options for evaluating the marginal likelihood in `BayesFactor` are unreliable and shouldn’t be used (e.g., Neal, 2018).

- The `ExtremalDep` package also allows for estimation of the generalized extreme value distribution with potential covariates for the location parameter and censoring below a marginal threshold, using a random walk Metropolis-Hastings algorithm. However, the user needs to provide starting values and default values for the variances of the multivariate normal proposals, `sigma`, and the code returns an error if there is no censoring and covariates are provided. The output is less user-friendly than other packages, as there are no methods associated with the returned list.

- `revdbayes` provides independent draws from the posterior at a fraction of the costs of the other packages. Unless one has to include covariates in the parameters, it is the recommended approach.

B.1. Evaluation of effectiveness

We look at computation time (Figure 8) and effectiveness (Table 15) of algorithms as measured by the effective sample size, computed using the `coda` method (based on autocorrelation of the chains). Alternative better methods exist based on running multiple chains, but we forgo these.

The `revdbayes` implementation is exact and fastest, thus should be privileged in any problem not involving covariates. `Stan` simulates posterior samples using a Hamiltonian Monte Carlo algorithm. The latter is much more efficient than Metropolis-Hastings random walk proposals since it uses information about the geometry of the posterior distribution: the programming language requires bespoke definitions of the extreme value models and some care is necessary for shapes close to zero for the GEV distribution.
Table 15: Effective sample size divided by the number of iterations (percentage).

(a) nonstationary generalized extreme value model

| package   | loc | loc (trend) | scale | shape |
|-----------|-----|-------------|-------|-------|
| evdbayes  | 10.1| 4.6         | 9.5   | 13.2  |
| extRemes  | 6.1 | 6.9         | 6.9   | 4.9   |
| STAN      | 90.0| 86.3        | 100.0 | 69.7  |
| texmex    | 6.6 | 7.0         | 7.0   | 7.2   |

(b) generalized Pareto model

| package       | scale | shape |
|---------------|-------|-------|
| extRemes      | 4.5   | 4.3   |
| MCMC4extremes | 6.6   | 7.1   |
| STAN          | 48.3  | 43.8  |
| texmex        | 12.4  | 10.3  |

Of all the remaining packages, \texttt{texmex} gives the best performance because it uses proposals informed by the maximum a posteriori estimate. This wouldn’t necessarily work with a multimodal objective function, but seems to do a good job in the simple scenarios we considered (and which are supported by the package). While we cannot know if we have converged to the target posterior distribution, the chains appear stationary.

The algorithm for \texttt{MCMC4Extremes} is fast, considering the number of observations it samples, but the implementation is crude and inefficient, including a burn-in period of 50K simulations, contrary to what the documentation states. The function is also not customizable.

The performance of \texttt{extRemes} is more dependent on tuning parameters than other implementations. Initial trials with the default parameter revealed problems: while the model starts at the MLE (so close to the stationary distribution), the default standard deviation of the normal random walk proposals are particularly ill-suited to the Venice sea level example. Trace plots (not shown) revealed lack of stationarity with default tuning parameters. With adapted proposals (and vague priors), the output seems satisfactory, but the effective sample size is subpar compared to other methods.

The \texttt{evdbayes} package includes a generalized Pareto model, but the latter also has a location parameter so is not directly comparable with other outputs.

Figure 8: Swarm plot of execution time (including preliminary optimization if necessary) of different numerical routines for a generalized extreme value model with linear trend in location fitted to the Venice sea level data (left) and the generalized Pareto distribution fitted to the Eskdalemuir rainfall data (right).
Data availability
The datasets analysed in Section 2 are available from the mev package.
## C. Software version

| Package                  | Version | License  | Package        | Version | License  |
|--------------------------|---------|----------|----------------|---------|----------|
| BMAmevt                  | 1.0.5   | GPL (≥ 2) | jointPm        | 2.3.2   | GPL (≥ 2) |
| climextRemes             | 0.3.0   | BSD-3-clause, † | laeken        | 0.5.2   | GPL (≥ 2) |
| coda                     | 0.19-4  | GPL (≥ 2) | lax            | 1.2.0   | GPL (≥ 2) |
| ComplPlAndFl               | 1.0.3-6 | GPL (≥ 2) | lite          | 1.1.0   | GPL (≥ 2) |
| copula                   | 1.1-2   | GPL (≥ 3), † | lmom       | 2.9     | GPL-1.0  |
| CTRE ⋆                   | 0.1.0   | GPL-3    | lmomco        | 2.4.7   | GPL      |
| erc                      | 1.0.1   | GPL (≥ 2) | Lmoments     | 1.3-1   | GPL-2    |
| erf ‡                    | 0.0.1   | GPL-3    | lmomRFA      | 3.5     | GPL-1.0  |
| eva                      | 0.2.6   | GPL (≥ 2) | loo           | 2.6.0   | GPL (≥ 3) |
| ev                          | 2.3-6.1 | GPL-3   | longevity ‡   | 2023-03.22 | GPL-3   |
| evdhabayes               | 1.1-3   | GPL (≥ 2) | MCMC4Extremes | 1.1     | GPL-2    |
| evgam                    | 1.0.0   | GPL-3    | mev           | 1.15    | GPL-3    |
| evir                     | 1.7-4   | GPL (≥ 2) | mgcv          | 1.8-42  | GPL (≥ 2) |
| evmix                    | 2.12    | GPL-3    | mvPot         | 0.1.5   | GPL-2    |
| evtclass                  | 1.0     | GPL-3    | POT           | 1.1-10  | GPL (≥ 2) |
| exdex                    | 1.2.1   | GPL (≥ 2) | ptsuite      | 1.0.0   | GPL-3    |
| ExtremalDep               | 0.0.4-0 | GPL (≥ 2) | QRM          | 0.4-31  | GPL (≥ 2) |
| extremefit                | 1.0.2   | GPL-2    | qrmtools      | 0.0-16  | GPL (≥ 3), † |
| ExtremeRisks             | 0.0.4   | GPL (≥ 2) | RandomFields ⋆ | 3.3-14  | GPL (≥ 3) |
| extRemes                 | 2.1-3   | GPL (≥ 2) | rbm ‡         | 1.0.0   | MIT      |
| extremeStat              | 1.5.5   | GPL (≥ 2) | Rels          | 1.0.12  | GPL (≥ 2) |
| extremis                  | 1.2.1   | GPL (≥ 3) | Renext       | 3.1-3   | GPL (≥ 2) |
| extremogram               | 1.0.2   | GPL-3    | revdbayes    | 1.5.1   | GPL (≥ 2) |
| EVcopula ‡               | 0.1     | GPL-3    | RobExtremes  | 1.2.0   | LGPL-3   |
| evt0                     | 1.1-1   | GPL (≥ 2) | RTDE         | 0.2-1   | GPL (≥ 2) |
| evtrail ⋆                | 0.1.0   | MIT †    | SimCop       | 0.7-0   | GPL (≥ 2) |
| fCopulae             | 4022.85 | GPL (≥ 2) | spatialADAI ‡ | 0.1-0   | none     |
| fExtremes                  | 4021.83 | GPL (≥ 2) | SpatialExtremes | 2.1-0   | GPL (≥ 2) |
| futureheatwaves          | 1.0.3   | GPL-2    | tailDepFun   | 1.0.1   | GPL-3    |
| GEVcdn                   | 1.1.6-2 | GPL-3    | tea          | 1.1     | GPL-3    |
| graphicalExtremes         | 0.2.0   | GPL-3    |              |         |          |
| gremes ⋆                 | 0.1.1   | GPL-2    | texmex       | 2.4.8   | GPL (≥ 2) |
| hkevp                    | 1.1.5   | GPL    | threshr ‡     | 1.0-3   | GPL (≥ 2) |
| IDf                      | 2.1.2   | GPL (≥ 2) | TLMoments    | 0.7-5.3 | GPL (≥ 2) |
| INLA ‡                  | 22.12.16 | GPL-2 | txxtreme     | 0.3-3  | GPL (≥ 2) |
| ismex                    | 1.42    | GPL (≥ 2) | VaRES        | 1.0-2   | GPL (≥ 2) |

Table 16: List of R packages, software licenses and version numbers at the time of the review. Additional file licenses are denoted with a † and packages archived from CRAN at the time of writing are denoted with a star ⋆. Packages available only from Github repository or personal websites are denoted with a ‡.