A modeler's guide to extreme value software

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

Compiled July 4, 2023

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.

Keywords Extreme values · Software · Threshold selection · R programming language

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.

Léo R. Belzile
Département de sciences de la décision, HEC Montréal, Canada, E-mail: `leo.belzile@hec.ca`

Christophe Dutang
Université Grenoble Alpes, CNRS, Grenoble INP, LJK, Grenoble, France

Paul J. Northrop
Department of Statistical Science, University College London, United Kingdom

Thomas Opitz
Biostatistics and Spatial Processes, INRAE, Avignon, France

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 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

$$\lim_{n\to\infty} \Pr\left(\frac{\max_{i=1}^{n} Y_i - b_n}{a_n} \leq x\right) = \begin{cases} \exp\left\{-\left(1+\xi\frac{x-\mu}{\sigma}\right)_+^{-1/\xi}\right\}, & \xi \neq 0, \\ \exp\left\{-\exp\left(-\frac{x-\mu}{\sigma}\right)\right\}, & \xi = 0, \end{cases} \tag{2.1}$$

where $x_+ = \max\{x, 0\}$. The right-hand side of Equation (2.1) 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 ξ 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

$$\lim_{u\to x^*} \frac{\Pr(a_u^{-1}Y > x + u)}{\Pr(a_u^{-1}Y > u)} = \bar{G}(x), \tag{2.2}$$

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

$$\bar{G}(x) = \begin{cases} (1+\xi x/\sigma_u)_+^{-1/\xi}, & \xi \neq 0, \\ \exp(-x/\sigma_u), & \xi = 0, \end{cases} \tag{2.3}$$

with $\sigma_u = \sigma + \xi(u-\mu)$. The right-hand side of Equation (2.3) is the survival function of the generalized Pareto distribution with scale σ_u and shape $\xi \in \mathbb{R}$. The unconditional distribution of F above u is $\Pr(Y > x + u) \approx \bar{G}(x)\Pr(Y > u)$. 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 $\boldsymbol{\theta}$ denote the p-vector of parameters of the extreme value model under consideration, e.g., $\boldsymbol{\theta} = (\mu, \sigma, \xi)^\top$ for the generalized extreme value distribution. We can approximate the log likelihood $\ell(\boldsymbol{\theta})$ by taking the limiting relations of, e.g., Equations (2.1) and (2.2), 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_n, b_n, 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 $\partial\ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ must be zero when evaluated at the MLE unless $\widehat{\xi} = -1$. 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 2.1 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, $I_{\xi\xi} = f(\xi)/\xi^{-4}$ (Prescott and Walden 1980), and the limit as $\xi \to 0$ is well-defined, but this expression is numerically unstable when $\xi \approx 0$. High precision functions such as `log1p` 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 estimation procedures and the exponential sub-case, for which the profile likelihood is unbounded, can be easily handled separately. The left panel of Figure 2.1 shows profile log likelihoods for two simulated datasets, including one for which $\widehat{\xi}$ lies on the boundary of the parameter space.

Reparametrization: We can sometimes reparametrize models to facilitate interpretation and make explicit the equivalence between various representations of the extremal types theorem. Suppose we model the n_u largest observations from the observed n sample, denoted $y_{(n)} \geq \cdots \geq y_{(n-n_u+1)} > u \geq y_{(n-n_u)}$. Coles (2001, Section 7.5) suggests to write the log-likelihood obtained through the limiting inhomogeneous Poisson point process as

$$\begin{aligned}\ell(\mu,\sigma,\xi;\boldsymbol{y}) = -n_u \log(c\sigma) - \sum_{i=1}^{n_u} \left(1+\frac{1}{\xi}\right) \log\left\{1+\xi\left(\frac{y_{(n-i+1)}-\mu}{\sigma}\right)\right\}_+ \\ - c\left\{1+\xi\left(\frac{u-\mu}{\sigma}\right)\right\}_+^{-1/\xi}, \qquad \mu,\xi \in \mathbb{R}, \sigma > 0. \qquad (2.4)\end{aligned}$$

The constant c is introduced as a way to relate the parameters of the point process likelihood to those of the generalized extreme value distribution fitted to the maximum of blocks of m observations if one sets $c = n/m$. This parametrization however induces strong correlation between the parameters (μ, σ, ξ) so isn't suitable for optimization: the right panel of Figure 2.1 shows how the support constraints lead to a multiple local maxima.

The MLE of the parameters of the inhomogeneous Poisson point process are notably hard to obtain because of this: the optimization in packages such as **ismev** (Heffernan and Stephenson. 2018; Coles 2001) or **evd** (Stephenson 2002) often fails to converge, mostly because of poor starting values. The invariance property of maximum likelihood estimators means that we can reparametrize the model to facilitate optimization: for example, Moins et al. (2023) propose a reparametrization that ensures orthogonality of the parameters of Equation (2.4), but the following trick can also help facilitate convergence: if the estimated probability of exceedance is small, the Poisson approximation implies

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

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

$$\mu_0 = u - \sigma_0\{(n_u/c)^{-\widehat{\xi}} - 1\}/\widehat{\xi}, \qquad \sigma_0 = \widehat{\sigma}_u (n_u/c)^{\widehat{\xi}}, \qquad \xi_0 = \widehat{\xi}.$$

Regularity conditions and implications: Moments of some of the kth 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 $\widehat{\xi} <$

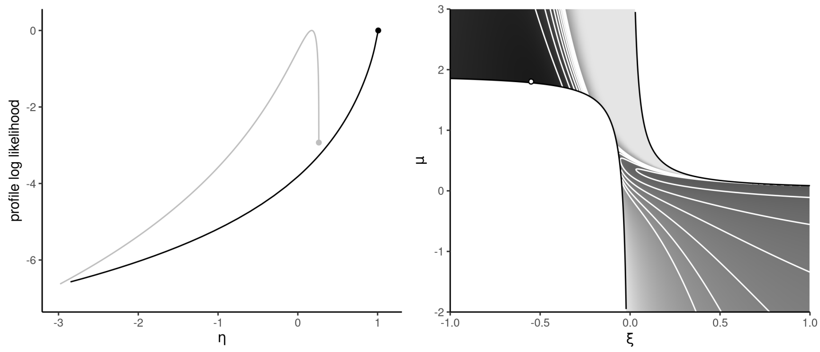

Fig. 2.1 Left: profile log likelihood of $\eta = -\xi/\sigma$ for a generalized Pareto distribution with scale σ and shape ξ. The lines show one data set for which the conditional maximum likelihood of the shape parameter lies on the boundary of the parameter space ($\widehat{\xi}_{\widehat{\eta}} = -1$, black) and one where it exceeds -1 (grey). Right: conditional log likelihood surface for the inhomogeneous Poisson process at $\widehat{\sigma}$ for simulated data (larger values have darker grey-scale shade). The white dot indicates the maximum likelihood estimate, while the hyperbola defines the feasible region of the parameter space given by the support constraints.

-1, as there exists a combination of parameters that lead to infinite log likelihood values. This means one should restrict the parameter space $\mathbb{S}$ to $\mathbb{S} \cap \{\xi : \xi \geq -1\}$ 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 $\widehat{\mu}_{\xi=-1}(x_1, \ldots, x_n) = \overline{x}$, the sample mean, and $\widehat{\sigma}_{\xi=-1}(x_1, \ldots, x_n) = \max_i x_i - \overline{x}$. Similarly, for the generalized Pareto distribution, $\widehat{\sigma}_{\xi=-1}(x_1, \ldots, x_n) = \max_i x_i$. For the likelihood of the r-largest order statistics fitted to vectors of size r, $\widehat{\sigma}_{\xi=-1}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \{\max_i x_{(n),i} - \overline{x}_{(n-r+1)}\}/r, \widehat{\mu}_{\xi=-1} = \max_i x_{(n),i} - \widehat{\sigma}$, where $\overline{x}_{(n-r+1)}$ 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 $\widehat{\xi} > -1$. 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 2.2 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 ξ, different from other packages because it fails more often when ξ 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 2.2 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 $\widehat{\xi} < -1$: 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 non-

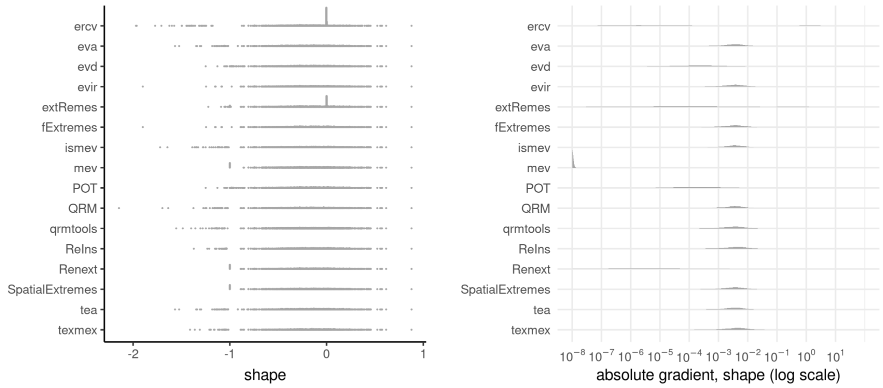

Fig. 2.2 Left: sampling distribution (dot plots) of generalized Pareto shape parameter estimates according to different packages. Right: absolute value of log gradient $\partial\ell/\partial\xi$ evaluated at the maximum likelihood estimator $(\widehat{\sigma}, \widehat{\xi})$ on the log-scale with base 10. Results for samples for which the numerical routines failed to converge are omitted.

stationary 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., B-splines), with a penalty that controls the wiggliness of the estimated predictor functions. Fitting regression (or *multilevel*) models is natural in the Bayesian setting, and many of the packages discussed in the next section have capabilities for fitting multilevel models. There usually is a natural Bayesian interpretation to such penalties: for example, quadratic penalty terms correspond to multivariate Gaussian prior distribution on the regression coefficients.

The obvious difficulty for numerical maximization of the log likelihood is again the presence of support constraints, since there are now potentially as many inequality constraints as there are observations. A general advice for models with covariates is that inputs should be centered and scaled to facilitate the optimization. Table 2.1 provides the list of packages allowing for regression models: the value in the column 'type' is either 'linear' for generalized linear models, 'GAM' for generalized additive models or 'neural network'; the column 'link' takes values in 'custom' for user-supplied functions or a link function as in base R. The **ismev**, **texmex**, **eva** and **extRemes** packages allow users to provide a model matrix (containing one covariate in each of its columns) for each parameter of the generalized Pareto and generalized extreme value distributions, thus enabling generalized linear modelling of the parameters, while the **evd** package only allows for linear modelling of the location parameter of the generalized extreme value distribution and bivariate counterparts; in both cases, no penalty terms are added to the log likelihood, while **texmex** allows for L_1 and L_2 penalties for the coefficients. The **lax** package (Northrop and Yin 2021) supplements the functionality of these, and other, packages by providing robust sandwich estimation of parameter covariance matrix and log likelihood adjustment (Chandler and Bate 2007) for their fitted model objects. The **GEVcdn** package uses a neural network to relate the parameters of the generalized extreme value distribution with covariates (Cannon 2010), while the recent **pinnEV** package allows fitting of "partially-interpretable"

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

Table 2.1 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.

neural networks for parameters using suitably penalized log likelihood functions for the generalized Pareto (different parametrizations), the blended generalized extreme value model and point process representation of the latter. The Matlab package **PPL-model** performs penalised piecewise-linear peaks-over-threshold regression modelling using one- or two-dimensional covariates (Barlow et al. 2023).

The scale parameter of the generalized Pareto distribution, σ_u, varies with the threshold u: it is recommendable to pay special attention to the parametrization of the scale and shape functions with covariates to ensure that the threshold stability property, which is used for extrapolation, is not lost (Eastoe and Tawn 2009). It may be tempting to use directly the likelihood of eq. (2.4) instead (see Northrop et al. 2016). Chavez-Demoulin and Davison (2005) use an orthogonal reparametrization (η, ξ), where $\eta = \sigma(1+\xi)$ along with bootstrap routines for uncertainty quantification; their generalized additive modelling framework is available via **ismev**.

Many general packages implement generalized additive modelling with some support for extreme value distributions, including **VGAM** (Yee and Stephenson 2007). The recent **evgam** package (Youngman 2022), dedicated to extreme value models, uses the methodology of Wood et al. (2016) for general distributions to marginalize out the regression coefficients using Laplace's method to obtain estimates of the hyperparameters (e.g., variance and autocorrelation of regression coefficients) controlling the penalty strength and shape — these are estimated simultaneously with all of the other parameters through maximum likelihood. The **evgam** package builds on generic model building tools available in packages such as **mgcv** (Wood 2017) and provides state-of-the-art methodology tailored for extremes, including generalized additive models for extreme value distributions, quantile regression and in addition functionalities for obtaining return levels for nonstationary models. Carrer and Gaetan (2022) propose an extension of the **gamlss** package for regression modelling of parameters of the extended generalized Pareto model of Naveau et al. (2016).

2.4 Bayesian modelling

2.4.1 Generalities

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

$$p(\boldsymbol{\theta} \mid \boldsymbol{Y}) = \frac{p(\boldsymbol{Y} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})}{\int p(\boldsymbol{Y} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}}, \tag{2.5}$$

is proportional, as a function of $\boldsymbol{\theta}$, to the product of the likelihood and the priors in the numerator. The posterior density $p(\boldsymbol{\theta} \mid \boldsymbol{Y})$ usually does not correspond to any well-known distribution family and the integral appearing in the denominator of Equation (2.5) is therefore untractable in general. Posterior inferences about the components of $\boldsymbol{\theta}$ further involve marginalizing out the other components. For instance, to obtain the posterior density $p(\theta_1 \mid \boldsymbol{Y})$ of the first parameter in $\boldsymbol{\theta}$, we have to evaluate the $(m-1)$-dimensional integral $\int p(\boldsymbol{\theta} \mid \boldsymbol{Y})\, \mathrm{d}(\theta_2, \ldots, \theta_m)$. 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 $p(\boldsymbol{\theta} \mid \boldsymbol{Y})$, 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 $p(\boldsymbol{Y} \mid \boldsymbol{\theta}^{(b)})$ one new observation for each draw of $\boldsymbol{\theta}^{(b)}$ 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 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 ξ may need to be truncated (Northrop and Attalides 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 ξ 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 2.2 lists packages for Bayesian univariate models, where the 'covariates' column lists the parameters which are allowed to depend on covariates (`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

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

Table 2.2 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/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 **evdbayes**, also provide `S3` methods (notably `plot` and `summary`). All packages return a matrix of posterior draws.

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.

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 nonstationary 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 **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 ξ and potentially second-order regular variation indices, ρ. 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 $S(x) = x^{-1/\xi} L_F(x)$ and the quantile function as $Q(1 - 1/x) = x^{\xi} L_U(x)$, with L_F and L_U slowly varying functions, meaning $\lim_{x\to\infty} L(tx)/L(x) = 1$ 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,

$$H_{k,n} = \frac{1}{k} \sum_{j=1}^{k} \log \left(\frac{Y_{(n-j+1)}}{Y_{(n-k)}} \right). \tag{2.6}$$

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 (2.6) 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), **ReIns** (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 (Wager 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 **ReIns** 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 adaptative 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 θ and probability weighted moments of the form $\mathrm{E}[Y^p F(Y)^q \{1-F(Y)\}^r]$ 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

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

where $H_{k,n}$ is the Hill estimator eq. (2.6) of the shape parameter and the threshold is $Y_{(n-k)}$,the $(n-k)$th order statistic. **ReIns** 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 **Expectrem**, which also allows for the possibility to use bias-reduced tail index estimators.

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

Table 2.3 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 2.3 for an overview of available implementations). In the univariate case, the focus is often on the shape parameter ξ: we choose the lowest threshold above which we judge the point estimates of ξ 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 (2.3) for threshold excesses or the inhomogeneous Poisson process model, using a frequentist or Bayesian analysis. In the generalized Pareto case, the threshold-independent scale parameter $\sigma_u^* = \sigma_u - \xi u$ 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 $\mathrm{E}(Y - v \mid Y > v)$ 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 2.4 summarises the functionality of R packages in terms of these plots.

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	`mrl.plot`
mev	`tstab.egp`, `tstab.gpd`	1,3	yes	MLE/B	`automrl`
POT	`tcplot`	1	no	MLE	`mrlplot`
QRM	`xiplot`	1	no	MLE	`MEplot`
ReIns	`1Dmle`	1	—	MLE	`MeanExcess`
texmex	`egp3RangeFit`, `gpdRangeFit`	1,3	no	MLE/B	`mrl`
threshr	`stability`	1	yes	MLE	—

Table 2.4 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.

The `lmomplot` function in the **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 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

type	methods	package	function
penultimate	Northrop and Coleman (2014)	**mev**	`NC.diag`
	Papastathopoulos and Tawn (2013)	**mev**	`tstab.egp`
goodness-of-fit	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)	**threshr**	`ithresh`
mixture	Hu and Scarrott (2018)	**evmix**	—
	Durrieu et al. (2015)	**extremefit**	`·paretomix`
	Naveau et al. (2016)	**mev**	`fit.extgp`

Table 2.5 Overview of formal threshold selection methods and numerical implementations

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 χ^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).

Predictive performance. The **threshr** 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 $\{\boldsymbol{Y}_i\}_{i\geq 1}$, where each vector $\boldsymbol{Y}_i$ has marginal distribution functions F_j ($j = 1,\ldots,D$). By analogy with the univariate case, we consider the random vector of componentwise maxima $\boldsymbol{M}_n = (M_{n,1},\ldots,M_{n,D})$, where $M_{n,j} = \max\{Y_{1,j},\ldots,Y_{n,j}\}$. If there exists sequences of location and scale vectors $\boldsymbol{a}_n \in \mathbb{R}^D_+$ and $\boldsymbol{b}_n \in \mathbb{R}^D$ such that

$$\lim_{n\to\infty} \Pr\left\{\boldsymbol{a}_n^{-1}(\boldsymbol{M}_n - \boldsymbol{b}_n)\right\} = G(\boldsymbol{y}),$$

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 $\boldsymbol{\mu} = \boldsymbol{0}_D$ and $\boldsymbol{\sigma} = \boldsymbol{1}_D$. With $t(\boldsymbol{y})$ denoting a transformed vector whose jth component is $(1+\xi_j y_j)_+^{1/\xi_j}$, the limiting max-stable distribution is

$$G(\boldsymbol{y}) = \exp\left\{-D\int_{\mathbb{S}_D} \max\left\{\frac{\boldsymbol{w}}{t(\boldsymbol{y})}\right\}\mathrm{d}H(\boldsymbol{w})\right\}, \tag{3.1}$$

where the so-called spectral measure H is a probability measure on the D-simplex $\mathbb{S}_D = \{\boldsymbol{\omega} \in \mathbb{R}^D_+ : \|\boldsymbol{\omega}\|_1 = 1\}$. The distribution H must only satisfy the moment constraints $\mathrm{E}(S_j) = 1/D$ ($j = 1,\ldots,D$) for $\boldsymbol{S} \sim H$: the set of probability measures satisfying these is infinite, unlike in the univariate case. The **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 **mev**.

Likelihood-based estimation. The likelihood of a simple max-stable random vector $\boldsymbol{Z}$ with a parametric model for V is obtained by differentiating the distribution function $\exp\{-V(\boldsymbol{z})\}$ with respect to each $z_1,\ldots,z_D$. The number of terms in the likelihood is the Dth 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 **evd** and **copula** packages provide functionalities that are restricted to the bivariate setting, while **ExtremalDep** (Beranger et al. 2023) includes composite likelihood estimation via it's function `fExtDep` for a variety of models. The **SpatialExtremes** and **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 $\mathrm{O}(D)$ 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 $\mathrm{O}(D^2)$ 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 $\mathrm{O}(2^D)$ 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 approach, whereby data are first transformed to standardized margins and then dependence parameters are fitted separately. The function `fbvevd` in **evd** allows the user to pass fixed values for some parameters. The **tailDepFun** package contains routines for fitting the continuous updating weighted least squares estimator, along with goodness-of-fit tests, for multivariate and functional models including max-linear models (Einmahl et al. 2018).

A different avenue is to estimate an equivalent form of H in eq. (3.1) termed the Pickands dependence function (cf., Falk et al. 2011, p. 150). The latter has properties, notably convexity and known values on the corners of the simplex, that can be enforced to improve estimation. The **evd** package allows users to estimate nonparametrically the bivariate dependence function based on the estimators of Pickands (1981) and Caperaa et al. (1997) for block maxima; additional options correct for boundary and convexity constraints. The function `An` in **copula** provides generalization of estimators of Pickands dependence function to higher dimensions (Gudendorf and Segers 2012). Multivariate estimators based on Bernstein polynomials that guarantee convexity (Marcon et al. 2017b) are provided by the `beed` procedure in **ExtremalDep**, along with the `madogram` estimator. Bayesian estimation is also available in the bivariate case, imposing a prior on the order of the Bernstein polynomials. The package also includes a procedure for computing pointwise confidence intervals using a nonparametric bootstrap. The `plot_ExtDep.np` function with parameter `type="Qsets"` from **ExtremalDep** provides credible intervals for bivariate extreme quantile regions (Beranger et al. 2021a), estimated using an extension of this approach. Lastly, **fCopulae** (Wuertz et al. 2023) provides parametric dependence function, correlation coefficient and tail dependence measures for bivariate extreme value copulas.

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** (`evCopula` 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 (3.1) 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 Λ on a risk region $\mathcal{R} \subset \mathbb{R}_+^D \setminus \{\mathbf{0}_D\}$, i.e., the positive orthant excluding the origin (Resnick 1987). Assuming the intensity measure is absolutely continuous, the intensity function $\lambda(\boldsymbol{x}) = \partial^D \Lambda(\boldsymbol{x})/(\partial x_1 \cdots \partial x_D)$ exists and we can define a density over $\mathcal{R}$ by renormalizing $\lambda(\boldsymbol{x})$ 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 $\mathcal{R} = \{\boldsymbol{x} \in \mathbb{R}_+^D : x_i > u\}$; if $\xi = \mathbf{1}_D$, then $\mathcal{R} = \{\boldsymbol{x} \in \mathbb{R}_+^D : \|\boldsymbol{x}\|_1 > u\}$ has risk measure $\Lambda(\mathcal{R}) = Du^{-1}$ irrespective of the model for Λ. 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 $\mathcal{R} = \{\boldsymbol{x} \in \mathbb{R}^D_+ : \max_{j=1}^D x_j > u\}$. 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 strategies 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 χ. 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 `bvtcplot` 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 jth component, i.e., consider a limiting model for $\boldsymbol{Y}_{-j} \mid Y_j > u$. 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 $\boldsymbol{Y} \mapsto \boldsymbol{Y}^\star$, and then mapped to radius and pseudo-angles $(R, \boldsymbol{\Omega})$, with, e.g., $R = \|\boldsymbol{Y}^\star\|_1$ and $\boldsymbol{\Omega} = \boldsymbol{Y}^\star_{-D}/R$. Since R and $\boldsymbol{\Omega}$ become stochastically independent as R tends to infinity, one can focus on modelling the spectral measure $H(\boldsymbol{\omega})$ appearing in eq. (3.1). **ExtremalDep**, through `fExtDep`, 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 quantiles 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 $\chi = \lim_{v\to 1} \chi(v)$ (Coles et al. 1999), where

$$\chi(v) = \frac{\Pr[\min_i\{F_i(Y_i) > v\}]}{1-v}. \tag{3.2}$$

The latter is used to assess whether extremes are asymptotically independent ($\chi = 0$) or dependent ($\chi > 0$). Equation (3.2) suggests replacing the unknown distribution functions by their empirical counterpart 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 η, which can be used to characterize the speed of decay for asymptotically independent variables. With random vectors transformed to unit Pareto margins, say $\boldsymbol{Y}^{\mathrm{p}}$, the structural variable $T = \min_{j=1}^{D} Y_j^{\mathrm{p}}$ is such that, for large u (Ledford and Tawn 1996, eq. 5.6),

$$\Pr\left(T > u + t \mid T > u\right) \approx \frac{L(u+t)}{L(u)}(1+t/u)^{-1/\eta}, \tag{3.3}$$

with $L(x)$ a slowly varying function. The coefficient of tail dependence can be estimated by fitting a generalized Pareto distribution with shape η and scale ηu to exceedances of T above u. If data are transformed to the exponential scale instead, the scale parameter of the structural variable is η 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 multivariate setting, the coefficients η_C for subsets $C \subset \{1, \ldots, D\}$ satisfy ordering constraints (de Haan and Zhou 2011, § 4.2).

In the bivariate setting, it is customary to consider $\bar{\chi} = 2\eta - 1$ instead of η, which gives $\bar{\chi} \in (-1, 1]$ (Coles et al. 1999). The **evd** package function `chiplot` provides plots of χ and $\bar{\chi}$ based on the empirical distribution of the minimum, with approximate pointwise standard errors through the delta-method. The **mev** package provides various estimators of η and χ, while **graphicalExtremes** includes empirical estimators `emp_chi` that can be used to obtain empirical estimates of the dependence matrix 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 $Z_\omega = \min\{Y_1^{\mathrm{P}}, Y_2^{\mathrm{P}}\omega/(1-\omega)\}$ for $\omega \in (0,1)$ a fixed angle. Under a regular variation assumption, the distribution of Z_ω 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 special case. The **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 considered in Wadsworth and Tawn (2013) and implemented in `lambdadep` function of the **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 ω, 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 extremes 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 extreme 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 3.1 gives basic information about the direct estimators of the extremal index that feature in this section, while Table 3.2 summarises implementations 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 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** (`declust`) use an estimate of

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 3.1 Overview of some direct estimators of the extremal index with associated references and 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	`kgaps`	yes	`choose_uk`
	SPM	`spm`	yes	`choose_b`
fExtremes	runs	`runTheta`	no	`exindexPlot`
	blocks 1	`clusterTheta`	no	`exindexesPlot`
	blocks 2	`blocktheta`	no	
	intervals	`ferrosegersTheta`	no	
mev	ILS, FS	`ext.index`	no	`ext.index`
	K-gaps	`ext.index`	no	`infomat.test`, `ext.index`
POT	runs	`fitexi`	no	`exiplot`
revdbayes	K-gaps	`kgaps_post`	yes	—
texmex	FS	`extremalIndex`	yes	`extremalIndexRangeFit`
tsxtreme	runs	`thetaruns`	yes	—

Table 3.2 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.

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 3.3 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 `evmc` 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 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

reference	package	function(s)	area
Fawcett and Walshaw (2012)	**texmex**	`declust`, `evm`	**m**
Fawcett and Walshaw (2012)	**lite**	`flite`	**m**
Durrieu et al. (2019)	**extremefit**	`hill.ts`	**m**
Davis and Mikosch (2009)	**extremogram**	`extremogram1`, `bootconf1`, ...	**e**
Lugrin et al. (2016)	**tsxtreme**	`depfit`, `dep2fit`	**d**
Smith et al. (1997)	**evd**	`evmc`	**d**
Smith et al. (1997)	**POT**	`fitmcgpd, simmc`	**d**
Noven et al. (2018)	**ev.trawl**	`FullPL, rtrawl`	**d**
Hees et al. (2021)	**CTRE**	`MLestimates`	**d**

Table 3.3 Overview of packages and main functions for modelling time series extremes by area: marginal modelling (**m**); exploratory analysis (**e**); dependence modelling (**d**).

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 θ_2, tail correlation function based on the χ 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 $Y(\boldsymbol{s})$ for stochastic processes indexed by $\boldsymbol{s} \in \mathcal{S} \subset \mathbb{R}^d$, representing the process of the original event data. Usually we have a random sample of observations $Y_i(\boldsymbol{s}_j)$ for $j = 1, \ldots, D$ locations observed at $i = 1, \ldots, n$ time points and denote a realization in space by $\boldsymbol{Y} = \{Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_D)\}$, 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 $a_n(\boldsymbol{s}) > 0$ and $b_n(\boldsymbol{s})$ such that the law of the scaled maximum converges for all finite-dimensional distributions,

$$\lim_{n\to\infty} \Pr\left[a_n(\boldsymbol{s})\{M_n(\boldsymbol{s}) - b_n(\boldsymbol{s})\} \leq x\right] = \Pr\{Z(\boldsymbol{s}) \leq x\}, \quad s \in \mathcal{S}, \tag{4.1}$$

with $Z(\boldsymbol{s})$ a nondegenerate limit process, then $Z(\boldsymbol{s})$ 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 $Y^\star(\boldsymbol{s})$, 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 $\lim_{x\to\infty} x \Pr[g\{Y(\boldsymbol{s})\} > x] = 1$. For example, we can choose $Y^\star(\boldsymbol{s}) = g_{\boldsymbol{s}}\{Y(\boldsymbol{s})\} = 1/[1 - F_{\boldsymbol{s}}\{Y(\boldsymbol{s})\}]$, where $F_{\boldsymbol{s}}$ is the marginal distribution of $Y(\boldsymbol{s})$. Risk-Pareto processes (Ferreira and de Haan 2014; Thibaud and Opitz 2015; Dombry and Ribatet 2015; de Fondeville and Davison 2018; Engelke et al. 2019) arise asymptotically when a functional r of the standardized process $Y^\star(\boldsymbol{s})$ exceeds a threshold that tends towards the upper endpoint of the probability distribution of r.

Typically, summary functionals are homogeneous, meaning $r(t\boldsymbol{x}) = tr(\boldsymbol{x})$ for $t > 0$; examples include the average $r(\boldsymbol{x}) = |\mathcal{S}|^{-1} \int_{\mathcal{S}} x(\boldsymbol{s})\,\mathrm{d}s$, the minimum $r(\boldsymbol{x}) = \min_{\boldsymbol{s}\in\mathcal{S}} x(\boldsymbol{s})$, the maximum $r(\boldsymbol{x}) = \max_{\boldsymbol{s}\in\mathcal{S}} x(\boldsymbol{s})$, or the median. Convergence is assumed in the space of continuous functions over compact $\mathcal{S}$, such that the distribution of the functional $r[g_{\boldsymbol{s}}\{Y(\boldsymbol{s})\}]$ is well defined. Functional convergence of maxima in (4.1) implies functional convergence to r-Pareto processes $Z_r(\boldsymbol{s})$:

$$\lim_{u\to\infty} \Pr\left[u^{-1}Y^\star(\boldsymbol{s}) \leq x \mid r\{Y^\star(\boldsymbol{s})\} \geq u\right] = \Pr\{Z_r(\boldsymbol{s}) \leq x\}, \quad s \in \mathcal{S}. \tag{4.2}$$

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 $Y(\boldsymbol{s}_1)$ and $Y(\boldsymbol{s}_2)$ at locations $\boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathcal{S}$. 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 (3.3), if considered for D sites $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_d$, 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 $1/D < \eta(\boldsymbol{s}_1, \boldsymbol{s}_2) < 1$ is possible; for example, the class of max-infinitely divisible processes (Huser et al. 2021), which is useful for flexible modelling of locationwise 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. (3.1), 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 $F_{\boldsymbol{s}}$ 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., $\xi(\boldsymbol{s}), \mu(\boldsymbol{s}), \sigma(\boldsymbol{s})$. Different options exist in the peaks over threshold setting. A common approach is to fix a high, potentially nonstationary threshold $u(\boldsymbol{s})$, and then estimate the threshold exceedance probability $p(\boldsymbol{s}) = \Pr\{Y(\boldsymbol{s}) > u(\boldsymbol{s})\} = 1 - F_{\boldsymbol{s}}\{u(\boldsymbol{s})\}$ and the generalized Pareto parameters $\xi(\boldsymbol{s}), \sigma(\boldsymbol{s})$ based on observations of the exceedances $Y(\boldsymbol{s}) - u(\boldsymbol{s}) > 0$.

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 approximations 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 (2.5) 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 unpublished package **spatialADAI**. The Matlab package **SpatialConditionalExtremesSatellite** fits univariate and multivariate spatial conditional extremes models (Shooter et al. 2021, 2022). An INLA-based implementation for Bayesian conditional extremes models for spatial and spatio-temporal data is provided as supplementary material of Simpson et al. (2023). The implementation of many Bayesian extreme value models in the literature is achieved with standard MCMC algorithms that are tailored to the particular data application, but often generic and easily reusable or reproducible code is not provided, which hinders reproducibility.

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

methods	package	functions	scope
	copula	rCopula*	**b**, **m**
Tajvidi and Turlach (2018)	**SimCop**	—	**b**
Stephenson (2003)	**evd**	rbvevd, rmvevd	**b**, **m**
Dombry et al. (2016)	**mev**	rmev, rmevspec	**a**, **m**, **f**
Engelke and Hitz (2020)	**graphicalExtremes**	rmstable	**m**, **f**
Beranger et al. (2017)	**ExtremalDep**	rExtDep, rExtDepSpat	**m**, **f**
Padoan and Bevilacqua (2015)	**CompRandFld**	RFsim*	**f**
Dombry et al. (2013)	**SpatialExtremes**	condrmaxstab	**f**
Dombry et al. (2016)	**SpatialExtremes**	rmaxstab	**f**
Schlather et al. (2015)	**RandomFields**	RFsimulate*	**f**
Reich and Shaby (2012)	**hkevp**	hkevp.rand	**f**
Ballani and Schlather (2011)	**BMAmevt**	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 4.1 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.

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
Gudendorf and Segers (2012)	**copula**	An	**b**, **m**	ang
Einmahl and Segers (2009) and	**extremis**	angcdf	**b**	ang
de Carvalho et al. (2013)	**mev**	angmeas	**m**	ang
Marcon et al. (2017b)	**ExtremalDep**	madogram, beed	**m**	ang
Beranger et al. (2021a)	**ExtremalDep**	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**	fbvpot	**b**	pot
Engelke et al. (2019)	**graphicalExtremes**	fmpareto_graph_HR	**m**	pot
Heffernan and Tawn (2004)	**texmex**	mex	**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)	**ExtremalDep**	fExtDep, fExtDepSpat	**m**, s	max

Table 4.2 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.

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. 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 **evtclass** 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, DGECR-2022-00461). The authors declare that they have no conflict of interest. They thank three anonymous referees, the associate editor and B. Béranger 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 24(3):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

Ballani F, Schlather M (2011) A construction principle for multivariate extreme value distributions. Biometrika 98(3):633–645, doi:10.1093/biomet/asr034

Barlow AM, Mackay E, Eastoe E, Jonathan P (2023) A penalised piecewise-linear model for non-stationary extreme value analysis of peaks over threshold. Ocean Engineering 267:113265, doi:10.1016/j.oceaneng.2022.113265

Barnett V (1976) The ordering of multivariate data (with discussion). Journal of the Royal Statistical Society Series A (General) 139(3):318–355, doi:10.2307/2344839

Beirlant J, Vynckier P, Teugels JL, et al. (1996) Excess functions and estimation of the extreme-value index. Bernoulli 2(4):293–318, doi:10.2307/3318416

Beirlant J, Dierckx G, Guillou A (2011) Bias-reduced estimators for bivariate tail modelling. Insurance: Mathematics and Economics 49(1):18–26, doi:10.1016/j.insmatheco.2011.01.010

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

Beranger B, Padoan S, Marcon G (2023) `ExtremalDep`: Extremal Dependence Models. R package version 0.0.4-0

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

Bücher A, 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

Caeiro F, Gomes MI (2014) On the bootstrap methodology for the estimation of the tail sample fraction. In: Proceedings of COMPSTAT 2014, pp 545–552

Caeiro F, Gomes MI (2016) Threshold selection in extreme value analysis. In: Dey DK, Yan J (eds) 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 hydroclimatology. Hydrological Processes 24(6):673–685, doi:10.1002/hyp.7506

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, Warchoł 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, Moriña 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:211007051`

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) extremis: 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

do Nascimento FF, Moura e Silva WV (2016) MCMC4Extremes: Posterior distribution of extreme value models in R. R package version 1.1

Dombry 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

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

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

Dombry C, Ribatet M, Stoev S (2018) Probabilities of concurrent extremes. Journal of the American Statistical Association 113(524):1565–1582, doi:10.1080/01621459.2017.1356318

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 (2020) 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 JHJ, 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 JHJ, 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):871–932, 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:doi.org/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`
Grama I, Spokoiny V, et al. (2008) Statistics of extremes by oracle estimation. The Annals of Statistics 36(4):1619–1648, doi:10.1214/07-AOS535
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 (1990) Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. Journal of Multivariate Analysis 32(2):177–203, doi:10.1016/0047-259X(90)90080-2
Hall P, Welsh AH (1985) Adaptive estimates of parameters of regular variation. 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

Heffernan JE, Stephenson AG (2018) `ismev`: An introduction to statistical modeling of extreme values. R package version 1.42

Heffernan JE, Tawn JA (2004) A conditional approach for multivariate extreme values (with discussion). Journal of the Royal Statistical Society: Series B (Statistical Methodology) 66(3):497–546, doi:10.1111/j.1467-9868.2004.02050.x

Hill B (1975) A simple general approach to inference about the tail of a distribution. Annals of Statistics 3:1163–1174, doi:10.1214/aos/1176343247

Ho ZWO, Dombry C (2019) Simple models for multivariate regular variation and the Hüsler–Reiss Pareto distribution. Journal of Multivariate Analysis 173:525–550, doi:10.1016/j.jmva.2019.04.008

Hofert M, Hornik K, McNeil AJ (2022) `qrmtools`: Tools for Quantitative Risk Management. R package version 0.0-16

Hosking JRM (1990) l-moments: Analysis and estimation of distributions using linear combinations of order statistics. Journal of the Royal Statistical Society Series B (Methodological) 52(1):105–124, doi:10.1111/j.2517-6161.1990.tb01775.x

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

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

Hosking JRM, Wallis JR (1987) Parameter and quantile estimation for the generalized Pareto distribution. Technometrics 29(3):339–349, doi:10.1080/00401706.1987.10488243

Hosking JRM, Wallis JR (1997) Regional Frequency Analysis: An Approach Based on L-Moments. Cambridge University Press, doi:10.1017/CBO9780511529443

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

Huser 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

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, Dombry 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-infinitely 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) `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 modeling 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) evt0: 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

Moins 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, Kossinna P, Jayasinghe D, Wijeratne D (2019) ptsuite: 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 ℓ_p-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) `tea`: 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, Bevilacqua M (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 JY (ed) Extreme Value Modeling and Risk Analysis: Methods and Applications, CRC Press, Boca Raton, FL, pp 131–151

Resnick S, Stărică C (1997) Smoothing the Hill estimator. Advances in Applied Probability pp 271–293, doi:10.2307/1427870

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

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

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

Rootzén 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, Bitner-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 LP, Scheirer WJ, Boult 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, Fougères 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

Sebille Q (2016) **hkevp**: 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, Sørbye SH (2017) Penalising model component complexity: A principled, practical approach to constructing priors. Statistical Science 32(1):1–28, doi:10.1214/16-STS576

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/asv045

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, Scheibel 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, Dombry C, Cai JJ, Engelke S (2021) Gradient boosting for extreme quantile regression. `arXiv:210300808`

Vignotto E, Engelke S (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

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

Wuertz 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