

Geostatistical Analysis in Extremes: An Overview

M. Manuela Neves

Abstract Classical statistics of extremes is very well developed in the univariate context for modeling and estimating parameters of rare events. Whenever rain, snow, storms, hurricanes, earthquakes, and so on, happen the analysis of extremes is of primordial importance. However such rare events often present a temporal aspect, a spatial aspect or both. Classical geostatistics, widely used for spatial data, is mostly based on multivariate normal distribution, inappropriate for modeling tail behavior. The analysis of spatial extreme data, an active research area, lies at the intersection of two statistical domains: extreme value theory and geostatistics. Some statistical tools are already available for the spatial modeling of extremes, including Bayesian hierarchical models, copulas and max-stable random fields. The purpose of this chapter is to present an overview of basic spatial analysis of extremes, in particular reviewing max-stable processes. A real case study of annual maxima of daily rainfall measurements in the North of Portugal is slightly discussed as well the main functions in R environment for doing such analysis.

1 Motivation and Introduction

This chapter is related to a talk presented in a session of the “International Conference Planet Earth. Mathematics of Energy and Climate Change” entitled *The role of Statistics of Extremes in Society*. Through the ages natural disasters and catastrophes have happened, causing deaths and destruction. Remember, for example, the 1755 Lisbon earthquake and tsunami and the 2005 hurricane Katrina, New Orleans, see Fig. 1.

The scientific community has been worried about whether anything could be done for society to be at least better prepared for those occurrences.

On 18.01.12, Nicolas Guerin [22], from École Polytechnique Fédérale de Lausanne, wrote in *News Mediacom*:

... “*The problem of extremes is that there are so few events, by definition...explains EPFL mathematician Anthony Davison. It’s thus necessary to*

M.M. Neves (✉)
CEAUL, and ISA, University of Lisbon, Lisbon, Portugal
e-mail: manela@isa.ulisboa.pt

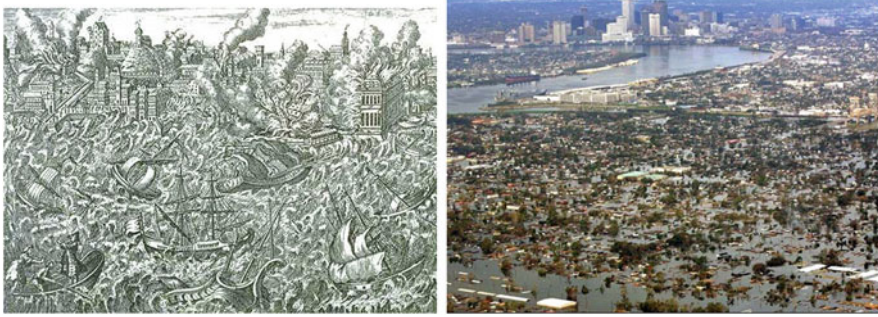


Fig. 1 The 1755 Lisbon earthquake and tsunami (*left*) and the 2005 hurricane Katrina, New Orleans (*right*)

create specific models that are different from those that use innumerable mean values...

For several years now, the scientists have noted that the increase in extreme events associated with climate change appears to be having much more of an impact on society than the increase in mean temperatures. Natural disasters are accompanied by a significant human and economic cost...

Extreme value theory (EVT) is the branch of probability and statistics dedicated to characterizing the behavior of the extreme observations. An extreme observation is a datum that has low probability of occurrence, but which can be very large (or small).

EVT has its beginnings in the early to middle part of the last century. It formally began with the paper by Dodd [14], followed by papers of Fréchet [18], Fisher and Tippett [17] Gumbel [23] and von Mises [42], to cite the pioneering and most relevant works.

Rare events such as the risk of flooding, potential crop damage from drought, health effects of extreme air pollution, storms, and so on, may cause severe impacts on human life as well as on ecosystems. In weather and climate studies as well as in other fields, often what one wants to characterize is not the usual behavior, but the extreme events.

Assessing the behavior of rare events such as wind speed, precipitation and temperature presents unique statistical challenges, and requires one to characterize the tail of the distribution of the quantity of interest.

Emil Gumbel (1891–1966) was the pioneer in the application of statistics of extremes. He wrote “...*The aim of a statistical theory of extreme values is to analyze observed extremes and to forecast further extremes*”. Gumbel [24] presents several applications of EVT on real world problems in engineering and in meteorological phenomena. There appear first applications in hydrology. “...*It seems that the rivers know the theory...*” is a remarkable expression from Gumbel.

Recently, a special issue of the journal *Extremes* (2010) **13:2** on Statistics of Extremes in Weather and Climate, shows the relevance of EVT in the area.

Nowadays, extreme value analysis appears in quite diversified areas revealing the importance of statistics of extremes in applications. Many excellent books, presenting both the methodological basis and a great emphasis in the applications must be referred to. Besides Gumbel [24], we can mention Tiago de Oliveira (ed.) [12] that is still a reference today with a wide range of contributions and applications of statistics of extremes. It was the result of a remarkable meeting that took place thirty years ago in Vimeiro and fortunately was remembered in 2013, celebrating that conference and also dedicated to Ivette Gomes, a highly recognized international researcher in Statistics of Extremes. More recently, other books emphasizing the applications appeared and deserve to be mentioned: Coles [4], Finkenstadt and Rootzén (eds.) [16], Castillo et al. [3], Beirlant et al. [1], Reiss and Thomas, [33] and Gomes et al. [21].

Statistical modeling of extremes was based initially on limiting families of distributions for maxima of a sequence, X_1, \dots, X_n , of independent and identically distributed (i.i.d.) random variables from an unknown distribution function (d.f.), F . Given that the distribution of the maxima is highly dependent of the unknown form of F , similar to the central limit theory, researchers tried to obtain sequences $\{a_n > 0\}$ and $\{b_n\} \in \mathbf{R}$ such that $M_n := \max\{X_1, \dots, X_n\}$, linearly normalized by those constants, had a non-degenerate limiting distribution.

Univariate EVT is well developed, but is it well recognized that many extreme events, particularly in the environment, environmental health, climate, hydrology or meteorology occur in a place and/or in a time. In those areas of application we are faced with the task of analyzing data that are geographically referenced and show a correlated structure that needs to be adequately modelled.

Spatial data are measurements or observations taken at specific locations or within specific regions. The dependence structure of those data needs to be adequately captured. Geostatistics is an active area of research with important applications in the environment, agriculture and public health. Classical methods are well known and explored under the Gaussian model.

Regarding extreme values there was thus a need to develop methods for analyzing and characterizing spatial extreme data. Spatial extreme theory is an area that lies at the intersection of EVT and geostatistics. While classical geostatistics is usually applied to situations where only one realization of the process is taken, the spatial extreme approach needs multiple realizations underlying the subset of extreme data analyzed.

After this general motivation and introduction, Sect. 2 introduces the basic notions in EVT and in Standard Geostatistics. In Sect. 3 a background in spatial extremes is given as well as some most common statistical models in the max-stable process approach for spatial extremes. Section 4 is devoted to slightly discuss an application to annual maxima of daily rainfall data, with particular emphasis to

the use of R software. Applications to rainfall data have been done recently by Smith and Stephenson [39], Padoan et al. [31] and Davison et al. [8].

2 Fundamental Notions and Basic Results

Classical EVT has been one of the most fast developing areas in the last decades. The underlying mathematical basis is well established, see e.g. Leadbetter et al. [26], Embrechts et al. [15] and de Haan and Ferreira [11], to cite only a few books. To model the tail of a distribution, where extreme events often occur, has been a challenge for researchers. There is now a well established set of methods to model the tail of a distribution.

2.1 Main Limiting Results in EVT

For the univariate EVT the main and well known result is: Let X_1, \dots, X_n be independent replications from an unknown d.f., F , and define $M_n := \max\{X_1, \dots, X_n\}$. Fréchet [18], Fisher and Tippet [17], Gumbel [23], and von Mises [42], in Fig. 2, obtained the first results concerning the existence of a non-degenerate law of the maximum of that series, suitably normalized.

Gnedenko [20] and later on de Haan [9] gave necessary and sufficient conditions for the existence of sequences $\{a_n > 0\}$ and $\{b_n\} \in \mathbf{R}$ such that,

$$\lim_{n \rightarrow \infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = \lim_{n \rightarrow \infty} F^n(a_n x + b_n) = EV_\xi(x),$$

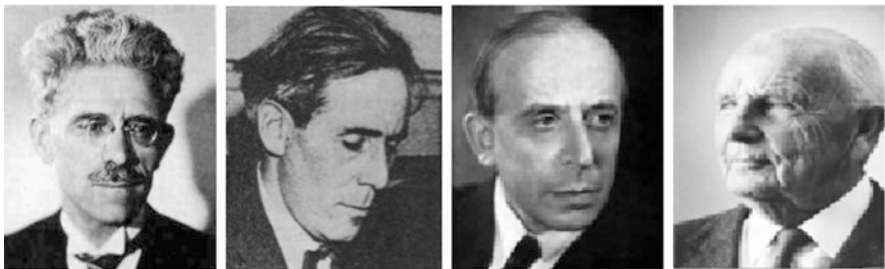


Fig. 2 Fréchet (1878–1973), Gumbel (1891–1966), von Mises (1883–1953) and Weibull (1887–1979) (from left to right)

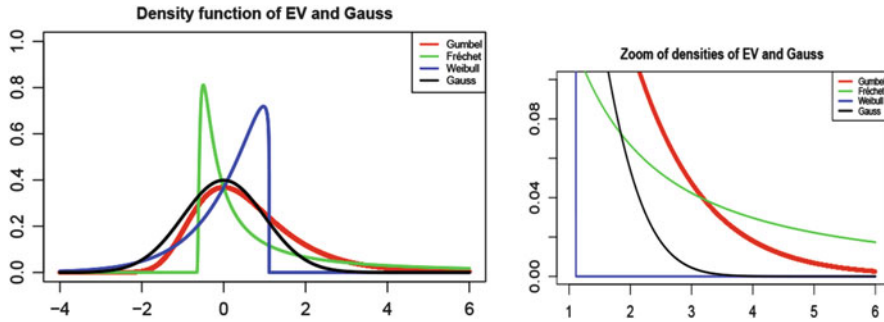


Fig. 3 Gumbel, Fréchet and Weibull p.d.f. (left) and zoom of Gumbel, Fréchet, Weibull and Gauss p.d.f. (right)

$\forall x \in \mathbf{R}$, where EV_ξ is a nondegenerate distribution function. This function, called Extreme Value d.f., is given by

$$EV_\xi(x) = \begin{cases} \exp[-(1 + \xi x)^{-1/\xi}], & 1 + \xi x > 0 & \text{if } \xi \neq 0 \\ \exp[-\exp(-x)], & x \in \mathbf{R} & \text{if } \xi = 0. \end{cases} \quad (1)$$

In applications, the d.f. in (1) can also present a more general form with a location parameter $\lambda \in \mathbf{R}$ and a scale parameter $\delta \in \mathbf{R}^+$.

The EV_ξ incorporates the three (Fisher-Tippett) types: Gumbel type: $\Lambda(x) = \exp(-\exp(-x)) \equiv EV_0(x)$, $x \in \mathbf{R}$, ($\xi = 0$), the limit for exponential tailed distributions; Fréchet type: $\Phi_\xi(x) = \exp(-x)^{-1/\xi} \equiv EV_\xi(\frac{x-1}{\xi})$, $x > 0$, $\xi > 0$, the limit for heavy tailed distribution and Weibull type: $\Psi_\xi(x) = \exp(-x)^{1/\xi} \equiv EV_\xi(\frac{-x-1}{\xi})$, $x < 0$, $\xi < 0$, the limit for short tailed distributions. These three families of models were combined as the d.f. EV_ξ in (1) by von Mises [42].

If $\xi = 0$, the *right endpoint*, $x^* := \sup\{x : F(x) < 1\}$, can then be either finite or infinite. If $\xi > 0$, F has an infinite *right endpoint*. If $\xi < 0$, F has a finite *right endpoint*, x^* . The shape parameter, ξ , is then directly related to the weight of the right tail, $\bar{F} := 1 - F$, of the underlying model F . As ξ increases the right tail becomes heavier. Figure 3 shows the behavior of the right-tails for the three different types of EV models, together with the Gauss model for comparison.

Jointly with the knowledge of methodological aspects of extreme values theory, the interest for having free, accurate and simple software has increased tremendously, motivated by the wide range of areas of application of EVT.

R is an environment and a programming language for statistical computing and graphics. It is a free and open source project. Several packages for extreme value analysis are already available, with a large set of functions, among which we mention: `evd`, Stephenson [40], with functions for statistical analysis of extremes, including multivariate extremes and Bayesian methods; `ismev`, Stephenson [41], with functions for classical extreme value analysis, fitting extreme value distribution to “block maxima”, as well as generalized Pareto distribution to excesses over a

high threshold and `extRemes`, with a graphical user interface based on `ismev`, are perhaps the most well known. Other packages are nevertheless available, see Gilleland, et al. [19]. They published a very nice software review, comparing the available statistical software and presenting the main characteristics of the main packages for extreme value analysis.

2.2 Standard Geostatistics

If the quantity of interest, e.g. the rainfall level, is observed at different locations spread over a region, it is necessary to know how to take into account the spatial pairwise dependence among sites, for an adequate analysis of data. There are three types of spatial data:

- Geostatistical data or point referenced data are measurements taken at fixed locations $\mathbf{s} \in \mathbf{K} \subset \mathbf{R}^d$. These locations are generally spatial continuous. Data may be modeled as values from a spatial process.
- Lattice data or areal referenced data where \mathbf{K} is again a fixed subset but observations are associated with spatial regions and with well defined boundaries.
- Spatial points patterns where locations are now the variable of interest, so \mathbf{K} is itself random and observation sites may be treated as random.

We shall consider geostatistical data modeled by a stochastic process $\{\mathbf{Y}(\mathbf{s})\}$ where $\mathbf{s} \in \mathbf{K}$ and \mathbf{K} is a compact subset in \mathbf{R}^d . Data are observed at $D = \{\mathbf{s}_1, \dots, \mathbf{s}_D\} \subset \mathbf{K} \subset \mathbf{R}^d$. Usually $d = 2$ and we shall assume it throughout. Essential elements for exploring and modeling spatial data are: stationarity, isotropy and the variogram, key elements of the “Matheron school”, see Cressie [7].

Let us consider a spatial process with mean, $\mu(\mathbf{s}) = E[\mathbf{Y}(\mathbf{s})]$ and variance $Var[\mathbf{Y}(\mathbf{s})]$ finite for all $\mathbf{s} \in \mathbf{K} \subset \mathbf{R}^2$.

Usually it is assumed second-order stationarity what implies that covariance relationship between values of the process at any two locations can be summarized by the covariance function $C(\mathbf{h}) = Cov(Y(\mathbf{s}), Y(\mathbf{s} + \mathbf{h}))$, for all $\mathbf{h} \in \mathbf{R}^2$, such that \mathbf{s} and $\mathbf{s} + \mathbf{h}$ both lie within \mathbf{K} . That function depends only on the separation vector \mathbf{h} .

Assuming also that $E[\mathbf{Y}(\mathbf{s} + \mathbf{h}) - \mathbf{Y}(\mathbf{s})] = 0$, with \mathbf{s} and $\mathbf{s} + \mathbf{h} \in \mathbf{K}$, the variogram is defined as:

$$E[\mathbf{Y}(\mathbf{s} + \mathbf{h}) - \mathbf{Y}(\mathbf{s})]^2 = Var[\mathbf{Y}(\mathbf{s} + \mathbf{h}) - \mathbf{Y}(\mathbf{s})] = 2\gamma(\mathbf{h}),$$

where $\gamma(\mathbf{h})$ is designated as the semivariogram and is a crucial measure for quantifying spatial dependence in the data. If the semivariogram depends only on the length of \mathbf{h} , and not on the orientation, $\gamma(\mathbf{h}) = \gamma(\|\mathbf{h}\|)$, the process is said to be isotropic, i.e., roughly speaking “as it looked the same in all directions”. Table 1 summarizes the main models for isotropic semivariograms.

Table 1 Theoretical semivariograms

Model	Semivariogram ($\gamma(t)$)
Linear	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 t, & t > 0; \\ 0, & \text{otherwise.} \end{cases}$
Spherical	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2, & t \geq 1/\phi; \\ \tau^2 + \sigma^2 (\frac{3}{2}\phi t - \frac{1}{2}(\phi t)^3), & 0 < t \leq 1/\phi; \\ 0, & \text{otherwise.} \end{cases}$
Exponential	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 (1 - \exp(-\phi t)), & t > 0; \\ 0, & \text{otherwise.} \end{cases}$
Powered exponential	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 (1 - \exp(- \phi t ^p)), & t > 0; \\ 0, & \text{otherwise.} \end{cases}$
Matérn at $\nu = 3/2$	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 (1 - (1 + \phi t) \exp(-\phi t)), & t > 0; \\ 0, & \text{otherwise.} \end{cases}$

In an exploratory phase of a geostatistical analysis, the dependence is investigated via an empirical covariogram or empirical semivariogram.

Assuming stationarity and isotropy the simplest semivariogram estimator is the *moments estimator*, Matheron [27],

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{(s_i, s_j) \in N(h)} \{Y(s_i) - Y(s_j)\}^2,$$

where $N(h) = \{(s_i, s_j) : ||s_i - s_j|| = h\}$ and $|N(h)|$ denotes the cardinality of $N(h)$.

A few steps can be summarized for basic geostatistical analysis: remove trends in mean and (perhaps) in variance; transform residuals to standard normal margins; use graphical techniques to assess likely form for the semivariogram function; fit a suitable semivariogram model; make inferences using weighted least squares (kriging), likelihood or Bayes procedures; make predictions using the fitted correlation to obtain a map of predictions, based on a fitted normal model.

In R environment several packages are available for geostatistical analysis. Exploratory analysis, modeling and extrapolating are possible through several functions in `spatial`, `gstat`, `sp`, `MASS` and `geoR`, for example.

- `gstat` includes functions as: `variogram`—calculates sample (experimental) variograms; `plot.variogram`—plots an experimental variogram with automatic detection of lag spacing and maximum distance; `fit.variogram`—iteratively fits an experimental variogram; `krige`—a generic function to make predictions by inverse distance interpolation, ordinary kriging, OLS regression, regression-kriging and co-kriging; `krige.cv`—runs `krige` with cross-validation, see Pebesma [32] and Bivand et al. [2] for a complete overview of `gstat` functions and examples.

- `geoR`—extensively described by Diggle and Ribeiro Jr. [13] and Ribeiro Jr. et al. [35], with a series of tutorials.

Usually it is supposed that $\{\mathbf{Y}(\mathbf{s})\}$ follows a Gaussian process, so likelihood inference is realized under this assumption.

3 Geostatistical Analysis in Extremes

As mentioned in the previous section, Gaussian processes play a central role in modeling spatial processes. In this approach relevance is given to studying the central tendencies of the distribution rather than the distribution tails. However in many events already pointed out behind, the extremes are of main interest. The generalization of classical multivariate extreme value distributions to the spatial case is done through max-stable processes. Other statistical approaches have been developed for the spatial modeling of extremes, such as Bayesian hierarchical models and copulas, topics that will not be considered in this brief overview. A recent and very good survey on spatial extremes is Cooley et al. [6].

The purpose of this section is to review some max-stable processes, and mainly to show packages and functions already available in the R environment for the analysis of spatial extremes. First results date back to de Haan [10] and were developed by several authors, such as Smith [38], Schlather [36] and Kabluchko et al. [25], to mention only a few.

Max-stable processes follow a similar asymptotic motivation to the univariate EV distribution, providing a general approach to modeling process extremes incorporating temporal or spatial dependence. For the region \mathbf{K} under study and the location \mathbf{s} , we will assume $Y_1(\mathbf{s}), Y_2(\mathbf{s}), \dots$ as independent replicas of a stochastic process.

Consider that we have daily rainfall data collected in some stations and we are interested in considering the maximum of that quantity over a period of time (e.g. a year) in order to:

- Assess the dependence of the extreme precipitation levels between stations.
- Predict values at unobserved locations.
- Elaborate a map of the distribution of the maximum precipitation levels.

Interest is now to model extremes of a process $\{\mathbf{Y}(\mathbf{s})\}$ over spatial domain \mathbf{K} , where data are observed at sites $\mathbf{s}_d \in \{\mathbf{s}_1, \dots, \mathbf{s}_D\}$ within \mathbf{K} and at times $\mathbf{T} = \{t_1, \dots, t_n\}$. A difference between geostatistics and spatial extremes is that much of geostatistics is applied to situations where one has only one realization of the process $\{\mathbf{Y}(\mathbf{s})\}$. However to perform an extreme value analysis it is necessary that multiple realizations $\{Y_i(\mathbf{s})\}$ underlie the subset of extreme data which are analyzed.

An overview on some main max-stable models is here considered and briefly applied to a case study.

3.1 Some Models for Max-Stable Processes

Let us recall the definition of max-stable process by de Haan [10]. Let $\{Y_i(\mathbf{s})\}$, $\mathbf{s} \in \mathbf{D}$, $i = 1, \dots, n$, be independent replicas of a stochastic process $\{\mathbf{Y}(\mathbf{s})\}$ defined in $\mathbf{D} \subset \mathbf{R}^2$.

Definition 1 The stochastic process $\{\mathbf{Y}(\mathbf{s})\}$ is called max-stable if, for all $\mathbf{s} \in \mathbf{D}$ there exist normalizing sequences $\{a_n(\mathbf{s}) > 0\}$ and $\{b_n(\mathbf{s})\}$ such that, as $n \rightarrow \infty$,

$$\left\{ \frac{\max_{i=1, \dots, n} Y_i(\mathbf{s}) - b_n(\mathbf{s})}{a_n(\mathbf{s})} \right\} \rightarrow_d \{\mathbf{Y}^*(\mathbf{s})\}_{\mathbf{s} \in \mathbf{R}^2}$$

where $\{\mathbf{Y}^*(\mathbf{s})\}$ is identical in law to $\{\mathbf{Y}(\mathbf{s})\}$.

To characterize max-stable processes is a very difficult task. For max-stable processes with Fréchet unit margins de Haan [10] introduced a very useful representation that allowed the construction of parametric models for spatial extremes. Such a general representation is as follows: let $\{Y_i(\mathbf{s})\}_{i \in \mathbf{N}}$, be independent realizations of a stochastic process $\{\mathbf{Y}(\mathbf{s})\}$ with $E[\mathbf{Y}(\mathbf{s})] = 1$ and let ζ_i be points of a Poisson process Π with intensity $d\zeta/\zeta^2$ on $(0, \infty)$. Then

$$Z(\mathbf{s}) = \max_{i \geq 1} \zeta_i Y_i(\mathbf{s}), \quad \mathbf{s} \in \mathbf{D}, \tag{2}$$

is a max-stable process with unit-Fréchet margins. and the distribution function is determined by

$$P(Z(\mathbf{s}) \leq z(\mathbf{s}), \mathbf{s} \in \mathbf{D}) = \exp \left(- E \left[\sup_{\mathbf{s} \in \mathbf{D}} \left\{ \frac{Y(\mathbf{s})}{z(\mathbf{s})} \right\} \right] \right).$$

Different choices of the process $Y_i(\mathbf{s})$ lead to different models of max-stable processes. As examples let us mention the following models.

- Smith [38] proposed to take $Y_i(\mathbf{s}) = \varphi(\mathbf{s} - \mathbf{s}_i)$, where φ is a zero mean multivariate normal density with covariance matrix Σ , where the joint distribution at two sites is given by

$$P[Z(\mathbf{s}_1) \leq z_1, Z(\mathbf{s}_2) \leq z_2] = \exp \left[-\frac{1}{z_1} \Phi \left(\frac{a}{2} + \frac{1}{a} \log \frac{z_2}{z_1} \right) - \frac{1}{z_2} \Phi \left(\frac{a}{2} + \frac{1}{a} \log \frac{z_1}{z_2} \right) \right], \tag{3}$$

where $a = \sqrt{(\mathbf{s}_1 - \mathbf{s}_2)^T \Sigma^{-1} (\mathbf{s}_1 - \mathbf{s}_2)}$, is a dependence parameter, and Φ is the standard normal d.f.

- Schlather [36] proposed a more flexible class of max-stable processes by taking $Y_i(\mathbf{s})$ to be any stationary Gaussian random field with finite expectation. He considered

$$Z(\mathbf{s}) = \max_{i \geq 1} \zeta_i \max\{0, Y_i(\mathbf{s})\}$$

where $\mu = E[\max\{0, Y_i(\mathbf{s})\}] < \infty$, being the joint distribution at two sites given by

$$P[Z(\mathbf{s}_1) \leq z_1, Z(\mathbf{s}_2) \leq z_2] = \exp \left[-\frac{1}{2} \left(\frac{1}{z_1} + \frac{1}{z_2} \right) \left(1 + \sqrt{1 - \frac{2(\rho(h) + 1)z_1z_2}{(z_1 + z_2)^2}} \right) \right], \quad (4)$$

where $h = \|\mathbf{s}_1 - \mathbf{s}_2\|$ and $\rho(h)$ is chosen from Whittle-Matérn, Cauchy and Powered Exponential, see Schlather [36].

- As an example of another model let us consider a more recent proposal, the so-called Brown-Resnick process, studied in Kabluchko et. al. [25], who proposed an alternative specification for the $Y_i(\cdot)$ process, $Y(\mathbf{s}) = \exp(\epsilon_i(\mathbf{s}) - \sigma^2(\mathbf{s})/2)$, where $\epsilon(\mathbf{s})$ is a Gaussian process with stationary increments, being $\sigma^2(\mathbf{s})$ the variance of $\epsilon(\mathbf{s})$.

3.2 Spatial Dependence of Extremes

To use max-stable models we need to have information on how the dependence between two locations decreases, when the distance increases. It would be nice to have a *kind of variogram* for extremes of a stochastic process. However if we assume that Z is a unit Fréchet max-stable process, the variance (and even the mean) might be infinite.

A new function is now needed to reflect how evolves the spatial dependence of extremes. It provides sufficient information about extremal dependence and it is called *the extremal coefficient function*, Schlather and Tawn [37].

Definition 2 If $Z(\cdot)$ is a max-stable process with unit Fréchet margins the extremal coefficient function $\theta()$ is defined by

$$P[Z(\mathbf{s}_1) \leq z, Z(\mathbf{s}_2) \leq z] = \exp \left(-\frac{\theta(\|\mathbf{s}_1 - \mathbf{s}_2\|)}{z} \right),$$

where $1 \leq \theta(\|\mathbf{s}_1 - \mathbf{s}_2\|) \leq 2$.

The extremal coefficient function has the following meaning: If $\theta(\|\mathbf{s}_1 - \mathbf{s}_2\|) = 1$, we have perfect dependence; if $\theta(\|\mathbf{s}_1 - \mathbf{s}_2\|) = 2$, we have independence.

Extremal coefficient functions for the above models are:

- $\theta(\|\mathbf{s}_1 - \mathbf{s}_2\|) = 2\Phi\left(\frac{\sqrt{(\mathbf{s}_1 - \mathbf{s}_2)^T \Sigma^{-1}(\mathbf{s}_1 - \mathbf{s}_2)}}{2}\right)$, for the Smith model covering the whole range of dependence;
- $\theta(\|\mathbf{s}_1 - \mathbf{s}_2\|) = 1 + \sqrt{\frac{1 - \rho(\|\mathbf{s}_1 - \mathbf{s}_2\|)}{2}}$, for Schlather model that has upper bound of $1 + \sqrt{1/2}$. A drawback of this model is that independence of extremes can not be attained because $\theta \in [1; 1.8333]$ and $\theta = 2$ is never attained;
- $\theta(\|\mathbf{s}_1 - \mathbf{s}_2\|) = 2\Phi\left(\sqrt{\gamma(\|\mathbf{s}_1 - \mathbf{s}_2\|)/2}\right)$, for the Brown-Resnick process. As $\gamma(\|\mathbf{s}_1 - \mathbf{s}_2\|) \rightarrow 0$, we have $\theta(\|\mathbf{s}_1 - \mathbf{s}_2\|) \rightarrow 1$, while if $\gamma(\|\mathbf{s}_1 - \mathbf{s}_2\|)$ is unbounded, then $\theta(\|\mathbf{s}_1 - \mathbf{s}_2\|) \rightarrow 2$ as $\|\mathbf{s}_1 - \mathbf{s}_2\| \rightarrow \infty$.

Another measure of dependence between two locations is given by a “kind” of variogram. The first idea was to consider the madogram, a tool in classical geostatistics, Matheron [28], defined as:

$$v(\|\mathbf{s}_1 - \mathbf{s}_2\|) = E [|Z(\mathbf{s}_1) - Z(\mathbf{s}_2)|].$$

The madogram requires the finiteness of the first-moment and for stationary max-stable processes with unit Fréchet margins, mean and variance may be not finite and that mean value does not exist theoretically. Consequently, variogram-based approaches, specially designed for extremes, have been proposed:

- the F-Madogram, Cooley et al. [5],

$$v_F(\|\mathbf{s}_1 - \mathbf{s}_2\|) = \frac{1}{2} E [|F\{Z(\mathbf{s}_1)\} - F\{Z(\mathbf{s}_2)\}|].$$

The F-madogram is related to the extremal coefficient function as:

$$v_F(\|\mathbf{s}_1 - \mathbf{s}_2\|) = \frac{\theta(\|\mathbf{s}_1 - \mathbf{s}_2\|) - 1}{\theta(\|\mathbf{s}_1 - \mathbf{s}_2\|) + 1}.$$

- λ -Madogram, Naveau et al. [29].

$$v_\lambda(\|\mathbf{s}_1 - \mathbf{s}_2\|) = \frac{1}{2} E [|F^\lambda\{Z(\mathbf{s}_1)\} - F^{1-\lambda}\{Z(\mathbf{s}_2)\}|], \quad 0 \leq \lambda \leq 1,$$

where $F(z) = \exp(-1/z)$ is the unit Fréchet d.f.

The F-Madogram is similar to the λ -Madogram when $\lambda = 0.5$. The F-Madogram has the advantage of suggesting an estimator directly from its definition, see Cooley et al. [5].

Naveau et al. [29] discussed the estimation of the madogram. For the F-Madogram it was considered the plug in, \hat{F} , an estimate of the d.f., at the specified location and the binned estimate of the F-madogram. For the λ -Madogram, the binned λ -Madogram estimator and the adjusted estimator have been proposed.

4 A Case Study

For 21 different stations, in the North of Portugal, daily precipitation has been recorded. Our data refers to the maximum annual values for 20 years (1977–1996), in each station. A preliminary analysis of these data was done in Neves and Prata Gomes [30]. It would be nice to have more years of observations but surprisingly in recent years some missing values were found.

Figure 4 shows the region of Portugal where data have been collected on the left, and locations of the stations pointed out on the right. In each site and location, the maxima annual values of daily precipitation were considered in our study.

We began our analysis by performing marginal analysis and transformation. Daily values of precipitation at a given station x are dependent, however, the maxima values at each hydrological year can be considered almost independent.

As an illustration of the graphical diagnostic GEV fitting, graphics in two locations were chosen and displayed in Figs. 5 and 6.

The transformation of the data at each station to the unit Fréchet distribution was then performed, through the function `gev2frech()` of the package `SpatialExtremes`, Ribatet [34].

To estimate the spatial dependence structure, estimates of values of the extremal coefficient at two stations s_1 and s_2 are evaluated. For this, the `fitextcoeff()` was used considering both Smith, [38], and Schlather-Tawn, [37], estimators. As far as we know only these estimators are available in R package. Work is

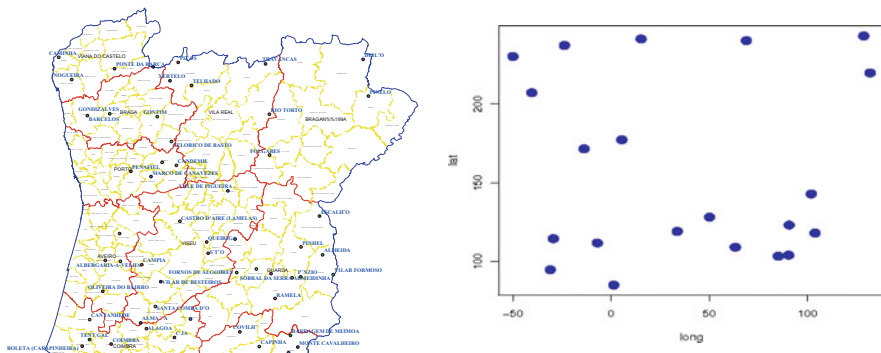


Fig. 4 Map of the North of Portugal (*left*) and the positions of the stations where data were recorded (*right*)

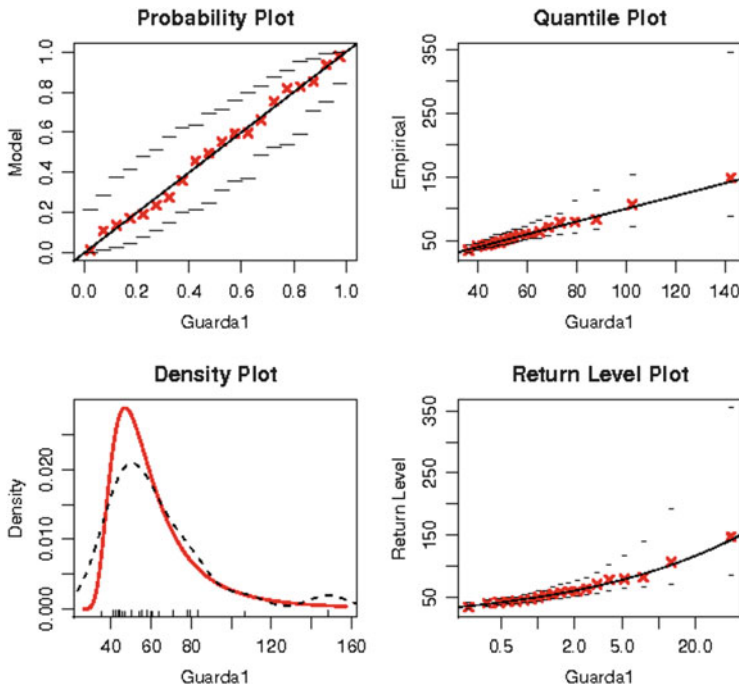


Fig. 5 GEV model diagnostic for data from Guarda

now in progress for including other functions in R environment. Figure 7 shows pairwise extremal coefficient estimates and lowess curves for Smith and Schlather-Tawn estimators and also the F-madogram, here obtained through `fmadogram()` function.

Trend surfaces were tried to be estimated in order to capture the spatial dependence structure by describing the marginal parameters as:

$$\lambda(x) = \beta_{o,\lambda} + \beta_{1,\lambda}lon(x) + \beta_{2,\lambda}lat(x)$$

$$\delta(x) = \beta_{o,\delta} + \beta_{1,\delta}lon(x) + \beta_{2,\delta}lat(x)$$

$$\xi(x) = \beta_{o,\xi}$$

where $lon(x)$ and $lat(x)$ denote the longitude and the latitude of the stations.

However, for our data, values of $\lambda(x)$, $\delta(x)$ and $\xi(x)$ showed a weak relationship with $lon(x)$ and $lat(x)$. Other trend surfaces need to be considered, but work is now in progress. Even though, considering the estimated matrix for Smith model as well as estimated sill and range for Schlather model, with the powered exponential correlation function, several simulations were performed on a 21×21 grid. Figure 8 displays one of those simulations for the Smith model and for the Schlather model.

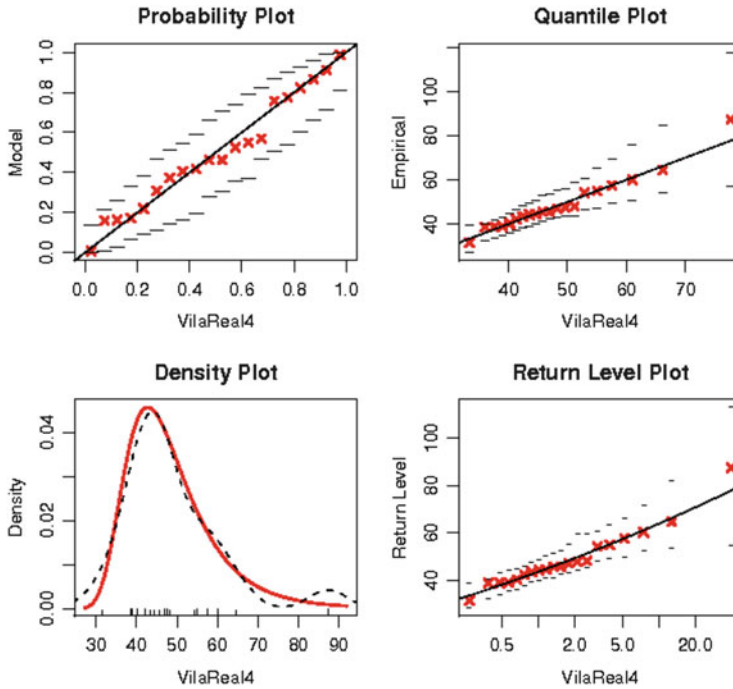


Fig. 6 GEV model diagnostic for data from Vila Real

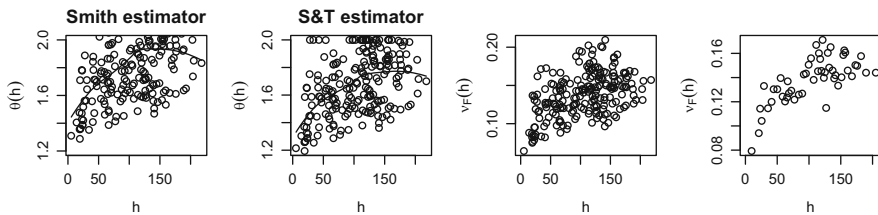


Fig. 7 Pairwise extremal coefficient estimates and lowess curves: Smith and Schlather-Tawn (left); F-madogram and binned madogram (right)

5 Concluding Remarks

This chapter was intended to introduce spatial models for extreme value analysis as well as to present an overview of functions available in the R software for doing that analysis. Max-stable processes are a natural generalization of multivariate extreme models and the most common way to deal with extreme value data in spatial statistics. Some procedures for estimating the spatial dependence available in the R

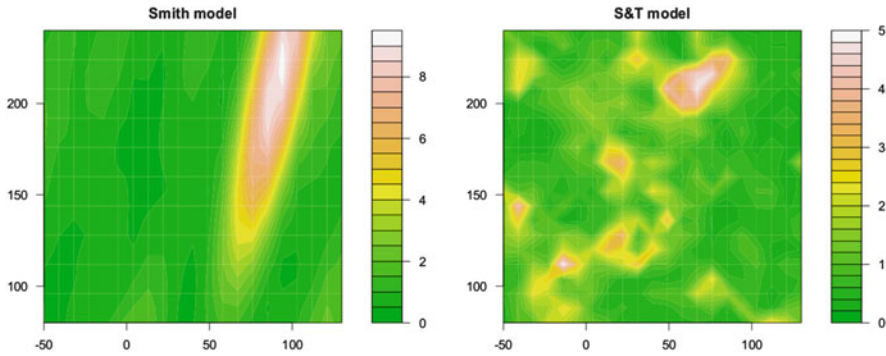


Fig. 8 One realization of the Smith (*left*) and Schlather and Tawn (*right*) models

environment were shown and some simulations using Smith and Schlather models were also performed in the R package `SpatialExtremes`.

Work is already in progress for including more max-stable models in R, as well as other approaches for modeling spatial dependence.

Some difficulties related to the amount of data available still remain. How to deal with missing values, in a given real situation, is another challenging point.

Acknowledgements Research partially supported by National Funds through **FCT**—Fundação para a Ciência e a Tecnologia, projects PEst-OE/MAT/UI0006/2014 (CEAUL)

References

1. Beirlant, J., Goegebeur, Y., Teugels, J., Segers, J.: *Statistics of Extremes: Theory and Applications*. Wiley, England (2005)
2. Bivand, R., Pebesma, E., Gomez-Rubio, V.: *Applied Spatial Data Analysis with R*, 2nd edn. Springer, New York (2013)
3. Castillo, E., Hadi, A.S., Balakrishnan, N., Sarabia, J.M.: *Extreme Value and Related Models in Engineering and Science Applications*. Wiley, New York (2005)
4. Coles, S.: *An Introduction to Statistical Modeling of Extreme Values*. Springer, London (2001)
5. Cooley D., Naveau P., Poncet P.: Variograms for spatial max-stable random fields. In: Springer (ed.) *Dependence in Probability and Statistics. Lecture Notes in Statistics Edition*, vol. 187, pp. 373–390. Springer, New York (2006)
6. Cooley, D., Cisewski, J., Erhardt, R.J., Jeon, S., Mannshardt, E., Omolo, B.O., Sun, Y.: A survey of spatial extremes: measuring spatial dependence and modeling spatial effects. *Revstat Stat. J.* **10**, 135–165 (2012)
7. Cressie, N.A.C.: *Statistics for Spatial Data*. Wiley, New York (1993)
8. Davison, A.C., Padoan, S.A., Ribatet, M.: Statistical modelling of spatial extremes (with discussion). *Stat. Sci.* **27**, 161–186 (2012)
9. de Haan, L.: On regular variation and its application to the weak convergence of sample extremes. Thesis, University of Amsterdam/Mathematical Centre Tract 32 (1970)
10. de Haan, L.: A spectral representation for max-stable processes. *Ann. Probab.* **12**(4), 1194–1204 (1984)

11. de Haan, L., Ferreira, A.: *Extreme Value Theory: An Introduction* Springer Science+Business Media, LLC, New York (2006)
12. Tiago de Oliveira, J. (ed.): *Statistical Extremes and Applications*. D. Reidel, Dordrecht (1983)
13. Diggle, P.J., Ribeiro, Jr. P.J.: *Model-Based Geostatistics*. Springer, New-York (2007)
14. Dodd, E.L.: The greatest and the least variate under general laws of error. *Trans. Am. Math. Soc.* **25**, 525–539 (1923)
15. Embrechts, P., Kluppelberg, C., Mikosch, T.: *Modelling Extremal Events for Insurance and Finance*, 3rd edn. Springer, Berlin (2001)
16. Finkenstadt, B., Rootzen, H. (eds.): *Extreme Values in Finance, Telecommunications and the Environment*. Chapman & Hall/CRC, London (2004)
17. Fisher R.A., Tippett, L.H.C.: Limiting forms of the frequency distribution of the largest and smallest member of a sample. *Proc. Camb. Philos. Soc.* **24**, 180–190 (1928)
18. Fréchet, M.: Sur la loi de probabilité de l'écart maximum. *Ann. Soc. Polon. Math.* **6**, 93–116 (1927)
19. Gilleland, E., Ribatet, M., Stephenson, A.G.: A software review for extreme value analysis. *Extremes* **16**(1), 103–119 (2013)
20. Gnedenko, B.V.: Sur la distribution limite d'une série aléatoire. *Ann. Math.* **44**, 423–453 (1943)
21. Gomes, M.I., Fraga Alves, M.I., Neves, C.: *Análise de Valores Extremos: uma Introdução*. Edições SPE, Lisboa (2013)
22. Guerin, N.: Climate and the statistics of extremes. Available in <http://actu.epfl.ch/news/climate-and-the-statistics-of-extremes/> (2012). Cited 25 June 2014
23. Gumbel, E.J.: Les valeurs extrêmes des distributions statistiques. *Ann. Inst. Henri Poincaré* **5**(2), 115–158 (1935)
24. Gumbel, E.J.: *Statistics of Extremes*. Columbia University Press/Dover Publications, New York (1958)
25. Kabluchko, Z., Schlather, M., de Haan, L.: Stationary max-stable fields associated to negative definite functions. *Ann. Prob.* **37**, 2042–2065 (2009)
26. Leadbetter, M., Lindgren, G., Rootzén, H.: *Extremes and related properties of random sequences and series*. Springer, New York (1983)
27. Matheron, G.: *Traité de géostatistique appliquée*. Tome I: Mémoires du Bureau de Recherches Géologiques et Minières, no. 14. Editions Technip, Paris (1962)
28. Matheron, G.: Suffit-il, pour une covariance, d'être de type positive. *Sciences de la Terre, série informatique géologique* **26**, 51–66 (1987)
29. Naveau, P., Guillou, A., Cooley, D., Diebolt, J.: Modeling pairwise dependence of maxima in space. *Biometrika* **96**(1) 1–17 (2009)
30. Neves, M. and Prata Gomes, D.: Geostatistics for spatial extremes. A case study of maximum annual rainfall in Portugal. *Procedia Environ. Sci.* **7**, 246–251 (2011)
31. Padoan, S.A., Ribatet, M., Sisson, S.A.: Likelihood-based inference for max-stable processes. *J. Am. Stat. Assoc.* **105**, 263–277 (2010)
32. Pebesma, E.J.: Multivariable geostatistics in S: the `gstat` package. *Comput. Geol.* **30**, 683–691 (2004)
33. Reiss, R.-D., Thomas, M.: *Statistical Analysis of Extreme Values: From Insurance, Finance, Hydrology and other Fields*. Birkhauser, Basel (2007)
34. Ribatet, M.: *A User's Guide to the SpatialExtremes Package*. École Polytechnique Fédérale de Lausanne, Switzerland (2011)
35. Ribeiro, Jr., P.J., Christensen, O.F., Diggle, P.J.: `geoR` and `geoRglm`: software for model-based geostatistics. In: *Proceedings of DSC*, vol. 2 (2003)
36. Schlather, M.: Models for stationary max-stable random fields. *Extremes* **5**(1) 33–44 (2002)
37. Schlather, M., Tawn, J.: A dependence measure for multivariate and spatial extremes: properties and inference. *Biometrika* **90**(1) 139–156 (2003)
38. Smith, R.: *Max-stable processes and spatial extremes*. Unpublished manuscript (1990)
39. Smith, E.L., Stephenson, A.G.: An extended Gaussian max-stable process model for spatial extremes. *J. Stat. Plann. Inference* **139**, 1266–1275 (2009)

40. Stephenson, A.G.: `evd`: Extreme value distributions. *R News* **2**(2), 31–32 (2002). <http://CRAN.R-project.org/doc/Rnews/>
41. Stephenson, A.G.: `ismev`: an introduction to statistical modeling of extreme values. R package version 1.38 (2012). <http://CRAN.R-project.org/package=ismev>
42. von Mises, R.: La distribution de la plus grande de n valeurs. *Revue Math. Union Interbalcanique* **1**, 141–160 (1936) [Reprinted in *Selected Papers Volumen II*, pp. 271–294. American Mathematical Society, Providence (1954)]