



HAL
open science

Modeling extreme rainfall A comparative study of spatial extreme value models

Quentin Sebillé, Anne-Laure Fougères, Cécile Mercadier

► **To cite this version:**

Quentin Sebillé, Anne-Laure Fougères, Cécile Mercadier. Modeling extreme rainfall A comparative study of spatial extreme value models. *Spatial Statistics*, 2017, Volume 21 (Part A), pp.187-208. hal-01702688

HAL Id: hal-01702688

<https://hal.science/hal-01702688>

Submitted on 7 Feb 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Modeling extreme rainfall

A comparative study of spatial extreme value models

Quentin Sebille*, Anne-Laure Fougères*, Cécile Mercadier*

January 15, 2017

Abstract

In this paper, focus is done on spatial models for extreme events and on their respective efficiency regarding the estimation of two risk measures: one extrapolating marginal distributions and one summarizing the spatial bivariate dependence of extremes. A wide comparison is performed on an innovative simulation plan that has been specifically designed from a daily precipitation data set. The objective of this paper is twofold: firstly, pointing out the inherent properties of each model, and secondly, advising users on how to choose the model depending on the specific type of risk.

Keywords: Spatial modeling of extreme events; extreme value theory; max-stable processes; hierarchical models; spatial prediction; precipitation data.

Classification 2010: 60G70; 62G32; 62H11; 62P12.

1 Introduction

Analyses of extreme values of environmental variables such as precipitation are of great importance since it involves human lives as well as considerable losses of money when a catastrophic event occurs. For instance, between May 28th and June 7th 2016, extreme precipitation events affected a part of Europe including France. This huge and sudden amount of rainfall caused nineteen deaths and, for France only, one billion Euro in damages. Accurate risk measures for such extreme phenomena are therefore needed to prevent from this type of scenario.

The risk estimation of these events is challenging because they involve values that are beyond the range of the observations. For this purpose, adapted tools come from extreme value theory. See for instance de Haan and Ferreira (2006), Beirlant et al. (2004), Finkenstädt and Rootzén (2004) and Coles (2001). Since precipitation phenomena have a spatial feature and data are generally observed at several stations, the dedicated setting to handle this question is that of (conditionally) max-stable spatial models. Detailed and helpful reviews on these models are Cooley et al. (2012), Davison et al. (2012) or Ribatet (2013).

Five spatial models have been selected among the most popular and the most recent in the literature. This choice includes Bayesian and frequentist concepts, and goes from simple to more complex spatial dependence structure. Within this paper, the main goal is to answer: *Which of the five competing models yields the best spatial prediction for extreme behaviors of simulated processes mimicking precipitation data?* Addressing this question supposes in particular a careful consideration of the data sets involved, as well as a relevant choice of performance criteria.

Two comparative criteria adapted to extreme events prediction are evaluated. The estimation of rare events at a location where no data is available is handled first; this induces the capacity of spatial

*Univ Lyon, Université Claude Bernard Lyon 1, CNRS UMR 5208, Institut Camille Jordan, 43 blvd. du 11 novembre 1918, F-69622 Villeurbanne cedex, France.

extrapolation of the extreme behavior when looking at marginal information only. A clear and well known way to summarize this marginal information into a concrete risk measure is the return level. Then a second and complementary criterion is the measure of extreme sets for the bivariate distribution at a pair of locations. This aims at capturing the spatial dependence structure of extremes. One could also look at indices involving more than two dimensions, but most of the dependent models for extremes have explicit formulae only in the bivariate case, so it would induce heavier numerical calculations.

Several options can be chosen to define the terms of comparison. One possible option could be to start from an expert point of view and compare each model with an a priori value of the previous criteria. To depart from a subjective choice, an intensive simulation study has been preferred. *How could one choose inside each of the five models a meaningful representative mimicking rainfall data?* Such a question has not a straight answer. A parametric bootstrap procedure has been set up here. More precisely:

- A real precipitation data set is considered over a central-east region of France on which each of the five spatial models is fitted. These fitted models are then fixed to play the role of extreme rainfall generators.
- Both criteria (return level and bivariate extremal dependence) are evaluated on each generator so that every single generating process has its own true value of a criterion.
- The five spatial models are again considered and fitted on each generated sample drawn from one of the five generators. Both criteria are then evaluated on each fitted model and finally compared to the corresponding (known) true value.

To our knowledge, this way to proceed is original. We believe it offers a better insight into the comparative study than applying a given methodology to a single data set, or estimating a given process inside its own class of model only.

The remainder of this article is organized as follows. The theoretical background of extreme value theory is addressed in the following section, with an emphasis on the so-called block maxima approach and on the five (conditionally) max-stable spatial models that we consider. Section 3 describes the simulation study, the two criteria used to evaluate each case and the results. Conclusions are drawn in Section 4, where some recommendations are provided with respect to different purposes.

2 Notations and models

2.1 Definition of max-stable processes

Let \mathcal{S} be a compact subset of \mathbb{R}^d that represents the spatial region of interest, d being a positive integer. Consider a random process $Y(\cdot) = \{Y(s)\}_{s \in \mathcal{S}}$ defined over \mathcal{S} , with continuous sample paths. Write $Y_1(\cdot), \dots, Y_T(\cdot)$ for independent copies of $Y(\cdot)$. The process Y is called *max-stable* if for each $T > 1$, there exist continuous functions $a_T(\cdot) > 0$ and $b_T(\cdot) \in \mathbb{R}$ such that:

$$\bigvee_{t=1}^T \frac{Y_t(\cdot) - b_T(\cdot)}{a_T(\cdot)} \stackrel{d}{=} Y(\cdot),$$

where $\stackrel{d}{=}$ denotes equality in distribution. Such max-stable processes arise as non degenerate limits for pointwise maxima of stochastic processes on \mathcal{S} , and have been introduced by de Haan (1984). In particular, for each $s \in \mathcal{S}$, the random variable $Y(s)$ follows a generalized extreme value distribution $\text{GEV}(\mu(s), \sigma(s), \xi(s))$, where the location $\mu(s) \in \mathbb{R}$, scale $\sigma(s) > 0$ and shape $\xi(s) \in \mathbb{R}$ parameters are indexed over space \mathcal{S} . Recall that the cumulative distribution function (cdf) of a $\text{GEV}(\mu, \sigma, \xi)$ random variable Y is:

$$\Pr(Y \leq y) = \begin{cases} \exp\left(-\left[1 + \xi \frac{y-\mu}{\sigma}\right]_+^{-1/\xi}\right) & \text{if } \xi \neq 0 \\ \exp\left(-\exp\left[-\frac{y-\mu}{\sigma}\right]\right) & \text{if } \xi = 0, \end{cases}$$

where $z_+ = \max(0, z)$ for all $z \in \mathbb{R}$. Thanks to the one-to-one mapping:

$$Z(s) = \left[1 + \xi(s) \frac{Y(s) - \mu(s)}{\sigma(s)} \right]^{1/\xi(s)}, \quad (1)$$

one transforms $Y(\cdot)$ into a *simple max-stable process* $Z(\cdot)$, that is with unit Fréchet margins, corresponding to a GEV(1,1,1) process at each spatial location in \mathcal{S} .

The joint cdf of $Z(\cdot)$ at a set of sites $\{s_1, \dots, s_n\} \subset \mathcal{S}$ is given by:

$$\Pr(Z(s_1) \leq z_1, \dots, Z(s_n) \leq z_n) = \exp[-V_{s_1, \dots, s_n}(z_1, \dots, z_n)], \quad (2)$$

in terms of an *exponent measure* V_{s_1, \dots, s_n} , also denoted V when there is no ambiguity. This exponent measure contains the information about the spatial dependence of extremes (see e.g de Haan and Ferreira, 2006, Chapter 9). The two limit cases are the independence case, with $V(z_1, \dots, z_n) = \sum_{i=1}^n z_i^{-1}$ and the total positive dependence case, with $V(z_1, \dots, z_n) = \bigvee_{i=1}^n z_i^{-1}$. A common measure of spatial dependence between the set of sites $\{s_1, \dots, s_n\}$ is the extremal coefficient $\theta(s_1, \dots, s_n)$ of Schlather and Tawn (2003) defined by:

$$\Pr(Z(s_1) \leq z, \dots, Z(s_n) \leq z) = \Pr(Z(s_1) \leq z)^{\theta(s_1, \dots, s_n)}.$$

One can interpret $\theta(s_1, \dots, s_n) \in [1, n]$ as the equivalent number of components of $\{Z(s_1), \dots, Z(s_n)\}$ that are independent. The extremal coefficient is linked to the exponent measure via $\theta(s_1, \dots, s_n) = V_{s_1, \dots, s_n}(1, \dots, 1)$.

In practice, the description of a max-stable process is done in two parts. First, the marginal distribution is captured by the processes that represent the GEV parameters. Classically, inference on the marginals is carried out under a linear model involving covariates. Such details are postponed until later (as in Equation (4) for instance). Second, the spatial dependence of extremes is measured by $V_{s_1, \dots, s_n}(\cdot)$, or summarized through $\theta(s_1, \dots, s_n)$. Parametric models for V can be helpful and some examples are presented in the next section.

2.2 Spectral representation and parametric models

Max-stable processes can be described thanks to the following spectral representation, due to de Haan (1984). Let $\{\zeta_j\}_{j \in \mathbb{N}}$ be a Poisson point process on $(0, \infty)$ with intensity measure $\zeta^{-2} d\zeta$ and consider independent copies $\{W_j(s), s \in \mathcal{S}\}_{j \in \mathbb{N}}$ of a stationary process $W(\cdot)$ assuming $\sup_{s \in \mathcal{S}} W(s) < \infty$ with $\mathbb{E}[W(s)_+] = 1$. Then the process $Z(\cdot)$ defined for each $s \in \mathcal{S}$ by

$$Z(s) = \max_{j \geq 1} \zeta_j W_j(s), \quad (3)$$

is max-stable with unit Fréchet margins. Its joint cdf is expressed as:

$$\Pr(Z(s) \leq z(s), s \in \mathcal{S}) = \exp\left(-\mathbb{E}\left[\sup_{s \in \mathcal{S}} \frac{W(s)}{z(s)}\right]\right).$$

Different choices for the so-called spectral processes $W_j(\cdot)$ in (3) lead to different max-stable models. Three such max-stable models are now presented. Along this subsection, the dependence structure between two given sites s_1 and s_2 will be expressed skipping the notation in terms of the s_i 's and h will denote the Euclidean distance between s_1 and s_2 .

2.2.1 The Extremal Gaussian Process: EGP

A first possible choice is to take $W(\cdot) = \sqrt{2\pi}\varepsilon(\cdot)_+$, in terms of a standard stationary Gaussian process $\varepsilon(\cdot)$ with correlation function $\rho(\cdot)$. The resulting max-stable process, defined by Schlather (2002), is

called the *Extremal Gaussian Process* (EGP). The associated bivariate exponent measure is:

$$V_{\text{EGP}}(z_1, z_2) = \frac{1}{2} \left(\frac{1}{z_1} + \frac{1}{z_2} \right) \left(1 + \sqrt{1 - 2 \frac{[\rho(h) + 1]z_1 z_2}{(z_1 + z_2)^2}} \right).$$

If the $\varepsilon_j(\cdot)$'s are assumed isotropic, then so is the max-stable process $Z(\cdot)$. The bivariate extremal coefficient of the EGP can be shown to be

$$\theta_{\text{EGP}}(h) = 1 + \sqrt{\frac{1 - \rho(h)}{2}}.$$

A shortcoming of the EGP is that it cannot account for asymptotic independence, even when the distance h between two sites increases to infinity. Due to $\rho(\cdot)$ being positive definite, the bivariate extremal coefficient, in dimension $d = 2$, does not span over the interval $[1, 2]$ but instead spans the interval of $[1, 1.838]$ (see A for details).

Different correlation functions $\rho(\cdot)$ can be chosen. We work with the powered exponential form $\rho(h) = \exp[-(h/\lambda)^\nu]$, where $\nu \in (0, 2]$ and $\lambda > 0$ are respectively the smoothness and the range parameters. These parameters need to be estimated when fitting the EGP, and we denote them by ν_{EGP} and λ_{EGP} , respectively.

2.2.2 The Brown-Resnick Process: BRP

A second possibility is to take $W(\cdot) = \exp[\varepsilon(\cdot) - \gamma(\cdot)]$, where here $\varepsilon(\cdot)$ is a Gaussian process with stationary increments and semivariogram $\gamma(\cdot)$. Then, representation (3) leads to a process $Z(\cdot)$ originally introduced by Brown and Resnick (1977), see also Kabluchko et al. (2009). This latter process is traditionally called the *geometric Gaussian process* or the *Brown-Resnick Process*, and will be denoted by BRP in the following. The associated bivariate exponent measure is

$$V_{\text{BRP}}(z_1, z_2) = \frac{1}{z_1} \Phi \left(\sqrt{\frac{\gamma(h)}{2}} + \frac{\log(z_2/z_1)}{\sqrt{2\gamma(h)}} \right) + \frac{1}{z_2} \Phi \left(\sqrt{\frac{\gamma(h)}{2}} + \frac{\log(z_1/z_2)}{\sqrt{2\gamma(h)}} \right),$$

and its extremal coefficient is given by

$$\theta_{\text{BRP}}(h) = 2\Phi \left(\sqrt{\frac{\gamma(h)}{2}} \right).$$

When $\varepsilon(\cdot)$ is a fractional Brownian motion, the semivariogram $\gamma(\cdot)$ is given by $\gamma(h) = (\frac{h}{\lambda})^\nu$, where $\lambda > 0$ and $\nu \in (0, 2]$ are respectively the range and smoothness parameters. These parameters must be estimated to describe the spatial structure, and they are denoted by λ_{BRP} and ν_{BRP} , respectively.

2.2.3 The Extremal- t Process: ETP

Opitz (2013) obtained the only possible max-stable limit for asymptotically dependent elliptical processes, namely the *Extremal- t Process*, denoted ETP here. Let $\delta > 0$, $m_\delta = \sqrt{\pi} 2^{1-\delta/2} \Gamma(\frac{\delta+1}{2})^{-1}$ and $\varepsilon(\cdot)$ be a standard stationary Gaussian process with correlation function $\rho(\cdot)$. As for the EGP, we work with the powered exponential form for $\rho(\cdot)$.

The spectral process of the extremal- t model is $W(\cdot) = m_\delta \varepsilon(\cdot)_+^\delta$ and the corresponding exponent measure V for a pair of sites $\{s_1, s_2\}$ is given by

$$V_{\text{ETP}}(z_1, z_2) = \frac{1}{z_1} T_{\delta+1} \left(-\frac{\rho(h)}{\varphi_\delta(h)} + \frac{1}{\varphi_\delta(h)} \left(\frac{z_2}{z_1} \right)^{1/\delta} \right) + \frac{1}{z_2} T_{\delta+1} \left(-\frac{\rho(h)}{\varphi_\delta(h)} + \frac{1}{\varphi_\delta(h)} \left(\frac{z_1}{z_2} \right)^{1/\delta} \right),$$

where $T_{\delta+1}(\cdot)$ is the cdf of the Student distribution with $(\delta + 1)$ degrees of freedom and $\varphi_\delta(h) = \sqrt{(1 - \rho(h)^2)/(\delta + 1)}$. The extremal coefficient of the ETP is given by

$$\theta_{\text{ETP}}(h) = 2T_{\delta+1}\left(\frac{1 - \rho(h)}{1 - \rho(h)^2}(\delta + 1)\right).$$

Note that both the EGP of Schlather (2002) and the BRP of Brown and Resnick (1977) are special cases of the ETP. Indeed, the EGP is obtained straightforwardly when $\delta = 1$, while the ETP converges towards the BRP when $\delta \rightarrow \infty$. The dependence parameters that need to be estimated are then: δ_{ETP} , the degree of freedom, and the parameters of the correlation function, namely λ_{ETP} and ν_{ETP} , respectively.

2.3 Hierarchical modeling

Hierarchical models typically assume that the observed process $Y(\cdot)$ is spatially independent conditionally on unobserved latent processes or variables. The interested reader may find examples of such hierarchical spatial models in Banerjee et al. (2004) for instance. Models based on hierarchical approaches are usually defined within the Bayesian paradigm. The estimation is generally performed through a Markov Chain Monte Carlo (MCMC) algorithm, which allows to sample from the posterior distribution of the parameters, given the data. Recent studies exploiting hierarchical modeling for extreme precipitation are for instance Apputhurai and Stephenson (2013) and Dyrddal et al. (2015).

2.3.1 The latent variable model: LVM

Davison et al. (2012) introduced a simple hierarchical structure for spatial extremes, called *the latent variable model* (LVM). Note that the usual lack of clear spatial pattern for the shape $\xi(\cdot)$ when dealing with precipitation data, jointly with the difficulty of estimating this parameter lead in this paper to consider $\xi(\cdot) \equiv \xi_0$. The random variables $\{Y(s)\}_{s \in \mathcal{S}}$ are assumed to be independent conditionally on latent processes that describe the GEV parameters. More precisely,

$$Y(s) | \{\mu, \sigma, \xi\} \stackrel{\text{indep}}{\sim} \text{GEV}(\mu(s), \sigma(s), \xi(s)),$$

where

$$\begin{cases} \mu(s) &= \beta_\mu^T c(s) + \varepsilon_\mu(s), \\ \sigma(s) &= \beta_\sigma^T c(s) + \varepsilon_\sigma(s), \\ \xi(s) &\equiv \xi_0, \end{cases} \quad (4)$$

where the random parts $\varepsilon(\cdot)$ are assumed to be independent stationary zero-mean Gaussian processes and where $c(s)$ denotes covariates associated to each position $s \in \mathcal{S}$. As a consequence, the mean function of the latent location parameter process $\mu(\cdot)$ (resp. $\sigma(\cdot)$) is written as a linear combination of covariates with unknown vector of coefficients β_μ (resp. β_σ). The covariates we use in this paper are the constant, the longitude, latitude and altitude. Note that Davison et al. (2012) consider the exponential form $\rho_\varepsilon(h) = \delta \cdot \exp(-h/\lambda)$ to model the correlation of the latent Gaussian processes. We follow this choice to be consistent with the choices made in Sections 2.2.1 and 2.2.3 when defining the EGP and ETP.

The assumption of conditional independence leads to consider a multivariate exponent measure $V_{\text{LVM}}(z_1, \dots, z_n) = \sum_{i=1}^n z_i^{-1}$ and the bivariate extremal coefficient $\theta_{\text{LVM}} \equiv 2$ that does not depend on the positions of s_1, \dots, s_n .

2.3.2 The Hierarchical Kernel Extreme Value Process: HKEVP

From Davison et al. (2012), one knows that the LVM is particularly appealing when the estimation of the marginal distributions is of interest, as it focuses on modeling the GEV parameters. The main

drawback is that the dependence structure of extremes is not considered since conditional independence is assumed.

The aim of this subsection is to present a model that describes both the marginal effect and the dependence structure within a Bayesian framework. The *Hierarchical Kernel Extreme Value Process* (HKEVP) has been introduced by Reich and Shaby (2012) and further developed in Shaby and Reich (2012) and Reich et al. (2014). It is defined as follows. Suppose that $Y(s) \sim \text{GEV}(\mu(s), \sigma(s), \xi(s))$ and model the margins as the LVM as described by (4). Again, exponential correlations are used for the latent Gaussian processes $\varepsilon_\mu(\cdot)$ and $\varepsilon_\sigma(\cdot)$ in order to be consistent with previous choices.

Consider now $Z(\cdot)$, the associated simple max-stable process. The HKEVP allows for spatial dependence combined with a nugget effect by assuming $Z(s) = U_\alpha(s)\vartheta_\alpha(s)$ for all $s \in \mathcal{S}$, where $\alpha \in (0, 1]$ is a parameter that controls the magnitude of the nugget effect.

- $U_\alpha(\cdot)$ is a spatially-independent process with common marginal distribution $\text{GEV}(1, \alpha, \alpha)$ representing the nugget effect.
- $\vartheta_\alpha(s) = \left(\sum_{\ell=1}^L A_\ell \omega_\ell^{1/\alpha}(s) \right)^\alpha$ describes the spatial dependence structure, constructed with:
 - deterministic positive kernel functions $\{\omega_1(\cdot), \dots, \omega_L(\cdot)\}$ satisfying $\sum_{\ell=1}^L \omega_\ell(s) = 1, \forall s \in \mathcal{S}$,
 - and associated independent and identically distributed (iid) random variables $\{A_1, \dots, A_L\}$ following the positive stable distribution with characteristic exponent α , denoted $\text{PS}(\alpha)$.

The kernels $\omega_\ell(\cdot)$ used in Reich and Shaby (2012) are rescaled Gaussian densities. More precisely, let $\mathcal{V} := \{v_1, \dots, v_L\} \subset \mathcal{S}$ be a set of knots, and

$$\omega_\ell(s) = \frac{K(s|v_\ell, \tau)}{\sum_{j=1}^L K(s|v_j, \tau)}, \quad \forall \ell \in \{1, \dots, L\},$$

where τ is the bandwidth parameter and

$$K(s|v_\ell, \tau) = \frac{1}{2\pi\tau^2} \exp \left[-\frac{1}{2\tau^2} (s - v_\ell)^T (s - v_\ell) \right].$$

The product of a fixed $\vartheta_\alpha(s)$ by a $\text{GEV}(1, \alpha, \alpha)$ distributed random variable $U_\alpha(s)$ gives a GEV distributed random variable $Z(s)$ whose parameters are known. More explicitly, by conditioning on $\vartheta_\alpha(s)$ (accordingly, over the random variables A_1, \dots, A_L), one gets:

$$Z(s) \mid \vartheta_\alpha(s) \sim \text{GEV}(\vartheta_\alpha(s), \alpha\vartheta_\alpha(s), \alpha),$$

and using relation (1) between processes $Y(\cdot)$ and $Z(\cdot)$ leads to the following hierarchical formulation:

$$\begin{aligned} Y(s) \mid \mu, \sigma, \xi, \alpha, \vartheta_\alpha &\stackrel{\text{indep}}{\sim} \text{GEV}(\mu^*(s), \sigma^*(s), \xi^*(s)), \\ \mu^*(s) &= \mu(s) + \frac{\sigma(s)}{\xi(s)} \left(\vartheta_\alpha(s)^{\xi(s)} - 1 \right), \\ \sigma^*(s) &= \alpha\sigma(s)\vartheta_\alpha(s)^{\xi(s)}, \\ \xi^*(s) &= \alpha\xi(s), \\ \vartheta_\alpha(s) &= \left(\sum_{\ell=1}^L A_\ell \omega_\ell^{1/\alpha}(s) \right)^\alpha, \\ A_1, \dots, A_L &\stackrel{\text{iid}}{\sim} \text{PS}(\alpha). \end{aligned}$$

If the conditioning random variable A follows the positive stable distribution as stated, then this model has the nice feature that the multivariate distribution obtained is both conditionally *and* unconditionally max-stable, and has an explicit exponent measure for any set of sites $\{s_1, \dots, s_n\}$ with

logistic form (Stephenson, 2009):

$$V_{\text{HKEVP}}(z_1, \dots, z_n) = \sum_{\ell=1}^L \left[\sum_{i=1}^n \left(\frac{\omega_{\ell}(s_i)}{z_i} \right)^{1/\alpha} \right]^{\alpha} .$$

Therefore, the bivariate extremal coefficient of the HKEVP is, for each $s_1 \neq s_2$:

$$\theta_{\text{HKEVP}}(s_1, s_2) = \sum_{\ell=1}^L \left(\omega_{\ell}(s_1)^{1/\alpha} + \omega_{\ell}(s_2)^{1/\alpha} \right)^{\alpha} .$$

The parameter α plays a central role in the HKEVP: it controls the degree of spatial dependence in the extremes by tuning the magnitude of both processes $U_{\alpha}(\cdot)$ and $\vartheta_{\alpha}(\cdot)$. If α tends to 0, the nugget is low, with $U_{\alpha}(\cdot)$ approximately equal to 1 everywhere, so that $Z(\cdot)$ is very smooth. On the opposite, if $\alpha = 1$, the random variables $\{A_1, \dots, A_L\}$ are degenerated to 1, and by using the normalized condition on the kernel functions, one gets therefore $\vartheta_{\alpha}(\cdot) \equiv 1$. The process $Z(\cdot)$ is then a full-nugget spatial process. Refer to (Reich and Shaby, 2012, Section 2.1) for additional comments.

When jointly α tends to 0 and L tends to infinity, the HKEVP converges to a pioneer max-stable model known as the Gaussian extreme value model, or Smith's model (Smith, 1990). The latter model has not been considered in the present paper because of its unrealistic over smooth feature and a resulting lack of fit on real data. See for instance Davison et al. (2012) or Davison et al. (2013), where the Gaussian extreme value model is compared with the EGP and the BRP, among others.

2.4 Inference procedures

2.4.1 Inference for max-stable processes

Let $y_t(s_i)$ be the t -th observation of $Y(s_i)$, for $t \in \{1, \dots, T\}$ and $i \in \{1, \dots, n\}$, where T stands for the number of years of study, and let y be the set of all observations. Under temporal independence, the likelihood of y under a max-stable spatial model is

$$L(y; \psi_{\text{GEV}}, \psi_V) = \prod_{t=1}^T g_n(y_t(s_1), \dots, y_t(s_n); \psi_{\text{GEV}}, \psi_V) , \quad (5)$$

where g_n is the n -variate max-stable density function. The latter depends on the n -multivariate marginal parameters μ, σ , and ξ that are completely determined through a set of spatial parameters ψ_{GEV} using for instance formulation (6), and whose exponent function V is driven by parameters ψ_V .

The computation of $g_n(y_t(s_1), \dots, y_t(s_n); \psi_{\text{GEV}}, \psi_V)$, for $t \in \{1, \dots, T\}$ is obtained by the transformation of the margins via (1) and by differentiating (2) with respect to (z_1, \dots, z_n) . One gets:

$$\frac{\partial^n}{\partial z_1 \dots \partial z_n} V_{s_1, \dots, s_n}(z_1, \dots, z_n) = \sum_{\Pi \in \mathcal{P}} \prod_{\pi \in \Pi} -V_{\pi}(z_1, \dots, z_n) \exp \{-V_{s_1, \dots, s_n}(z_1, \dots, z_n)\} ,$$

where \mathcal{P} denotes the set of all partitions of indices $\{1, \dots, n\}$ and V_{π} stands for the partial derivatives of the exponent measure V_{s_1, \dots, s_n} with respect to the elements of $\pi \in \Pi \subset \mathcal{P}$. Unfortunately, the cardinal of \mathcal{P} equals the Bell's number and is greater than 10^5 when $n \geq 10$, making the full likelihood quickly intractable even when dealing with a reasonable amount of positions.

For this reason, inference on max-stable processes has been done by maximizing the *composite likelihood* (Lindsay, 1988), and more specifically the *pairwise likelihood*, which is evaluated at all pairs of positions $\{s_i, s_j\}_{i, j \in \{1, \dots, n\}}$ and has the following form:

$$L_{\text{pairwise}}(y; \psi_{\text{GEV}}, \psi_V) := \prod_{t=1}^T \prod_{i=1}^{n-1} \prod_{j=i+1}^n g_2(y_t(s_i), y_t(s_j); \psi_{\text{GEV}}, \psi_V) .$$

In the latter formulation, $g_2(y_t(s_i), y_t(s_j); \psi_{\text{GEV}}, \psi_V)$ is the bivariate max-stable density function with spatial marginal parameters ψ_{GEV} and dependence parameters ψ_V .

Estimation procedures using maximization of the pairwise likelihood for max-stable processes (Padoan et al., 2010) have been implemented in the following two R packages: `SpatialExtremes` (Ribatet, 2015) and `CompRandFld` (Padoan and Bevilacqua, 2015). The `RandomFields` package (Schlather et al., 2016) allows to simulate from these models. In Section 3, we use `SpatialExtremes` which allows both fit and simulation at non-gridded locations.

2.4.2 Inference for hierarchical models

Hierarchical models such as the LVM and the HKEVP are generally well handled in a Bayesian setting, and inference is thus performed through a MCMC Metropolis-within-Gibbs procedure.

For the LVM, the posterior distribution of the GEV parameters is given by:

$$\pi(\mu, \sigma, \xi | y) \propto L(y; \mu, \sigma, \xi, \psi_{\text{GEV}}) \pi(\mu, \sigma, \xi | \psi_{\text{GEV}}) \pi(\psi_{\text{GEV}} | \dots),$$

where \propto stands for *proportional to* and:

- $\psi_{\text{GEV}} = \{\beta_\mu, \beta_\sigma, \delta_\mu, \delta_\sigma, \lambda_\mu, \lambda_\sigma, \xi_0\}$ is the set of parameters related to the latent Gaussian processes,
- $L(y; \mu, \sigma, \xi, \psi_{\text{GEV}})$ is the likelihood (5) with independent spatial structure,
- $\pi(\mu, \sigma, \xi | \psi_{\text{GEV}})$ is the product of the Gaussian densities of the latent processes μ, σ and ξ ,
- $\pi(\psi_{\text{GEV}} | \dots)$ denotes the prior densities associated to the parameters ψ_{GEV} . Hyperparameters of these prior distributions are symbolized by the dots.

For the HKEVP, the posterior distribution of parameters $(\mu, \sigma, \xi, \alpha, \tau)$ is given by:

$$\pi(\mu, \sigma, \xi, \alpha, \tau | y) \propto L(y | \mu, \sigma, \xi, A, \alpha, \tau, \psi_{\text{GEV}}) \pi(\mu, \sigma, \xi | \psi_{\text{GEV}}) \pi(A | \alpha) \pi(\psi_{\text{GEV}}, \alpha, \tau | \dots),$$

where :

- $L(y | \mu, \sigma, \xi, A, \alpha, \tau, \psi_{\text{GEV}})$ is the likelihood (5) with *independent* spatial structure, since conditioning over the process ϑ_α (according to A, α and τ) gives spatial independence,
- $\pi(\mu, \sigma, \xi | \psi_{\text{GEV}})$ is the product of the Gaussian densities of the latent processes μ, σ and ξ ,
- $\pi(A | \alpha)$ is the positive stable density with characteristic exponent α of the conditioning random variable A ,
- $\pi(\psi_{\text{GEV}}, \alpha, \tau | \dots)$ is the product of independent prior densities for the spatial marginal parameters ψ_{GEV} and for the two parameters α and τ of the exponent function V_{HKEVP} .

As far as we know, there was no package including the fitting procedure for the HKEVP. The only connected function is the routine `abba` included in the package `extRemes` of Gilleland and Katz (2011) and related to the recent paper of Stephenson et al. (2015). Therein, the model uses a CAR prior over a network of thousands of gridded locations and is therefore not designed to make prediction outside the observed set of sites. The authors of the HKEVP propose nonetheless an open code available on Reich's website¹. Sebille (2016) recently published on CRAN the R package `hkevp` that contains in particular a routine called `hkevp.fit` which fits the HKEVP model. This function is widely inspired by Reich & Shaby's code, and the main changes are listed in the reference manual of `hkevp`.

The MCMC algorithm associated to the LVM is available in the R package `SpatialExtremes` under the routine `latent`, and in the `hkevp` library under the function `latent.fit`. The latter allows to set

¹Homepage <http://www4.stat.ncsu.edu/~reich/>

the shape parameter $\xi(s)$ as spatially-invariant. To keep a maximum of coherences between the five models that are compared in this paper, we work with the `latent.fit` routine.

The two MCMC algorithms are applied for this paper with 30.000 iterations after a burn-in period of size 15.000, and with a thinning procedure of length 15 so that the resulting Markov chains are of length 1.000. Whenever pointwise estimation is needed, the median of the corresponding chain is taken.

3 Comparison of the spatial models

This section is threefold. Firstly, the design of the simulation study is outlined. Secondly, the spatial models are fitted on the resulting artificial simulated data, and the results are analyzed. Finally, additional characteristics of these models are discussed. One should emphasize here that – as far as we know – this is the first time that the HKEVP takes part into a wide comparative simulation study.

3.1 Designing the simulation as a precipitation data set

In this section, the five models described in Section 2, namely the LVM, HKEVP, EGP, BRP, and ETP are applied to a set of precipitation data recorded in France.

Along the paper, the models are sorted this way to respect an increasing “smoothness within dependence modeling”, going from conditional independence to a spatial dependence structure with finite conditioning and then to three continuous max-stable dependence structures.

3.1.1 The precipitation data

The data set used in this article is extracted from the one analyzed by Penot (2014) and composed of $n = 61$ rain gauges in central-east of France, including 10 stations from EDF/DTG and 51 stations from Météo France. A map of France is given on Figure 1a with the positions of each meteorological station in blue and the region \mathcal{S} in a red rectangle. Figure 1b indicates the elevation associated to each station with a color code. The latter shows that the considered region is relatively flat, with most of the positions located below 600m.

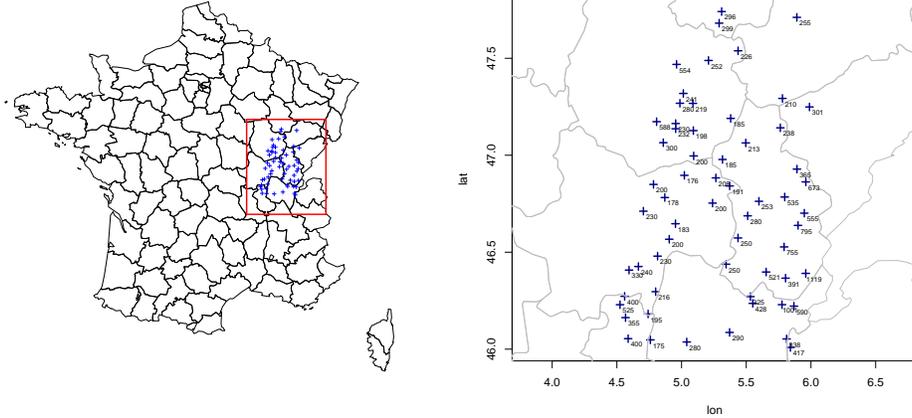
When available, annual maxima are taken at each station in the period between 1961 and 2005, which leads to $T = 45$ observations per site. For a given meteorological station, each annual maximum has been computed if there was less than 30 days missing in the corresponding year. Half the stations have complete recorded time series, and all of them have at least 30 non-missing annual maxima.

Pointwise estimations of the GEV distribution are performed at each site of the region \mathcal{S} . Figure 2 shows the estimated values of μ, σ and ξ on the map with a color code. From these maps, we can visually assess for spatial trends in the GEV parameters. A brief analysis indicates that the parameters μ and σ can be appropriately modeled using longitude, latitude and altitude as covariates. The shape parameter ξ at each location has no clear visual pattern, and a regression analysis failed to significantly link the pointwise estimates with any of the three covariates available. This parameter is then set as spatially-invariant for all models.

3.1.2 Fitting the models on the data

The five spatial models defined in Section 2 are fitted on the precipitation dataset described in Section 3.1.1. Motivated by the spatial trend analysis of the GEV parameters, we use the following marginal spatial model for the EGP, the BRP and the ETP:

$$\begin{cases} \mu(s) &= \beta_{\mu,0} + \text{lon}(s)\beta_{\mu,1} + \text{lat}(s)\beta_{\mu,2} + \text{alt}(s)\beta_{\mu,3} , \\ \sigma(s) &= \beta_{\sigma,0} + \text{lon}(s)\beta_{\sigma,1} + \text{lat}(s)\beta_{\sigma,2} + \text{alt}(s)\beta_{\sigma,3} , \\ \xi(s) &= \beta_{\xi,0} . \end{cases} \quad (6)$$



(a) Region of interest \mathcal{S} . (b) Elevation of the meteorological stations.

Figure 1: Map of France with positions of the meteorological stations and the region \mathcal{S} (a) and elevation of each station in \mathcal{S} (b).

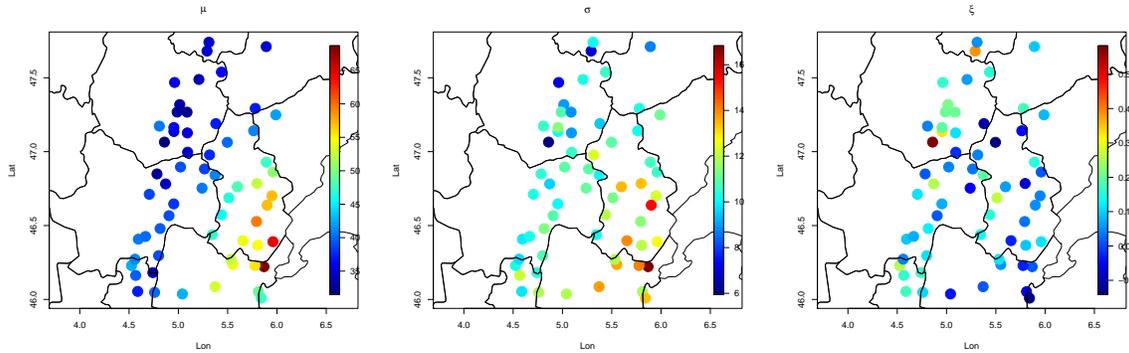


Figure 2: Pointwise estimates of GEV location (right), scale (middle) and shape (left) parameters.

For the LVM and the HKEVP, the same linear combination is taken to which are added the latent zero-mean Gaussian processes $\varepsilon_\mu(\cdot)$ and $\varepsilon_\sigma(\cdot)$, as previously described in (4).

Figure 3 shows the GEV Quantile-Quantile plots (QQ-plots) obtained from the five spatial models at each of the 61 positions. A color has been assigned to each station in order to distinguish roughly the results between one place to another. First, it can be pointed out that the pointwise estimation (upper-left figure) is relevant for every meteorological station, that accredits the GEV assumption. The addition of a spatial structure in the GEV parameters through a spatial model leads to a minor spoiling of the fitting quality, but this is necessary if the goal is to perform spatial extrapolation. The QQ-plots obtained by the three max-stable models are roughly equivalent: this is not surprising since they share the same marginal spatial modeling (6). The use of a hierarchical formulation increases the flexibility of the marginal estimation. The QQ-plots obtained by the LVM and the HKEVP (top middle and top right) show however that the corresponding fits are less accurate, overall on the right part of the distribution. The estimated value of ξ_0 might be an explanation for the second hierarchical model: it equals 0.20 for the HKEVP, while it is approximately equal to 0.09 for the four other ones. See discussions in Section 3.3.2.

The second step of the analysis consists in exploring the spatial bivariate dependence structure.

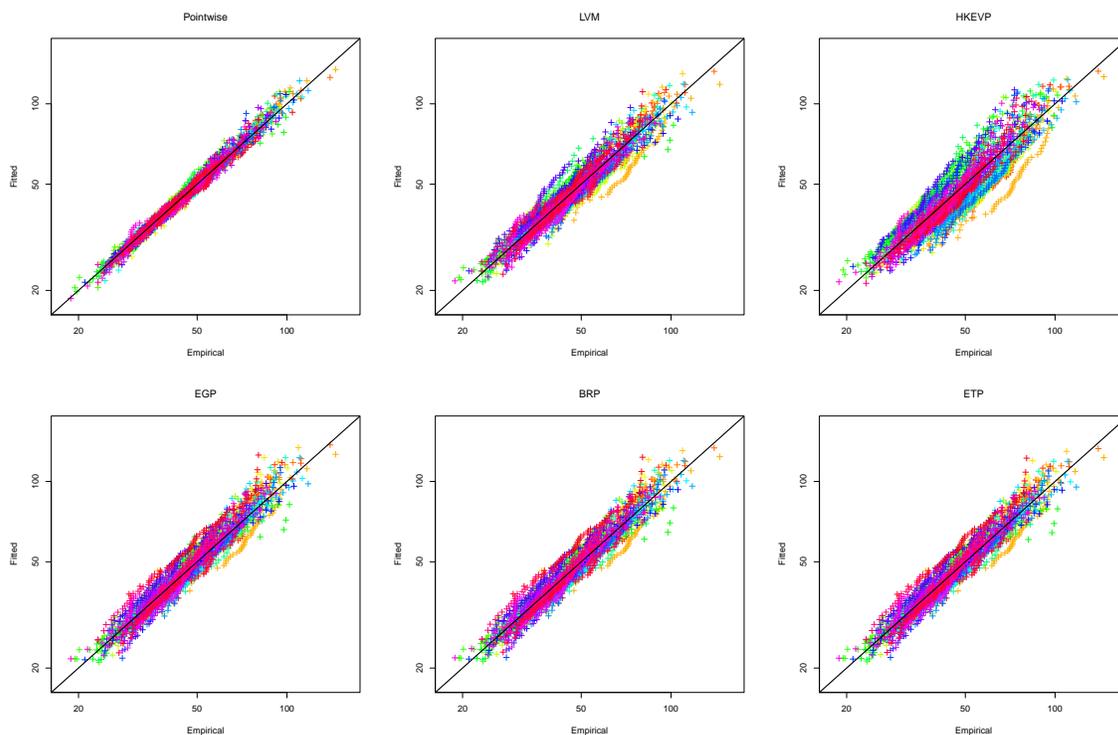


Figure 3: GEV Quantile-Quantile plots of observed data against fitted with pointwise GEV estimation (no spatial structure, top left figure) and each of the five spatial models. Colors are used to distinguish the meteorological stations.

This is done through the estimations of the extremal coefficient between two sites s_1 and s_2 . Recall that if ν is the F -madogram of Z , that is:

$$\nu(s_1, s_2) = \mathbb{E}|F(Z(s_1)) - F(Z(s_2))|,$$

where F denotes the marginal cdf of $Z(\cdot)$. Then the extremal coefficient θ is given for all $s_1, s_2 \in \mathcal{S}$ by:

$$\theta(s_1, s_2) = \frac{1 + 2\nu(s_1, s_2)}{1 - 2\nu(s_1, s_2)}.$$

The non parametric estimator of $\theta(s_1, s_2)$ is thus naturally derived from the madogram-based estimator of Cooley et al. (2006).

Figure 4 plots $\{\hat{\theta}(s_i, s_j)\}_{1 \leq i < j \leq n}$, the empirical estimates of the bivariate extremal coefficient for all pairs of sites. The corresponding extremal coefficients of each spatial model fitted on the annual maxima of precipitation data are also given in the same figure, with different colors to dissociate them. For EGP, BRP and EGP, the isotropic assumption allows to depict the curves of $\hat{\theta}$ in terms of h , while the estimation of θ is drawn for all pairs of sites when fitted by the anisotropic HKEVP (see several red points at each distance value abscissa). Note that the estimated posterior medians of α and τ were used as point estimates to compute the bivariate extremal coefficient for the HKEVP. The LVM has been omitted since the conditional independence assumption implies $\theta \equiv 2$.

One can deduce from Figure 4 that the four models with spatial dependence structure fit the data reasonably well. Two particular features can be commented on this figure. First, the EGP does not reach the limit value $\theta = 2$ (independence), even when the distance h becomes large. This model's drawback has been highlighted in Section 2.2.1. The limit value here is not 1.838 as stated

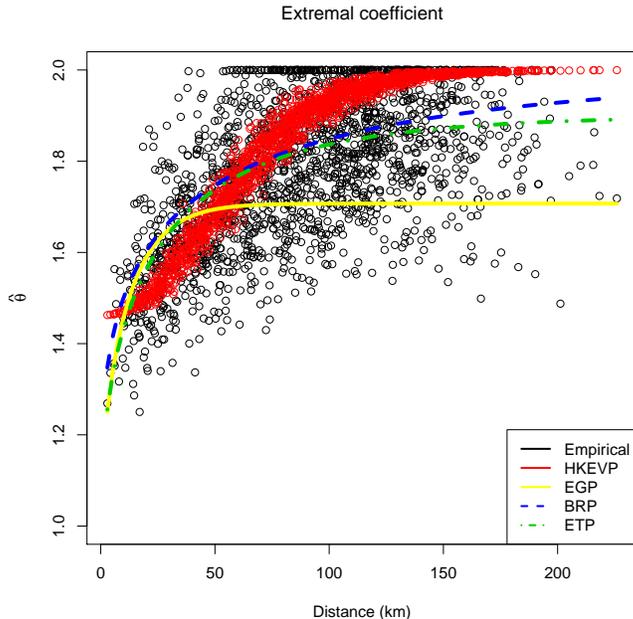


Figure 4: Empirical extremal coefficient $\hat{\theta}$ for the observed precipitation data (in black) and fitted ones for each model $\hat{\theta}_{\text{HKEVP}}$, $\hat{\theta}_{\text{EGP}}$, $\hat{\theta}_{\text{BRP}}$ and $\hat{\theta}_{\text{ETP}}$ (in color).

previously but 1.707, since we consider a powered exponential correlation function $\rho(\cdot)$ that is always positive. Secondly, the impact of the nugget effect can be seen on the HKEVP when h is close to zero: the bivariate extremal coefficient does not tend to 1 as it should, but is lower bounded. Indeed, $\lim_{\|s_1 - s_2\| \rightarrow 0} \theta_{\text{HKEVP}}(s_1, s_2) = 2^\alpha$, so that it equals one only when the nugget α is fading to zero.

Extreme rainfall data are known to exhibit a low range of spatial dependence. For instance, Fawcett and Walshaw (2014) found out that the dependence between extreme precipitation in Great Britain is genuine for a distance $h < 100$ km. In the present dataset, half of the inter-distances are lower than 100 km, and there is still some trace of extreme dependence between 100 km and 200 km.

3.2 Comparative study of the spatial models

Recall that in the previous section, the five models have been fitted to the precipitation data, so that they can now be used as data generators. The goal of this technique, close to the parametric bootstrap, is to work with simulated data so that we can compare estimated criteria with the known true value. Since we do not want to favor one particular model by setting an arbitrary simulation design, we use the fitted parameters of the five spatial models (obtained on the same dataset) to generate artificial data.

Despite the loss of precision in the marginal and the bivariate modelling observed respectively in Figure 3 and 4, we keep all five models and their fitted versions to create artificial data generators. Indeed, the goal is not to perfectly mimic the phenomena, but to work with simulations that are close enough to precipitation to be relevant. The five spatial models will compete through the simulation design presented below.

3.2.1 Simulated data sets

The aim of this section is to present the way to produce simulations of annual maxima of daily precipitation. Samples are drawn from the five models with parameters fixed from Section 3.1.2. Routines `rmaxstab` and `hkevp.rand` are used from packages `SpatialExtremes` and `hkevp` respectively. We denote by:

$$\mathbb{Y}_{\mathcal{G}}^{(k)} := \begin{bmatrix} Y_1^{(k)}(s_1|\mathcal{G}) & \dots & Y_1^{(k)}(s_n|\mathcal{G}) \\ \vdots & \ddots & \vdots \\ Y_{n_y}^{(k)}(s_1|\mathcal{G}) & \dots & Y_{n_y}^{(k)}(s_n|\mathcal{G}) \end{bmatrix}_{n_y \times n},$$

the k -th replicate, for $k \in \{1, \dots, K = 50\}$, coming from the generator model \mathcal{G} . Artificial data are generated on the $n = 61$ positions $\{s_1, \dots, s_n\}$ coinciding with the sites of Figure 1. The number of simulations $n_y = 45$ is the maximal length observed of yearly maxima series per station on this dataset. Equivalent datasets $\mathbb{Z}_{\mathcal{G}}^{(k)}$ are also used: they differ from $\mathbb{Y}_{\mathcal{G}}^{(k)}$ by being unit Fréchet distributed.

3.2.2 Two comparative criteria

The performances of the competing spatial models are measured on the two criteria that are described below.

The first criterion focuses on the quality of spatial extrapolation of the GEV marginal distribution. It consists in the estimation of the T -year return level at a “target” site s^* where no data is available. For any period of time T , this value is a quantile (denoted by $y_T(s^*)$) of the marginal distribution that we expect to be exceeded once over T years. It is therefore defined by $\Pr(Y(s^*) > y_T(s^*)) = 1/T$. Knowing the marginal parameters μ, σ and ξ evaluated at s^* , it is possible to explicitly compute its value:

$$y_T(s^*) = \mu(s^*) + \frac{\sigma(s^*)}{\xi(s^*)} \left[\log \left(\frac{T}{T-1} \right)^{-\xi(s^*)} - 1 \right]. \quad (7)$$

The linear model (6) allows direct extrapolation of marginal parameters at an ungauged site s^* , provided that all spatial covariates associated to s^* are known. Thanks to (7), the computation of the first criterion is straightforward for the EGP, the BRP and the ETP.

For the hierarchical models LVM and HKEVP, the first criterion can be estimated in different ways. We choose the median of the predictive posterior distribution of $y_T(s^*)$, by using a kriging estimator to obtain $\mu(s^*), \sigma(s^*)$ and $\xi(s^*)$ at each MCMC step. One could argue for using conditional simulations instead of kriging since this method looks more appropriate in a Bayesian framework. Here we chose kriging to be consistent with the three other models, that use interpolation to estimate the marginal parameters.

The second criterion we study provides a measure of the spatial bivariate dependence structure of a specific model. For this purpose, we use the bivariate extremal coefficient θ evaluated at a specific pair of sites (s_1, s_2) . This fixed pair is chosen so that the associated distance h is approximately 50 km, a range where all spatial models display some spatial dependence when fitted on the data (see Figure 4).

Apart from summarizing the dependence structure, the bivariate extremal coefficient may be used to compute a joint probability. For example, putting aside the LVM for which the following holds only conditionally to the marginal parameters, the probability that one of the two annual maxima $Y(s_1)$ and $Y(s_2)$ be greater than their respective T -years return levels $y_T(s_1)$ and $y_T(s_2)$ is given by:

$$\Pr \{Y(s_1) > y_T(s_1) \text{ or } Y(s_2) > y_T(s_2)\} = 1 - \exp \left(-\frac{\theta(s_1, s_2)}{z_T} \right),$$

where z_T is the T -years return level of a unit Fréchet random variable.

Note that the value $T = 100$ has been chosen in the comparison study. More generally, one could think about other criteria, that could be driven by specific practical needs. For example, instead of

splitting into two parts the marginal and the dependence behaviors, one could mix these characteristics into an additional criterion.

3.2.3 Estimating the comparative criteria with the five models

The five spatial models are all refitted to each set of artificial data $\mathbb{Y}_{\mathcal{G}}$, and the two criteria defined in Section 3.2.2 are estimated and compared to the true value associated to \mathcal{G} . Let \mathcal{F} denote a fitting model chosen among the five spatial models LVM, HKEVP, EGP, BRP, and ETP.

- As the first criterion involves a spatial extrapolation at an ungauged position, we use a κ -fold cross validation procedure. The 61 sites are randomly split into $\kappa = 6$ classes, represented in Figure 5 with different colors. The models \mathcal{F} are fitted on five classes of sites and the sixth plays the role of the ungauged positions. We obtain estimations of $y_{100}(\mathbf{s}_{\kappa}; \mathcal{G})$, the 100-years return level of $Y(\cdot)$ generated by \mathcal{G} at ungauged positions \mathbf{s}_{κ} of the κ -th class of sites. We repeat this procedure for each class κ , and compute, for each generator \mathcal{G} and each fitting model \mathcal{F} , the MSE:

$$\text{MSE}(\mathcal{F}; \mathcal{G}) = \frac{1}{6} \frac{1}{|\mathbf{s}_{\kappa}|} \sum_{\kappa=1}^6 \sum_{\mathbf{s}_{\kappa}} (y_{100}(\mathbf{s}_{\kappa}; \mathcal{G}) - \widehat{y}_{100}(\mathbf{s}_{\kappa}; \mathcal{F}))^2 .$$

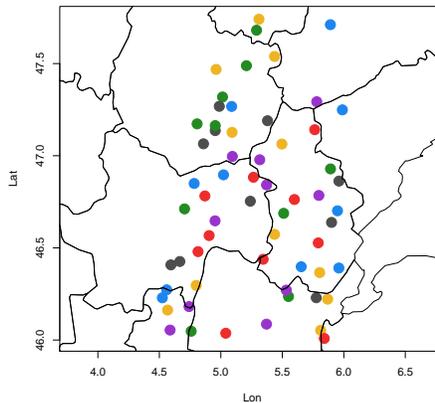


Figure 5: Classes of sites chosen randomly to perform the κ -fold cross validation on first criterion. Each color defines a different class.

- The second criterion, i.e. the bivariate extremal coefficient $\theta(s_1, s_2)$, is estimated by \mathcal{F} on the artificial set $\mathbb{Z}_{\mathcal{G}}$, with unit Fréchet margins. The results obtained are samples of $K = 50$ estimations of $\hat{\theta}_{(s_1, s_2)}(\mathcal{F}; \mathcal{G})$.

3.2.4 Results

The comparative studies are summarized in Figures 6 and 7, which respectively show $\text{MSE}(\mathcal{F}; \mathcal{G})$ and the boxplots of $\hat{\theta}_{(s_1, s_2)}(\mathcal{F}; \mathcal{G})$, for each fitting model \mathcal{F} and all generators \mathcal{G} . In both figures, \mathcal{G} is specified in the legend below each panel (from (a) to (e)) and \mathcal{F} is given below each bar or boxplot, keeping always the same ordering from left to right, namely LVM, HKEVP, EGP, BRP, and ETP. In Figure 7, the dashed red lines correspond to the true values $\hat{\theta}_{(s_1, s_2)}(\mathcal{G})$.

From these two figures, it is possible to roughly order the five spatial extreme value models by their performances over the two criteria defined in Section 3.2.2.

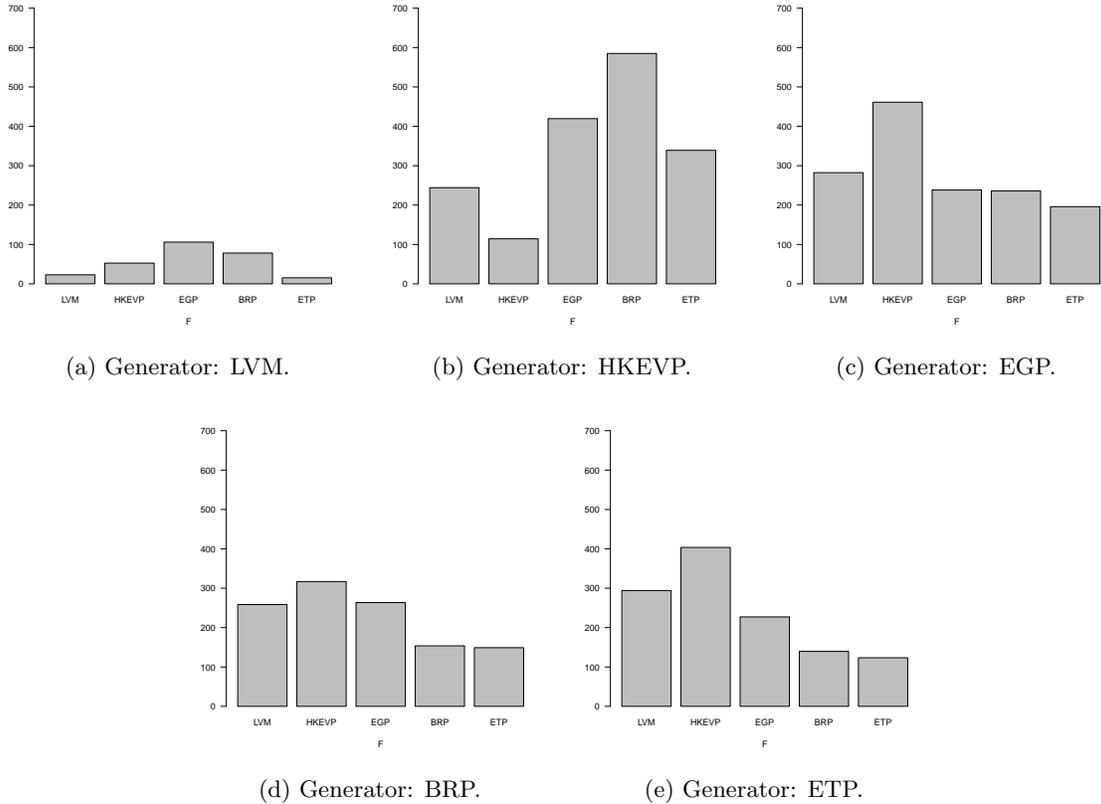


Figure 6: $\text{MSE}(\mathcal{F}; \mathcal{G})$ obtained on criterion 1 for all generators \mathcal{G} and fitting models \mathcal{F} .

The two “best” competitors may be the BRP of Brown and Resnick (1977); Kabluchko et al. (2009) and the ETP of Opitz (2013). Both methods show similar satisfying results in terms of the two criteria of comparison, with an exception on Figure 6b when the generator is the HKEVP. In this case, the MSE obtained on the first criterion is quite large for BRP. The ETP outperforms slightly the BRP in terms of marginal extrapolation, while the BRP seems more accurate regarding the second criterion. In particular (and rather surprisingly), the ETP shows the biggest bias on the extremal coefficient when artificial data come from the ETP itself. Based on these few observed differences, the ETP could be recommended for the spatial extrapolation of the 100-years return level, while the BRP should be preferred for the estimation of the extremal coefficient.

The HKEVP of Reich and Shaby (2012) has the particularity of performing well on both criteria for each of the hierarchical generators. Its main drawback seems to be its poor robustness when fitted on classical max-stable generators. Recall that when fitted on real data, the HKEVP is also the most atypic, with a shape parameter ξ that is twice the value of the other generators.

The LVM of Davison et al. (2012) assumes conditional independence, so that its performances on Figure 7 can not be discussed. Indeed, it fails detecting any trace of spatial dependence, and the extremal coefficient estimate reduces always to 2. Concerning the extrapolation of the return level, the LVM shows the best robustness and appears quite reliable, even if outperformed by the ETP and the BRP (see Figures 6d and 6e).

The EGP of Schlather (2002) fits in a rather satisfying way all the generators in terms of the first criterion. It is nonetheless never the best on Figure 6. As for the second criterion, Figure 7 illustrates the EGP’s drawback of being unable to model independence (see Section 2.2.1). On the one hand, as long as the true bivariate extremal coefficient is lower than the possible limit 1.707, the results are

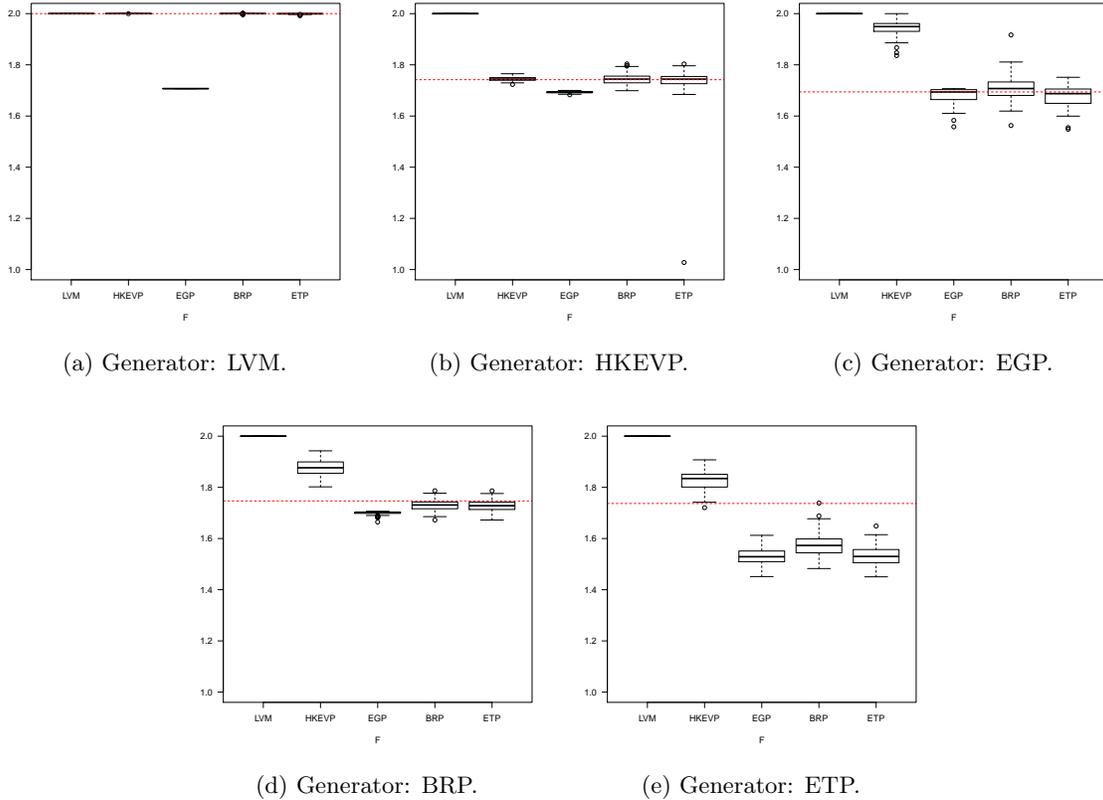


Figure 7: Estimations $\hat{\theta}_{(s_1, s_2)}(\mathcal{F}; \mathcal{G})$ of criterion 2 for all generators \mathcal{G} and fitting models \mathcal{F} .

rather similar to those of the ETP and the BRP (see e.g. Figure 7e). On the other hand, if the true value of θ is higher than 1.707, (Figures 7a, 7b and 7d), the EGP is degenerated toward its limit value. Recall that this ranking between EGP and ETP was expected, since the ETP generalizes the EGP.

Finally, note that these comments do not depend on the choice $T = 100$ in the definition of the two criteria. Other values of T lead to similar patterns of plots of Figure 6: Changing the return period only affects the amplitude of the resulting MSE. Extrapolating our conclusions to other distances between the locations s_1 and s_2 remains cumbersome. We did however a few trials from small to large distances between the two locations, and the results look quite similar to those presented on Figure 7. Besides, we also investigated adding a nugget effect in the max-stable models, but this yielded poorer performances.

3.3 Beyond the scope of this study: some additional comparisons

In the previous section, a comparative study has been driven between five spatial models based on two risk measures. The first one assesses the marginal behavior through the extrapolation of a 100-years return level, whereas the second evaluates the dependence structure with the bivariate extremal coefficient. In this section, we discuss other properties of the five spatial models that have not been fully taken into consideration yet. The aim is to guide the practitioner for a better choice of model, depending on the pursued goal.

3.3.1 Bayesian models

Two Bayesian models have been used: LVM and HKEVP. The results obtained when fitting these two models are Markov chains that, if convergence is assumed, represent a posterior distribution for each parameter (or for a function of parameters as the 100-years return level). To be feasible, the competition needed to focus on point estimates of the two criteria. The median of the posterior distributions has been chosen for the two hierarchical models. This choice looks rather restrictive for such models that can provide much more information on a given parameter. One could for example make use of conditional simulations to access the return level of the predictive distribution instead of computing its posterior median. One could also obtain an estimation of the uncertainty associated to the point estimates, which can be of great interest in most situations.

Bayesian models present however two drawbacks at the inference step:

- First, they have generally more input parameters than the non-Bayesian ones. For instance, the user of the function `latent` in `SpatialExtremes` must provide values that control initial steps, prior distributions, random walk jumps for candidate generation and, above all, the number of iterations to use in the algorithm for assessing convergence (burn-in period) and to provide a satisfactory sample of the posterior distribution. For functions `hkevp.fit` and `latent.fit`, default values are available for these arguments, though they remain to be defined with care.
- Secondly, these models are more time consuming to fit, especially when the desired length of the posterior samples is high. In our case, one estimation with this model on the real or artificial dataset takes around 30 minutes and one hour respectively with LVM and HKEVP. The model of Reich and Shaby (2012) is thus the heaviest in this case and the computational time increases with the number of sites n and the number of knots L .

Because of the time needed to get estimations from the LVM and the HKEVP, parallel computing is strongly advised to perform the comparative study of Section 3. We worked with R package `parallel` (Ripley et al., 2015), which uses the random number generator of L'Ecuyer et al. (2002) to assess for independence between worker processors.

As a short concluding remark, one could note that even though only the two hierarchical models LVM and HKEVP are genuinely Bayesian, some approaches exist to embed max-stable models in a Bayesian framework, cf. for instance Ribatet et al. (2012) or Thibaud et al. (2016).

3.3.2 More about the HKEVP

A few drawbacks can be pointed out for the HKEVP. They are listed here:

- As discussed by Castruccio et al. (2015), realizations from the HKEVP are dependent on the choice of spatial knots that define the kernels in the dependence structure. When fitting data with the HKEVP, the choice of knots has to be done carefully and has to be seen as a trade-off between efficient estimation and computational burden. Indeed, all values of the random effect are updated at each MCMC step, which represents $L \times T$ parameters. The impact of a misspecification of the knots set has been studied in Reich and Shaby (2012). The conclusions can be summarized saying that too few knots may lead to a larger bias in the estimation of the GEV parameters, while considering too many knots than necessary does not improve significantly these estimations but increases the computational time drastically.
- A second drawback is the non-mixing property of the model. Indeed, even with a great number of iterations and with a large burn-in period, a substantial thinning procedure is required to obtain stationary, non-correlated resulting Markov chains. Otherwise, the chains (in particular the ones of α and τ) show traces of dependence. In the comparison study, this aspect has been ignored because only point estimates were needed.

- Another drawback appears when the exact value of α is near 0, case where the spatial process is very smooth. In this case, the Markov chains are evolving very slowly due to the fact that the values of the random effect A are uniformly distributed over $\mathbb{R}_+ \setminus \{0\}$. However, it has to be noted that this feature appears when the observed process is perfectly smooth (as for the Gaussian extreme value model (Smith, 1990)). There is no reason to observe this singularity for annual maxima of a natural phenomenon such as precipitation.

Despite these shortcomings, the HKEVP model has two main advantages. The first one is that conditioning on the random effect A , we obtain independent responses in the hierarchical formulation. This allows us to use the full likelihood of the process rather than the composite likelihood. The second advantage is that its exponent function V is explicit for any set of sites. Comparatively, the exponent functions for the EGP, the BRP and the ETP are computed with multivariate Gaussian or Student cdf and thus are only explicit when evaluated at pairs of sites. If the goal is the evaluation of a multivariate probability on the spatial process $Y(\cdot)$, approximations must be made for this model (Genz and Bretz, 2009), which increases the computational cost. This is particularly the case for the ETP (see Thibaud and Opitz (2015) in a peaks-over-threshold approach).

3.3.3 Summary of the comparison

As a summary, Table 1 provides a visual assessment of the five models over the two criteria and the points discussed above.

	LVM	HKEVP	EGP	BRP	ETP
Marginal extrapolation	≈	≈	≈	✓	✓
Joint probability	✗	≈	✗	✓	✓
Bayesian approach	✓	✓	✗	✗	✗
Fast program	≈	✗	✓	✓	≈
Explicit multivariate cdf	✓	✓	✗	✗	✗

Table 1: Sketch of the characteristics of the five spatial models. A check ✓ (resp. a cross ✗) means that the model is performant (resp. not advised) for the corresponding criterion or satisfies (resp. not) the given characteristic. In some cases where the decision is difficult to make, a "≈" symbol is given.

4 Discussion and conclusion

4.1 General conclusions

In this article, five models for spatial extreme values are competing over two risk measures that represent usual interest in application: the extrapolation of a 100-years return level at an ungauged site and the estimation of the bivariate extremal coefficient. This paper can be regarded as a practical guide when fitting annual maxima of precipitation data. This work contrasts with existing comparisons of spatial extreme value methods as in Davison et al. (2012) or Davison et al. (2013) in two ways: on the one hand, it is the first study, as far as we know, that includes the hierarchical model of Reich and Shaby (2012); on the other hand, the use of simulated data that are tailored on real precipitation data enables the comparison to be more objective.

Results from Section 3.2.4 show dissimilarities between models and tend to discard some of them, depending on what is the main objective. Firstly, if the interest lies in the estimation of the marginal effect, one should prefer the ETP of Opitz (2013) or the BRP of Brown and Resnick (1977). One can besides keep in mind that the LVM of Davison et al. (2012) has some robust performances, even if slightly outperformed by the ETP and the BRP. Secondly, if the goal involves the modeling of the

joint dependence structure (e.g. a joint probability), the BRP or the ETP seem the best choices but the HKEVP of Reich and Shaby (2012) may also provide reliable estimates.

It is important to note that this comparative study has been made under circumstances that may influence the general conclusions. Namely, the whole simulation design was driven from the set of precipitation data using a parametric bootstrap. This procedure has been chosen for seek of objectivity. If this competition were realized with different data, the conclusions may have been slightly different. For instance, a previous version of the present article (Sebillé et al., 2016) was based on precipitation data that were sparsely located over France, thus exhibiting spatial asymptotic independence. In this case, unsurprisingly, the LVM of Davison et al. (2012) was always the best model in terms of the extrapolation of the 100-years return level, while the HKEVP was described as a good compromise towards both the marginal extrapolation and the estimation of the dependence structure.

As discussed in Section 3.3, the comparison between these spatial models may also be generally more complex because of several features that characterize each of them. For instance, the HKEVP is the only one which can give an explicit formulation of the dependence structure for an arbitrary set of sites. This allows in particular conditional sampling of the yearly maxima process, though conditional simulation on max-stable processes can also be performed, see Dombry et al. (2013) for instance. However, its inference via MCMC is less easy to handle, it involves arbitrary choices like the positions of the knots, and it demands more computational resources than the other four models to be properly fitted.

4.2 Return level maps with the best models

Using the conclusions of this comparison study, we now provide an interpolation map of the 100-years return level of precipitation over the studied region. To this end, we use a regular grid of positions that covers the region of interest \mathcal{S} . Elevation for each point is represented in Figure 8 with a color code. A

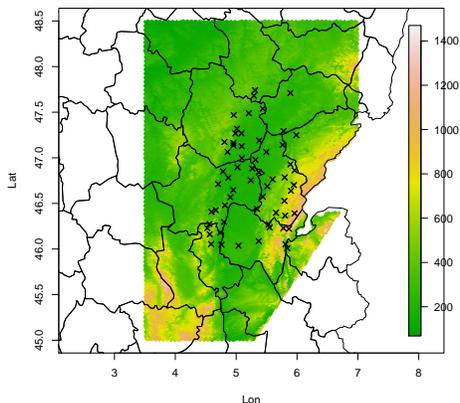


Figure 8: Elevation map of the central-east of France that covers the studied region \mathcal{S} where data are located.

part of the grid that corresponds to the Northern French Alps has been deliberately truncated because altitudes are much higher, making the prediction of the 100-years return level highly uncertain in this area. The effect on the interpolation map with this mountainous region was that only differences between plains and mountains were visible and variations in the whole regions were not illustrated well.

The ETP and the LVM are two satisfying ways to extrapolate the marginal parameters, coming furthermore from rather different constructions of spatial models for extremes. Hence, we use them to

produce the map of 100-years return level. Figure 9a (resp. 9b) displays the extrapolated map with the ETP (resp. LVM). Standard errors are shown on Figures 9c and 9d for the ETP and the LVM respectively. For LVM, it corresponds to the posterior standard deviation. The way of computing these errors for ETP is explained in B.

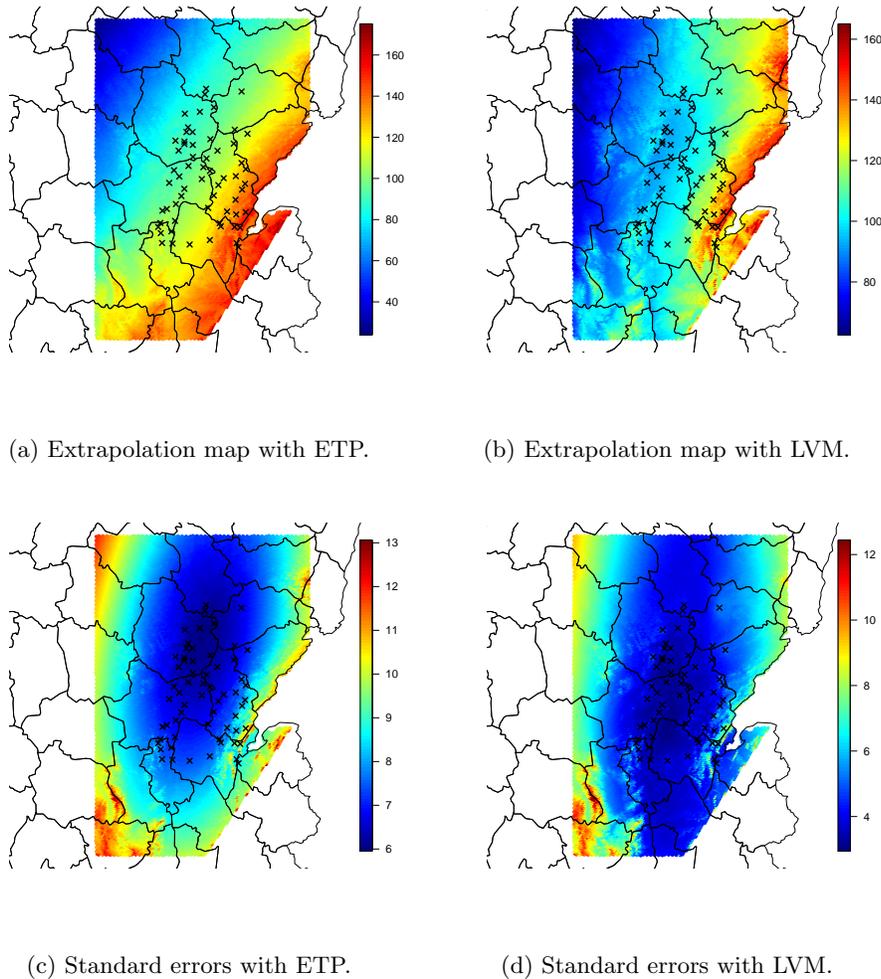


Figure 9: Maps of the extrapolated 100-years return level of precipitation and associated standard errors obtained with the ETP and the LVM.

The maps of Figures 9a and 9b show some dissimilarities between the extrapolation from the two spatial models. With ETP, the 100-years return level depends only on the spatial covariates which are the longitude, the latitude and the altitude. As a result, it increases with altitude and along the South-East direction. In the case of the LVM, the 100-years return level is affected by the variability in the latent Gaussian processes. The map of return levels show an increase towards the East direction and seems more correlated to the altitude covariate.

As expected, the more the prediction is made outside the convex hull of the stations, the greater the prediction error. Another feature that increases uncertainty in marginal extrapolation is the altitude: it can be seen in regions like the frontier with Switzerland or the upper-right part of the Auvergne massif (lower-left part of the region) where altitude is higher than 1.000 meters. Finally, we can see that the error obtained over the region \mathcal{S} is slightly lower for the LVM (between 4 mm and 6 mm over

most of the region) than for the ETP (between 6 mm and 10 mm over the same domain).

Acknowledgements

We would like to thank many contributors who have been helpful and supportive. First of all, we thank the authors of R packages `SpatialExtremes` and `RandomFields`, in particular Mathieu Ribatet for his very explicit functions around the EGP, the ETP and the BRP and the great help he provided us.

Secondly, we thank Aurélien Ribes from Météo France and Benjamin Renard from IRSTEA Lyon, who provided us useful tools to extract covariates information from the precipitation data set.

We also thank a lot Brian Reich and Benjamin Shaby for their model (HKEVP) and the indications they have given, allowing moreover the implementation of the R package `hkevp`.

This paper has been written during the PhD thesis of the first author. His thesis has been financed by EDF. We would like to thank warmly several researchers from EDF R&D, namely Anne Dutfoy, Marie Gallois, Thi Thu Huong Hoang, Sylvie Parey and Nicolas Bousquet for numerous fruitful discussions that improved substantially this work. The three authors were also partially supported by the french national programme LEFE/INSU.

We are grateful to Peter Craigmile who made numerous suggestions for improving the text, and to the two referees who provided many suggestions and helped to ameliorate the manuscript. Finally, we thank the R project which provides free material, and Météo France and EDF/DTG for allowing us to work freely on their data.

A Upper bound of θ for the EGP

By following (Matérn, 1986, p. 16), the correlation function $\rho(\cdot)$ of a Gaussian process satisfies:

$$\rho(h) \geq \inf_h \Lambda_{\frac{n-2}{2}}(h) ,$$

with $\Lambda_k(h) = \Gamma(k+1)(2/h)^k J_k(h)$ and $J_k(\cdot)$ being the Bessel function of the first kind.

In our case $h \in \mathbb{R}^2$, so the correlation function satisfies $\rho(h) > -0.403$. By expression

$$\theta_{\text{EGP}}(h) = 1 + \sqrt{\frac{1 - \rho(h)}{2}}$$

of the EGP extremal coefficient, the inequality $\theta_{\text{EGP}}(h) < 1.838$ is directly obtained. Recall that since we choose a powered exponential form for $\rho(\cdot)$ in the comparison, which is strictly positive, then the upper bound of θ_{EGP} becomes 1.707 in this case.

B Computation of standard errors

In Section 4.2, return levels have been computed for a set of ungauged locations s^* , with associated standard errors. This appendix describes how to obtain them from the ETP.

The estimator $\hat{\Psi}$ of the ETP parameters $\Psi = (\psi_{\text{GEV}}, \psi_V)$ is obtained by maximizing the pairwise likelihood (see Section 2.4) and is asymptotically normal (Ribatet, 2013):

$$\hat{\Psi} \sim \mathcal{N}(\Psi_0, H^{-1}(\hat{\Psi})J(\hat{\Psi})H^{-1}(\hat{\Psi})) ,$$

with $H(\Psi) = \mathbb{E}[\nabla^2 L_{\text{pairwise}}(y; \Psi)]$ the Hessian matrix and $J = \text{Var}(\nabla L_{\text{pairwise}}(y; \Psi))$ the variance score.

To compute standard errors for the 100-years return level extrapolated at s^* with ETP, we first need to get standard errors for $\hat{\Psi}$ by estimating $H(\hat{\Psi})$ and $J(\hat{\Psi})$:

- $\hat{H}(\hat{\Psi}) = \nabla^2 L_{\text{pairwise}}(y; \hat{\Psi})$ is obtained straightforwardly by evaluation of the Hessian matrix at $\hat{\Psi}$,
- $\hat{J}(\hat{\Psi}) = \sum_{t=1}^T \nabla L_{\text{pairwise}}(y_t; \hat{\Psi}) \nabla L_{\text{pairwise}}(y_t; \hat{\Psi})'$ (Varin and Vidoni, 2005).

The 100-years return level at any position s^* can be estimated by (7), which in terms of $\hat{\Psi}$ and a function h can be written $\hat{y}_{100}(s^*) := h(\hat{\Psi}, s^*)$. Standard errors associated to $\hat{y}_{100}(s^*)$ are then obtained using the Delta method:

$$\text{Var}(\hat{y}_{100}(s^*)) = \nabla h(\hat{\Psi}; s^*) \text{Var}(\hat{\Psi}) \nabla h(\hat{\Psi}; s^*)' .$$

References

References

- Apputhurai, P., Stephenson, A. G., 2013. Spatiotemporal hierarchical modelling of extreme precipitation in western australia using anisotropic gaussian random fields. *Environmental and ecological statistics* 20 (4), 667–677.
- Banerjee, S., Carlin, B., Gelfand, A., 2004. Hierarchical modeling and analysis for spatial data. Monographs on statistics and applied probability. Chapman & Hall.
- Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., 2004. *Statistics of Extremes: Theory and Applications*. Vol. 558. John Wiley & Sons.
- Brown, B. M., Resnick, S. I., 1977. Extreme values of independent stochastic processes. *J. Appl. Probability* 14 (4), 732–739.
- Castruccio, S., Huser, R., Genton, M. G., 2015. High-order composite likelihood inference for max-stable distributions and processes. *Journal of Computational and Graphical Statistics* (just-accepted), 1–32.
- Coles, S. G., 2001. *An introduction to statistical modeling of extreme values*. Springer Series in Statistics. Springer-Verlag London Ltd., London.
- Cooley, D., Cisewski, J., Erhardt, R. J., Jeon, S., Mannshardt, E., Omolo, B. O., Sun, Y., 2012. A survey of spatial extremes: Measuring spatial dependence and modeling spatial effects. *Revstat* 10 (1), 135–165.
- Cooley, D., Naveau, P., Poncet, P., 2006. Variograms for spatial max-stable random fields. In: *Dependence in probability and statistics*. Springer, pp. 373–390.
- Davison, A. C., Huser, R., Thibaud, E., 2013. Geostatistics of dependent and asymptotically independent extremes. *Mathematical Geosciences* 45 (5), 511–529.
- Davison, A. C., Padoan, S. A., Ribatet, M., May 2012. Statistical Modeling of Spatial Extremes. *Statistical Science* 27 (2), 161–186.
- de Haan, L., 1984. A spectral representation for max-stable processes. *The annals of probability*, 1194–1204.
- de Haan, L., Ferreira, A., 2006. *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering. New York, NY: Springer.
- Dombry, C., Éyi-Minko, F., Ribatet, M., 2013. Conditional simulation of max-stable processes. *Biometrika* 100 (1), 111–124.

- Dyrrdal, A. V., Lenkoski, A., Thorarinsdottir, T. L., Stordal, F., 2015. Bayesian hierarchical modeling of extreme hourly precipitation in norway. *Environmetrics* 26 (2), 89–106.
- Fawcett, L., Walshaw, D., 2014. Estimating the probability of simultaneous rainfall extremes within a region: a spatial approach. *Journal of Applied Statistics* 41 (5), 959–976.
- Finkenstädt, B., Rootzén, H., 2004. *Extreme values in finance, telecommunications and the environment*. Chapman & Hall/CRC, Boca Raton.
- Genz, A., Bretz, F., 2009. *Computation of multivariate normal and t probabilities*. Vol. 195. Springer Science & Business Media.
- Gilleland, E., Katz, R. W., 2011. **extRemes**: New software to analyze how extremes change over time. R package.
- Kabluchko, Z., Schlather, M., De Haan, L., 2009. Stationary max-stable fields associated to negative definite functions. *The Annals of Probability*, 2042–2065.
- L’Ecuyer, P., Simard, R., Chen, E. J., Kelton, W. D., 2002. An object-oriented random-number package with many long streams and substreams. *Operations research* 50 (6), 1073–1075.
- Lindsay, B. G., 1988. Composite likelihood methods. *Contemporary mathematics* 80 (1), 221–39.
- Matérn, B., 1986. *Spatial variation*, vol. 36 of. *Lecture Notes in Statistics* 2.
- Opitz, T., 2013. Extremal t processes: Elliptical domain of attraction and a spectral representation. *Journal of Multivariate Analysis* 122, 409–413.
- Padoan, S. A., Bevilacqua, M., 2015. *Analysis of Random Fields Using CompRandFld*. URL <http://www.jstatsoft.org/v63/i09/>
- Padoan, S. A., Ribatet, M., Sisson, S. A., 2010. Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association* 105 (489), 263–277.
- Penot, D., 2014. *Cartographie des événements hydrologiques extrêmes et estimation schadex en sites non jaugés*. Ph.D. thesis, Hydrologie. Université Grenoble Alpes, 2014.
- Reich, B. J., Shaby, B. A., 2012. A hierarchical max-stable spatial model for extreme precipitation. *The annals of applied statistics* 6 (4), 1430.
- Reich, B. J., Shaby, B. A., Cooley, D., 2014. A hierarchical model for serially-dependent extremes: A study of heat waves in the western us. *Journal of Agricultural, Biological, and Environmental Statistics* 19 (1), 119–135.
- Ribatet, M., 2013. Spatial extremes: Max-stable processes at work. *Journal de la Société Française de Statistique* 154 (2), 156–177.
- Ribatet, M., 2015. **SpatialExtremes**: Modelling Spatial Extremes. R package version 2.0-2.
- Ribatet, M., Cooley, D., Davison, A. C., 2012. Bayesian inference from composite likelihoods, with an application to spatial extremes. *Statistica Sinica*, 813–845.
- Ripley, B., Tierney, L., Urbanek, S., 2015. **parallel**. R package version 3.3.1.
- Schlather, M., 2002. Models for stationary max-stable random fields. *Extremes* 5 (1), 33–44.

- Schlather, M., Malinowski, A., Oesting, M., Boecker, D., Stokorb, K., Engelke, S., Martini, J., Ballani, F., Moreva, O., Menck, P. J., Gross, S., Ober, U., Christoph Berreth, Burmeister, K., Manitz, J., Morena, O., Ribeiro, P., Singleton, R., Pfaff, B., R Core Team, 2016. RandomFields: Simulation and Analysis of Random Fields. R package version 3.1.12.
URL <http://CRAN.R-project.org/package=RandomFields>
- Schlather, M., Tawn, J. A., 2003. A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika* 90 (1), 139–156.
- Sebillé, Q., 2016. `hkevp`: A hierarchical model for Spatial Extremes. R package version 1.0.
- Sebillé, Q., Fougères, A.-L., Mercadier, C., 2016. A comparison of spatial extreme value models: application to precipitation data. Unpublished: <https://hal.archives-ouvertes.fr/hal-01300751>.
- Shaby, B. A., Reich, B. J., 2012. Bayesian spatial extreme value analysis to assess the changing risk of concurrent high temperatures across large portions of European cropland. *Environmetrics* 23 (8), 638–648.
- Smith, R. L., 1990. Max-stable processes and spatial extremes. Preprint, Department of Mathematics, University of Surrey, Guilford.
- Stephenson, A. G., 2009. High-dimensional parametric modelling of multivariate extreme events. *Australian & New Zealand Journal of Statistics* 51 (1), 77–88.
- Stephenson, A. G., Shaby, B. A., Reich, B. J., Sullivan, A. L., 2015. Estimating spatially varying severity thresholds of a forest fire danger rating system using max-stable extreme-event modeling. *Journal of Applied Meteorology and Climatology* 54 (2), 395–407.
- Thibaud, E., Aalto, J., Cooley, D. S., Davison, A. C., 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.
- Thibaud, E., Opitz, T., 2015. Efficient inference and simulation for elliptical pareto processes. *Biometrika* 102 (4), 855–870.
- Varin, C., Vidoni, P., 2005. A note on composite likelihood inference and model selection. *Biometrika* 92 (3), 519–528.