Continuous mixtures of Gaussian processes as models for spatial extremes

Lorenzo Dell'Oro[a,b,c], Carlo Gaetan[d], Thomas Opitz[e]

[a] *IMAG, Université de Montpellier, Montpellier, France*
[b] *LEMON, Inria, Montpellier, France*
[c] *LJAD, Université Côte d'Azur, Nice, France*
[d] *DAIS, Università Ca' Foscari di Venezia, Venice, Italy*
[e] *BioSP, INRAE, Avignon, France*

Abstract

Spatial modelling of extreme values allows studying the risk of joint occurrence of extreme events at different locations and is of significant interest in climatic and other environmental sciences. A popular class of dependence models for spatial extremes is that of random location-scale mixtures, in which a spatial "baseline" process is multiplied or shifted by a random variable, potentially altering its extremal dependence behaviour. Gaussian location-scale mixtures retain benefits of their Gaussian baseline processes while overcoming some of their limitations, such as symmetry, light tails and weak tail dependence. We review properties of Gaussian location-scale mixtures and develop novel constructions with interesting features, together with a general algorithm for conditional simulation from these models. We leverage their flexibility to propose extended extreme-value models, that allow for appropriately modelling not only the tails but also the bulk of the data. This is important in many applications and avoids the need to explicitly select the events considered as extreme. We propose new solutions for likelihood inference in parametric models of Gaussian location-scale mixtures, in order to avoid the numerical bottleneck given by the latent location and scale variables that can lead to high computational cost of standard likelihood evaluations. The effectiveness of the models and of the inference methods is confirmed with simulated data examples, and we present an application to wildfire-related weather variables in Portugal. Although not detailed here, the approaches would also be straightforward to use for modelling multivariate (non spatial) data.

Keywords: Asymptotic dependence, Conditional simulation, Copula, Location mixture, Parameter estimation, Scale mixture

1. Introduction

Gaussian processes play a central role in spatial and spatio-temporal statistics by providing a flexible and tractable framework for modelling dependence in discrete and continuous domains. Extreme-value processes, by contrast, arise as asymptotic limits of suitably normalised pointwise maxima or threshold exceedances of stochastic processes and underpin modern theory for spatial extremes through max-stable and generalized Pareto processes (Brown and Resnick, 1977; de Haan, 1984; Ferreira and de Haan, 2014). The connection between these two classes of models is both theoretical and practical. Large families of max-stable processes can be constructed as limits of transformed Gaussian processes, for example via spectral representations in which exponentiated or scaled Gaussian processes serve as building blocks for max-stable dependence structures (Smith,

1990; Kabluchko et al., 2009; Opitz, 2013). These constructions establish a direct link between Gaussian dependence and asymptotic extreme-value behaviour.

Despite these theoretical strengths, growing empirical and practical evidence suggests that max-stable processes are often ill-suited to the demands of modern data analysis especially in the environmental context. Their asymptotic nature implies that they are primarily designed for modelling block maxima or exceedances above very high thresholds, whereas many contemporary applications require inference and stochastic simulation also at moderate levels and across a wide range of the distribution. Moreover, max-stable models typically impose strong forms of tail dependence that may not be supported by observed data, and likelihood-based inference is computationally challenging, often relying on composite likelihoods that become inefficient as the number of spatial locations increases.

Recent work has therefore argued for a shift away from max-stable processes as default modelling tools for spatial extremes especially for environmental data (Huser et al., 2025). Alternative approaches aim to retain the key insights of extreme-value theory while relaxing rigid asymptotic assumptions and improving computational scalability. These include hybrid frameworks mixing max-stability with other dependence types (Wadsworth and Tawn, 2012; Bacro et al., 2016), generalized Pareto processes for threshold exceedances (de Fondeville and Davison, 2018), and random-scale constructions. The latter retain a structure similar to asymptotic models but capture tail dependence more flexibly, including cases that are degenerate in asymptotic frameworks since spatial dependence gradually weakens and ultimately vanishes when events become more extreme (Opitz, 2016; Huser et al., 2017; Huser and Wadsworth, 2019).

Classical finite mixtures of multivariate Gaussian distributions provide a powerful framework for density approximation and clustering (McLachlan and Peel, 2000), while more structured constructions such as scale mixtures, location mixtures, and mean–variance mixtures naturally accommodate heavy tails and heteroskedasticity (Andrews and Mallows, 1974; Barndorff-Nielsen, 1977). Further extensions incorporating skewness have been developed through skew-normal and skew-t formulations, offering additional flexibility for asymmetric data (Azzalini and Dalla Valle, 1996; Azzalini and Capitanio, 2014); a unified framework is presented in Arellano-Valle and Azzalini (2021). In spatial and spatio-temporal settings, Gaussian mixture and mixture-type models have been shown to effectively capture nonstationarity and complex dependence patterns beyond classical Gaussian processes (Fuentes, 2002; Reich et al., 2012).

From an extreme-value perspective, location and scale mixtures of Gaussian processes (Huser et al., 2017, Krupskii et al., 2018, including t-type processes, Morris et al., 2017, and mean–variance mixtures, Zhang et al., 2022b) form a bridge between classical Gaussian dependence and asymptotic extreme-value behaviour and offer a complementary mechanism for tail thickening by introducing latent random scaling that induces heavy-tailed behaviour and more realistic clustering of large values. Another recent approach to scale mixtures of spatial processes with asymmetry in the tail dependence is made by Krupskii (2026).

Despite their appealing modelling properties, the practical deployment of such rich mixture-of-Gaussian models is often limited by computational considerations. Likelihood-based inference typically involves high-dimensional vectors of latent variables and can lead to multimodal likelihood surfaces and expensive numerical integration. This challenge is exacerbated in spatial or multivariate contexts, where dependence structures further increase computational complexity. As a result, there is a strong need for computationally feasible inferential strategies, including efficient EM-type algorithms, Bayesian hierarchical formulations with scalable Markov chain Monte Carlo schemes,

or approximate methods such as variational inference (Dempster et al., 1977; Frühwirth-Schnatter, 2006; Blei et al., 2017).

When direct likelihood maximisation is infeasible, two main inference approaches are typically adopted for Gaussian location and scale mixtures: i) EM algorithms, which have recently been reviewed by Lee and McLachlan (2021). However, they need to be adapted to the specific distributions of the random scale or location variables, and have not yet been studied in the general case in which both variables are present. Moreover, they can usually only be applied to the original marginal scales, and not to copula models; ii) Bayesian methods such as that proposed by Zhang et al. (2022a) in the context of censored inference on the tails of a scale mixture of a transformed Gaussian process. Hence, developing inference procedures that preserve model flexibility while remaining computationally tractable is a key methodological objective that we try to address with this work.

This paper has three main objectives. First, in Section 3 we review the key theoretical and practical properties of Gaussian location-scale mixture models and introduce new constructions with enhanced tail flexibility and rich dependence structures. We also develop a general algorithm for conditional simulation from these models, detailed in Section 4.

Second, we address the computational challenge of likelihood-based inference for Gaussian location-scale mixtures. In particular, we propose a new two-step inferential method that exploits two different data transformations: in the first step, we use multivariate information to estimate the dependence parameter of the baseline Gaussian process; in the second step, we use univariate information to estimate the parameter of the random scale and location variables. Details are provided in Section 5. This strategy alleviates the numerical burden induced by the latent variables, enabling the scalable and practical implementation of this class of models. The new inferential method is evaluated on simulated data in Section 6.

Finally, we leverage this flexibility to develop extended extreme-value models that describe both the bulk and the extremes of the distribution, moving beyond purely extreme-centric frameworks. This is important for analysing compound events, where not all data contributing to an extreme aggregate are extreme individually, and for stochastic generation of complete datasets used as inputs in impact models. We illustrate the models by analysing fire weather data in Portugal, using daily data of the Fire Weather Index available for more than 20 years at around 500 pixels, see Section 7.

2. Notation and definitions

We first introduce various mathematical objects used repeatedly throughout this paper. Whenever a mathematical operation involves a vector, say $\mathbf{a}$, and a scalar, say b, such as $\mathbf{a} + b$ or $\mathbf{a}/b$, the operation must be interpreted component-wise.

2.1. Univariate distributions

Let X a random variable with cumulative distribution function (cdf) F, and survival function $\overline{F}$. In the special case that X is a standard Gaussian or Student random variable with degrees of freedom equal to ν, F is indicated by either a Φ or a T_ν. The density function of a d-dimensional multivariate Gaussian with mean vector $\mathbf{0}$ and variance-covariance matrix $\mathbf{\Sigma}$ is denoted as $\phi_d(\cdot; \mathbf{\Sigma})$, while $\phi := \phi_1$ refers to the univariate standard Gaussian density.

The notation $X \sim \mathrm{E}(\lambda)$ means that X follows an exponential distribution with rate parameter λ, while $X \sim \mathrm{AL}(\lambda_1, \lambda_2)$ means that X follows an asymmetric Laplace distribution (Kotz et al., 2001; Gong and Huser, 2022a), with cdf

$$F_X(x;\, \lambda_1, \lambda_2) = \begin{cases} 1 - \lambda_2(\lambda_1 + \lambda_2)^{-1} \exp(-\lambda_1 x), & \text{if } x \geq 0 \\ \lambda_1(\lambda_1 + \lambda_2)^{-1} \exp(\lambda_2 x), & \text{if } x < 0. \end{cases} \tag{1}$$

In the sequel, we exploit the fact that if X_i, $i = 1, 2$, are two independent exponential random variables with rate λ_i then $X_1 - X_2 \sim \mathrm{AL}(\lambda_1, \lambda_2)$.

A cdf F with upper endpoint $x^* = \infty$ is said to be *regularly varying* with index $\alpha \geq 0$ if $\bar{F}(tx)/\bar{F}(x) \to x^\alpha$ as $t \to \infty$, for any $x > 0$.

The random variable X with cdf F_X is heavier-tailed than the random variable Y with cdf F_Y if

$$\lim_{t \to \infty} \frac{\overline{F}_X(t)}{\overline{F}_Y(t)} = \infty.$$

2.2. Tail dependence coefficients for spatial processes

A common way to quantify the strength of the dependence in the upper tail region of two random random variable X and Y is through the bivariate *(upper) tail dependence coefficient* χ. It is defined, for $0 < p < 1$, as the limit of

$$\chi(p) = \mathbb{P}\left(F_Y(Y) > p \mid F_X(X) > p\right) = \frac{\mathbb{P}\left(F_X(X) > p, F_Y(Y) > p\right)}{\mathbb{P}(F_X(X) > p)}$$

when this limit exists (Joe, 1997), i.e.

$$\chi = \lim_{p \to 1^-} \chi(p) \in [0, 1].$$

If $\chi > 0$, X and Y are said to be *asymptotically dependent* (AD), while if $\chi = 0$ the two variables are *asymptotically independent* (AI). Therefore, this coefficient discriminates between the two extremal dependence classes, and provides a quantification of the extremal dependence within the class of asymptotically dependent bivariate distributions, but it does not discriminate among asymptotically independent distributions. An alternative measure that overcomes this limitation is $\bar{\chi}$ (Coles et al., 1999). It is defined, for $0 < p < 1$, as

$$\bar{\chi} = \lim_{p \to 1^-} \frac{2 \log \mathbb{P}(F_X(X) > p)}{\log \mathbb{P}(F_X(X) > p, F_Y(Y) > p)} - 1$$

with $-1 < \bar{\chi} \leq 1$. If X and Y are asymptotically dependent, $\bar{\chi} = 1$. On the other hand, for asymptotically independent variables, the value of this coefficient increases with the strength of the extremal dependence. In the case of independent variables, $\bar{\chi} = 0$. It follows that the pair $(\chi, \bar{\chi})$ provides information about the extremal dependence of (X, Y) both in case of asymptotic dependence and independence. The two coefficients can also refer to the lower tails; their values can be computed by replacing X and Y with $-X$ and $-Y$, respectively. In the following, when necessary, χ_U and $\bar{\chi}_U$ refer to the upper tail and χ_L and $\bar{\chi}_L$ to the lower tail, while when not specified, χ and $\bar{\chi}$ implicitly refer to the upper tail.

A real-valued spatial process defined on a domain $\mathcal{S}$ will be shortly denoted by $\{X(\mathbf{s})\}$ where $\mathbf{s}$ is a generic spatial location in $\mathcal{S}$.

The coefficients χ and $\bar{\chi}$ can also be evaluated for any pair of random variables $X(\mathbf{s}_1)$ and $X(\mathbf{s}_2)$ of $\{X(\mathbf{s})\}$, obtaining the coefficients $\chi(\mathbf{s}_1, \mathbf{s}_2)$ and $\bar{\chi}(\mathbf{s}_1, \mathbf{s}_2)$. For simplicity and when there is no ambiguity, we use χ and $\bar{\chi}$ to refer to $\chi(\mathbf{s}_1, \mathbf{s}_2)$ and $\bar{\chi}(\mathbf{s}_1, \mathbf{s}_2)$. Likewise, if there is a correlation function $\rho(\mathbf{s}_1, \mathbf{s}_2)$ for $\{X(\mathbf{s})\}$, it will be denoted ρ shortly.

3. Gaussian location-scale mixtures

3.1. Model definition

We define the general Gaussian location-scale mixture model as:

$$X(\mathbf{s}) = S + R\, W(\mathbf{s}), \quad \mathbf{s} \in \mathcal{S} \subset \mathbb{R}^d, \tag{2}$$

where $S \in \mathbb{R}$, $R \geq 0$, $\{W(\mathbf{s})\}$ is a standardized Gaussian process with means 0, variances 1 and correlation function $\rho(\mathbf{s}_1, \mathbf{s}_2)$, for two locations $\mathbf{s}_1, \mathbf{s}_2 \in \mathcal{S}$, and (S, R) is independent of $\{W(\mathbf{s})\}$. For notational simplicity, we also assume by default that S and R are independent of each other, although this is not strictly necessary; indeed, some model classes do not make this assumption: two examples are normal mean-variance mixtures (Barndorff-Nielsen et al., 1982; Zhang et al., 2022b) where $S = cR^2$ with $c \in \mathbb{R}$, and the multivariate skew-t distribution (Azzalini and Capitanio, 2003), that can be written as a Gaussian location-scale mixture with dependence between S and R (see Arellano-Valle and Azzalini, 2021).

The general specification (2) encompasses the following three subclasses of models, for which we present in the next subsection some noteworthy specification:

1. if $R = c_1 \in \mathbb{R}^+$, it corresponds to a Gaussian location mixture. Without loss of generality, we assume $c_1 = 1$;
2. if $S = c_2 \in \mathbb{R}$, it corresponds to a Gaussian scale mixture. Without loss of generality, we assume $c_2 = 0$;
3. if no constraint is placed on the values of S and R, we are in the general case of Gaussian location-scale mixtures.

In the presentation of the extremal dependence properties of a particular formulation of the model, the focus is directed towards the upper tail region of the distributions. Lower-tail properties are the same as upper-tail properties when S is a symmetric distribution; otherwise, we can study lower-tail properties by considering the upper-tail properties of the modified process $-S + RW(\mathbf{s})$.

To give a quick idea of the variety of bivariate dependence produced by different specifications of (2), Figure 1 shows firstly three sets of 800 simulated values from a two-dimensional Gaussian vector $(W_1, W_2)^T$ with different correlation, namely $\rho = 0.25, 0.50, 0.75$. Then each simulated set is modified by 800 simulated values of S and R yielding realisations from X_i, $i = 1, 2$. All datasets are presented on uniform margins, i.e. as empirical copulas, to facilitate comparison of dependence properties.

(LM2) $X_i = S + W_i$, with $S = E_1 - E_2$, $E_1 \sim \mathrm{E}(0.4)$, $E_2 \sim \mathrm{E}(2.5)$;

(SM3) $X_i = R\, W_i$ with $R = 1/\sqrt{E_3}$, $E_3 \sim \mathrm{E}(1)$;

(LSM2) $X_i = S + R\, W_i$ with $S = E_1 - E_2$ and $R = \sqrt{E_4}$, $E_4 \sim \mathrm{E}(1/2)$.

Section S2.2 of the supplementary material provides an additional visualisation of the extremal dependence structure of these four models and a comparison of their distributions based on bivariate characteristic functions.

The Gaussian location-scale mixture model (2) can serve as a copula model that provides only the dependence structure of the spatial process. In this case the "observed" process $\{Y(\mathbf{s})\}$ has specific marginal distributions. Assuming for notational simplicity that $\{X(\mathbf{s})\}$ and $\{Y(\mathbf{s})\}$ have space-invariant marginal distributions F_X and F_Y, it is possible to link the two processes with $Y(\mathbf{s}) = F_Y^{-1}\left[F_X\{X(\mathbf{s})\}\right]$, for each $\mathbf{s} \in \mathcal{S}$. For an example of F_Y we refer to Section 7.

3.2. Examples

We introduce examples of models from the three subclasses. Some of them were already discussed in previous literature, but we also propose new instances. Table 1 condenses the assumptions about the distributions of S and R, as well as the resulting multivariate and marginal distributions when they correspond to a known family in the literature. It also includes their extremal dependence properties.

3.2.1. Gaussian location mixtures: $X(\mathbf{s}) = S + W(\mathbf{s})$

A key property of these models is that they allow for asymmetry in the upper and lower tail regions. The following models were studied in Krupskii et al. (2018).

Location Mixture 1 (LM1). Let $S \sim \mathrm{E}(\lambda)$, where λ is the rate parameter. Then, $\{X(\mathbf{s})\}$ is AD with $\chi = \chi_L = \chi_U = 2\Phi(-\lambda\sqrt{1-\rho}/\sqrt{2})$ and $\bar{\chi} = \bar{\chi}_U = \bar{\chi}_L = 1$. As $\lambda \to \infty$, the finite-dimensional distribution of $\{X(\mathbf{s})\}$ converges in distribution to a multivariate Gaussian distribution and we get $\chi \to 0$. Note that finite-dimensional distributions of $\{X(\mathbf{s})\}$ can be expressed in closed form for any dimension (see Krupskii et al., 2018).

Location Mixture 2 (LM2). Let $S \sim \mathrm{AL}(\lambda_1, \lambda_2)$ (see Section 2). $\{X(\mathbf{s})\}$ is AD in both tails, with $\chi_U = 2\Phi(-\lambda_1\sqrt{1-\rho}/\sqrt{2})$, $\chi_L = 2\Phi(-\lambda_2\sqrt{1-\rho}/\sqrt{2})$ and $\bar{\chi}_U = \bar{\chi}_L = 1$. The finite-dimensional distributions of $\{X(\mathbf{s})\}$ can be expressed in closed form (see Krupskii et al., 2018). Three bivariate simulations from this model are shown in Figure 1.

3.2.2. Gaussian scale mixtures: $X(\mathbf{s}) = RW(\mathbf{s})$

Given the symmetric nature of the standard Gaussian distribution, which carries over to Gaussian scale mixture models, we do not distinguish between their upper and lower tail regions. We present five examples of models. Models SM1, SM2, and SM3 have been extensively studied in the literature; see, for example, the articles cited in the respective paragraphs. To the best of our knowledge, model SM4 has not yet been studied as a dependence model.

Scale Mixture 1 (SM1) or Laplace process. Let $R = \sqrt{E}$, with $E \sim \mathrm{E}(1/2)$. Then $\{X(\mathbf{s})\}$ is a Laplace process (Opitz, 2016). It is asymptotically independent with $\chi = 0$ and $\bar{\chi} = \sqrt{2(1+\rho)} - 1$, and its finite-dimensional distribution is available in closed form.

Scale Mixture 2 (SM2). Let $R = \sqrt{G}$, with $G \sim \mathrm{Gamma}(\alpha, 1)$, where α is the shape parameter. Then $\{X(\mathbf{s})\}$ is asymptotically independent with $\chi = 0$ and $\bar{\chi} = \sqrt{2(1+\rho)} - 1$. The finite-dimensional distribution of $\{X(\mathbf{s})\}$ is a multivariate extension of the variance gamma (Madan and Seneta, 1990), which has been widely studied in mathematical finance (Fischer et al., 2025). It can be seen as an interpolation between Gaussian and Laplace processes.

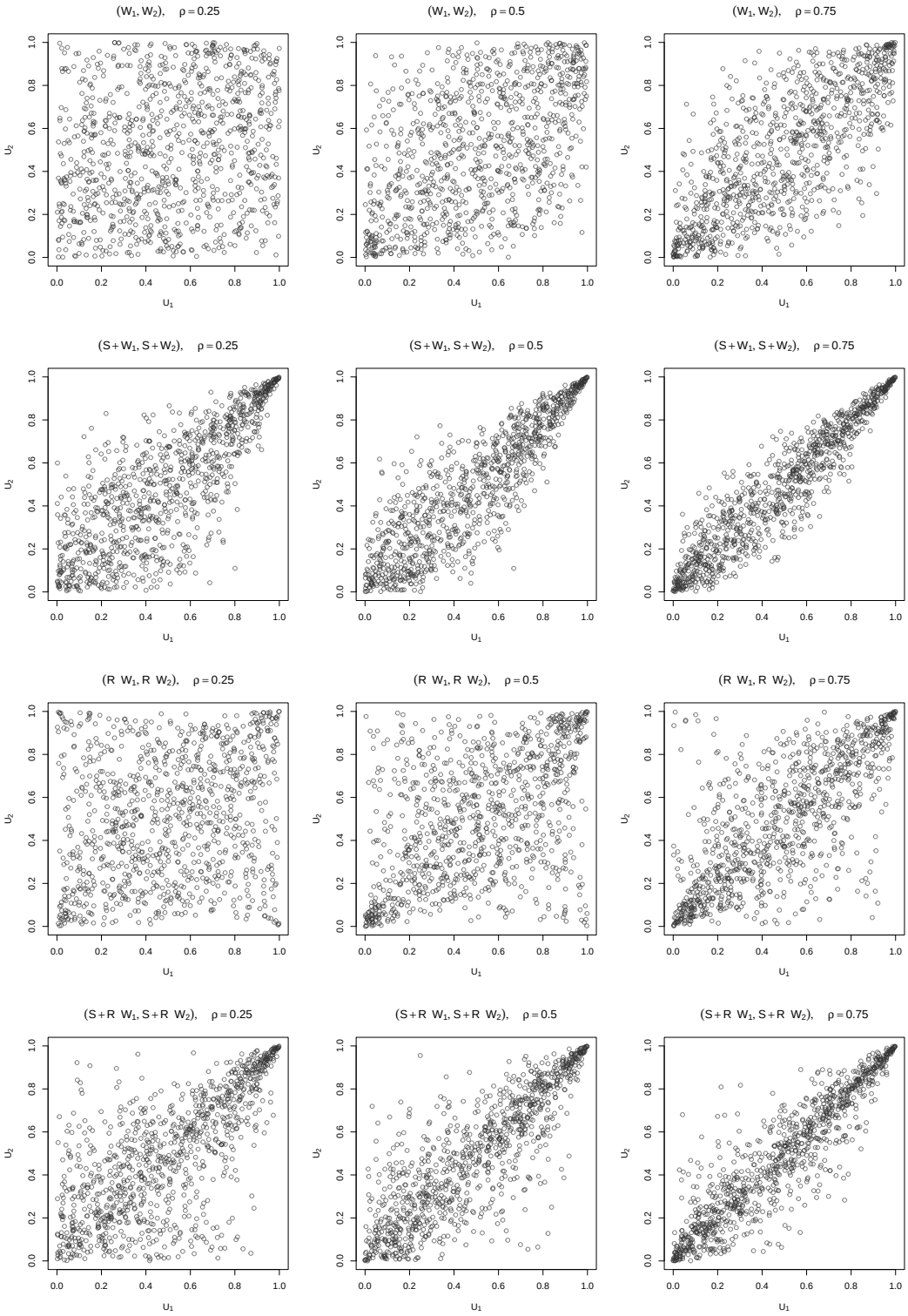


Figure 1: Simulation of 800 replications of three bivariate Gaussian random variables, with $\rho = \text{cor}(W_1, W_2) = 0.25$, 0.5, 0.75, respectively (top row), and of the corresponding models LM2, SM3 and LSM2 (second to fourth row). All variables were transformed to the uniform scale marginally.

Table 1: Examples of Gaussian location-scale mixture models detailed in Section 3.2, with the names of the resulting multivariate or marginal distributions for $\boldsymbol{X}$ when known, and their extremal dependence properties. The parameter $\rho \in (-1, 1)$ is the bivariate correlation of the Gaussian process $\{W(\mathbf{s})\}$ for two generic spatial locations.

Model	F_S, F_R	Resulting distribution	Extremal dependence	Bivariate χ	Bivariate $\bar{\chi}$
Gaussian	-	Multivariate Gaussian	AI	0	ρ
LM1	$S \sim \mathrm{E}(\lambda)$	Closed form but not known family	Upper tail: AD for $\lambda < \infty$ Lower tail: AI	$2[1 - \Phi(\lambda\sqrt{1-\rho}/\sqrt{2})]$	1
LM2	$S \sim \mathrm{AL}(\lambda_1, \lambda_2)$	Closed form but not known family	Upper tail: AD for $\lambda_1 < \infty$ Lower tail: AD for $\lambda_2 < \infty$	$\chi_U = 2[1 - \Phi(\lambda_1\sqrt{1-\rho}/\sqrt{2})]$ $\chi_L = 2[1 - \Phi(\lambda_2\sqrt{1-\rho}/\sqrt{2})]$	1
SM1	$R = \sqrt{E} \sim \mathrm{Rayleigh}(1)$ $E \sim \mathrm{E}(1/2)$	Multivariate Laplace	AI	0	$\sqrt{2(1+\rho)} - 1$
SM2	$R = \sqrt{G}$ $G \sim \mathrm{Gamma}(\alpha, 1)$	Mult. variance gamma	AI	0	$\sqrt{2(1+\rho)} - 1$
SM3	$R = 1/\sqrt{G}$ $G \sim \mathrm{Gamma}(\nu/2, \nu/2)$	Multivariate Student's t Gaussian limit as $\nu \to \infty$	AD for $\nu < \infty$	$2T_{\nu+1}\left(-\sqrt{\nu+1}\sqrt{\frac{1-\rho}{1+\rho}}\right)$	1
SM4	$R = \sqrt{E/G}$ $E \sim \mathrm{E}(1/2)$ $G \sim \mathrm{Gamma}(1/\gamma, 1/\gamma)$	Two-sided GPD (marg.) $X \sim \mathrm{SGPD}(1, \gamma)$ Laplace limit as $\gamma \to 0^+$	AD for $\gamma > 0$	$2T_{1/\gamma+1}\left(-\sqrt{1/\gamma+1}\sqrt{\frac{1-\rho}{1+\rho}}\right)$	1
SM5	$R \sim \mathrm{GPD}(1, \gamma)$	No closed form	AI for $\gamma \leq 0$ AD for $\gamma > 0$	$\gamma \leq 0: \ 0$ $\gamma > 0: \ 2T_{\gamma+1}\left(-\sqrt{\gamma+1}\sqrt{\frac{1-\rho}{1+\rho}}\right)$	$\gamma < 0 : \rho$ $\gamma = 0 : \ \sqrt[3]{4(1+\rho)} - 1$ $\gamma > 0 : \ 1$
LSM1	$S \sim \mathrm{E}(\lambda)$ $R = \sqrt{E}$ $E \sim \mathrm{E}(1/2)$	Exponential (S) + Laplace ($RW(\mathbf{s})$)	Upper tail: AI if $\lambda \geq 1$ Upper tail: AD if $\lambda < 1$ Lower tail: AI	$\lambda \geq 1$: 0 $\lambda < 1$: $\mathbb{E}\left[\min\left\{\frac{Z_1^\lambda}{\mathbb{E}(Z_1^\lambda)}, \frac{Z_2^\lambda}{\mathbb{E}(Z_2^\lambda)}\right\}\right]$, with $Z_i = \exp\{RW_i\}$, $i = 1, 2$	$\lambda > 1 : \max\{2/\lambda - 1, \rho\}$ $\lambda \leq 1 : 1$
LSM2	$S \sim \mathrm{AL}(\lambda_1, \lambda_2)$ $R = \sqrt{E}$ $E \sim \mathrm{E}(1/2)$	Asymmetric Laplace (S) + Laplace ($RW(\mathbf{s})$)	Upper tail: AI if $\lambda_1 \geq 1$ Upper tail: AD if $\lambda_1 < 1$ Lower tail: AI if $\lambda_2 \geq 1$ Lower tail: AD if $\lambda_2 < 1$	Upper tail: as above with $\lambda = \lambda_1$ Lower tail: as above with $\lambda = \lambda_2$ and $Z_i = \exp\{-RW_i\}$, $i = 1, 2$	As above with $\lambda = \lambda_1$ or $\lambda = \lambda_2$

Scale Mixture 3 (SM3) or Student's t process. Let $R = 1/\sqrt{G}$, with $G \sim \text{Gamma}(\nu/2, \nu/2)$, where $\nu/2$ is both the shape and the rate parameter. Then $\{X(\mathbf{s})\}$ is a Student's t process (Røislien and Omre, 2006). The finite-dimensional distribution of $\{X(\mathbf{s})\}$ is available in closed form and it is asymptotically dependent with $\chi = 2T_{\nu+1}\left[-\sqrt{(\nu+1)(1-\rho)/(1+\rho)}\right]$ and $\bar{\chi} = 1$ (see Demarta and McNeil, 2005). However, as $\nu \to \infty$, F_R converges to a Dirac distribution at 1, so $\{X(\mathbf{s})\}$ converges to the Gaussian process $\{W(\mathbf{s})\}$ and $\chi \to 0$. Therefore, this model allows for asymptotic independence as a limit, at the boundary of the parameter space. Three bivariate simulations from this model are shown in Figure 1.

Scale Mixture 4 (SM4) or elliptical generalized Pareto process. Let $R = \sqrt{E}/G$, with $E \sim \text{E}(1/2)$ and $G \sim \text{Gamma}(1/\gamma, 1/\gamma)$, where $1/\gamma$ is both the shape and the rate parameter. Then, the unidimensional distribution of $\{X(\mathbf{s})\}$ is a generalized double Pareto distribution (Armagan et al., 2013), also referred to as symmetric GPD (SGPD), two-sided GPD or type II compound Laplace distribution (Kotz et al., 2001), with density function

$$f(x;\sigma,\gamma) = \frac{1}{2\sigma}\left(1 + \frac{\gamma|x|}{\sigma}\right)^{-1/\gamma-1}, \quad x \in \mathbb{R},$$

denoted as $\text{SGPD}(\sigma,\gamma)$, with $\sigma = 1$ and $\gamma > 0$ by construction. The process $\{X(\mathbf{s})\}$ is AD with

$$\chi = 2T_{1/\gamma+1}\left[-\sqrt{(1/\gamma+1)(1-\rho)/(1+\rho)}\right];$$

see Appendix A for the proof.

As $\gamma \to 0$, G converges to a Dirac distribution at 1, so $\{X(\mathbf{s})\}$ converges to a Laplace process. Therefore, this model also allows for asymptotic independence as a limit, at the boundary of the parameter space.

Scale Mixture 5 (SM5). Let $R \sim \text{GPD}(1,\gamma)$, where $\gamma \in \mathbb{R}$ is the shape parameter. When $\gamma > 0$, $\{X(\mathbf{s})\}$ is asymptotically dependent with $\chi = 2T_{\gamma+1}[-\sqrt{(\gamma+1)(1-\rho)/(1+\rho)}]$ and $\bar{\chi} = 1$. When $\gamma = 0$, $\{X(\mathbf{s})\}$ is asymptotically independent with $\chi = 0$ and $\bar{\chi} = \sqrt[3]{4(1-\rho)} - 1$. When $\gamma < 0$, $\{X(\mathbf{s})\}$ is asymptotically independent with $\chi = 0$ and $\bar{\chi} = \rho$ (see Huser et al., 2017, for these results). Therefore, this model allows for a transition between the two extremal dependence classes in the interior of the parameter space.

3.2.3. Gaussian location-scale mixtures

Let $\{X(\mathbf{s})\} = \{X(\mathbf{s}),\ \mathbf{s} \in \mathcal{S}\}$, with $X(\mathbf{s}) = S + R\,W(\mathbf{s})$, $S \in \mathbb{R}$, $R > 0$, be a generic Gaussian location-scale mixture model. While independence between S and R is not required in general, we assume it in the following two models. A key property of these models is that they allow for asymmetry in the upper and lower tails, possibly with separate transition between extremal dependence regimes in the two tails.

To the best of our knowledge, models LSM1 and LSM2 are new proposals, although similar models have been studied by Gong and Huser (2022a) in the bivariate case and by Gong and Huser (2022b) in the spatial case. In those cases, the structure of the model was $S + g(W(\mathbf{s}))$, with g a marginal transformation of a standard Gaussian vector or process $\{W(\mathbf{s})\}$ to (asymmetric) Laplace distribution. This ensures that marginal parameters can be used to make either S or $g(W(\mathbf{s}))$ relatively heavier-tailed than the other term. Here, similar marginal structure is obtained

directly by a proper Gaussian location-scale mixture process without the need to apply any marginal transformation. This allows efficient conditional simulation following the approach in Section 4 and estimation with the method proposed in Section 5.

Location-Scale Mixture 1 (LSM1). Let $S \sim \mathrm{E}(\lambda)$, and let $R = \sqrt{E}$, with $E \sim \mathrm{E}(1/2)$. Note that $\{R\,W(\mathbf{s})\}$ is a Laplace process, with unidimensional Laplace(1) distribution, whose tail heaviness is exactly the same as that of S when $\lambda = 1$. Three cases can be distinguished; see Appendix A for the proof.

$\lambda < 1$**:** $\{X(\mathbf{s})\}$ is AD, with

$$\chi = (1-\lambda^2)\,\mathbb{E}\left[\min\{Z_1, Z_2\}^{\lambda}\right] \quad \text{and} \quad \bar{\chi} = 1, \tag{3}$$

where $Z_i = \exp\{R\,W(\mathbf{s}_i)\}$, $i = 1, 2$.

$\lambda > 1$**:** $\{X(\mathbf{s})\}$ is AI with

$$\chi = 0 \quad \text{and} \quad \bar{\chi} = \max\left\{\frac{2}{\lambda} - 1,\ \rho\right\}, \tag{4}$$

where $\rho = \mathrm{cor}(W(\mathbf{s}_1), W(\mathbf{s}_2))$.

$\lambda = 1$**:** $\{X(\mathbf{s})\}$ is AI, with $\chi = 0$ and $\bar{\chi} = 1$

Therefore, this model allows for a transition between the two extremal dependence classes for the upper tail, in the interior of the parameter space $\mathbb{R}^+$ of λ.

Location-Scale Mixture 2 (LSM2). Let $S \sim AL(\lambda_1, \lambda_2)$ (see Section 2), and let $R = \sqrt{E}$, with $E \sim \mathrm{E}(1/2)$. As $\lambda_2 \to \infty$, the finite-dimensional distribution of this model converges to that of the LSM1 model with $\lambda = \lambda_1$.

Since the tail heaviness of the asymmetric Laplace distribution F_S is governed by λ_1 in the upper tail and by λ_2 in the lower tail, and is the same as that of an exponential distribution with rate parameter λ_1 or λ_2, respectively, we can extend the asymptotic properties of model LSM1 to both tails of this model. In the upper tail, $\{X(\mathbf{s})\}$ is asymptotically independent if $\lambda_1 \geq 1$ and asymptotically dependent if $\lambda_1 < 1$, where χ_U and $\bar{\chi}_U$ are χ and $\bar{\chi}$ in (3) and (4), replacing λ with λ_1. In the lower tail, $\{X(\mathbf{s})\}$ is asymptotically independent if $\lambda_2 \geq 1$ and asymptotically dependent if $\lambda_2 < 1$, again replacing λ in (3) and (4) with λ_2 to obtain χ_L and $\bar{\chi}_L$; the only other difference from (3) is that here $Z_i = \exp\{-R\,W(\mathbf{s}_i)\}$, $i = 1, 2$.

Therefore, this model allows for a transition between the two extremal dependence classes in the interior of the parameter space, for both tails, and the extremal dependence in the two tails is governed by two different parameters, λ_1 for the upper tail and λ_2 for the lower tail.

4. Conditional simulation

In Gaussian location-scale mixture models, we can exploit the Gaussian structure $\{W(\mathbf{s})\}$ in Equation (2) to easily perform unconditional or conditional spatial simulations. Here, we present a general conditional simulation algorithm that can be used with any distribution of S and R. It generalizes the one proposed (Huser et al., 2017) for Gaussian scale mixtures.

The vector $\boldsymbol{X}$ of $m + h$ random variables $\boldsymbol{X} = (X(\mathbf{s}_1), \ldots, X(\mathbf{s}_{m+h}))^\top$ of the process (2) is partitioned into the vectors $\boldsymbol{X}_1 = (X(\mathbf{s}_1), \ldots, X(\mathbf{s}_m))^\top$ and $\boldsymbol{X}_2 = (X(\mathbf{s}_{m+1}), \ldots, X(\mathbf{s}_{m+h}))^\top$. By

Algorithm 1 Conditional simulation of Gaussian location-scale mixtures

Goal: Simulate $\boldsymbol{X}_2 \mid \boldsymbol{X}_1 = \mathbf{x_1}$ in a Gaussian location-scale mixture vector $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)$
Inputs: Conditioning components $\mathbf{x}_1$; Gaussian variance-covariance matrix $\boldsymbol{\Sigma}$

1. Simulate (S, R) from the conditional distribution of $(S, R) \mid \boldsymbol{X}_1$

$$f_{(S,R)|\boldsymbol{X}_1=\boldsymbol{x}_1}(r,s) = \frac{f_{(S,R,\boldsymbol{X}_1)}(s,r,\boldsymbol{x}_1)}{f_{\boldsymbol{X}_1}(\boldsymbol{x}_1)} = \frac{f_{(S,R,\boldsymbol{W}_1)}(S,R,(\boldsymbol{x}_1-s)/r)r^{-m}}{f_{\boldsymbol{X}_1}(\boldsymbol{x}_1)}$$
$$= \frac{f_{S,R}(s,r)\phi_m((\boldsymbol{x}_1-s)/r;\,\boldsymbol{\Sigma}_{1;1})r^{-m}}{f_{\boldsymbol{X}_1}(\boldsymbol{x}_1)}$$

   through a MCMC Metropolis-Hastings algorithm;
2. Compute $\boldsymbol{W}_1 = (\boldsymbol{X}_1 - S)/R$;
3. Simulate $\boldsymbol{W}_2$ from $\boldsymbol{W}_2 \mid \boldsymbol{W}_1 \sim \mathcal{N}(\boldsymbol{\mu}_{2|1}, \boldsymbol{\Sigma}_{2|1})$, with $\boldsymbol{\mu}_{2|1} = \boldsymbol{\Sigma}_{2;1}\boldsymbol{\Sigma}_{1;1}^{-1}\boldsymbol{W}_1$ and $\boldsymbol{\Sigma}_{2|1} = \boldsymbol{\Sigma}_{2;2} - \boldsymbol{\Sigma}_{2;1}\boldsymbol{\Sigma}_{1;1}^{-1}\boldsymbol{\Sigma}_{1;2}$;
4. Compute and return $\boldsymbol{X}_2 = S + R\,\boldsymbol{W}_2$.

analogy, we partition $\boldsymbol{W} = (W(\mathbf{s}_1), \ldots, W(\mathbf{s}_{m+h}))^\top$ into $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$. Let $\boldsymbol{\Sigma}_{i;j}$, $i, j = 1, 2$, be the corresponding block matrices stemming from $\boldsymbol{\Sigma}$, the correlation matrix of $\boldsymbol{W} = (\boldsymbol{W}_1^\top, \boldsymbol{W}_2^\top)^\top$. Suppose that we want to simulate from $\boldsymbol{X}_2$ conditional on the observed values $\boldsymbol{X}_1 = \boldsymbol{x}_1$. We can do so using Algorithm 1. For Monte–Carlo estimation of properties of the conditional distribution, such as point predictions at unobserved locations, multiple pairs (s, r) can be simulated in the Markov chain in Step 1, and Steps 2–4 can be run for each pair to obtain multiple realizations of $\boldsymbol{X}_2 \mid \boldsymbol{X}_1$. In point 1, independence between S and R, as in our case, can simplify the simulation procedure. Moreover, if interest is in simulating the corresponding vector $\mathbf{Y}_2$ on the scale of the observed process $Y(\mathbf{s})$ (see Section 3.1), this can be done via marginal transformation of the simulated conditional vector $\boldsymbol{X}_2 \mid \boldsymbol{X}_1$ using the unconditional marginal distributions F_Y and F_X of $\mathbf{Y}_2$ and $\boldsymbol{X}_2$, respectively. The conditional simulation algorithm is applied in Section 7.4 to real data, with an illustration in Figure 7. The median of the simulated data is used as point prediction on a hold-out region, and its values are compared to the observed ones in that region. We note that the potential applications of the proposed simulation method extend beyond point prediction. It generates the full conditional distributions at unobserved locations, enabling the calculation of threshold exceedance probabilities based on the flexible extremal dependence structure of the model class presented in this paper.

The MCMC Metropolis-Hastings algorithm (Robert and Casella, 2004) inside of Step 1 of Algorithm 1 creates a homogenous Markov chain, $(S^{(k)}, R^{(k)})$, $k = 1, 2, \ldots$, whose distribution converges to the distribution of $(S, R) \,|\, \boldsymbol{X}_1$. The transition from $(S^{(k)}, R^{(k)})$ to $(S^{(k+1)}, R^{(k+1)})$ is described in the following. First, propose (S', R') such that $S' = S^{(k)} + Z_S$ and $\log(R') = \log(R^{(k)}) + Z_R$, where $Z_S, Z_R \overset{\text{ind.}}{\sim} \mathcal{N}(0, 1)$. A logarithmic transformation could also be applied to S when its support is $\mathbb{R}^+$. Then, we set

$$\left(S^{(k+1)}, R^{(k+1)}\right) = \begin{cases} (S', R') & \text{with probability } \rho(S', R', S^{(k)}, R^{(k)}), \\ (S^{(k)}, R^{(k)}) & \text{otherwise,} \end{cases} \tag{5}$$

where the acceptance ratio is (see Appendix B for the derivation)

$$\rho(s', r', s^{(k)}, r^{(k)}) = \min\left\{ \frac{f_{S,R}(s', r')\phi_m((\boldsymbol{x}_1 - s')/r'; \boldsymbol{\Sigma}_{1;1}) r'^{-m+1}}{f_{S,R}(s^{(k)}, r^{(k)}))\phi_m((\boldsymbol{x}_1 - s^{(k)})/r^{(k)}; \boldsymbol{\Sigma}_{1;1}) r^{(k-m+1)}}, 1 \right\}.$$

In practice, the distribution of the initial states of the Markov chain may be quite different to the distribution of $(S, R) \mid \boldsymbol{X}_1$. To mitigate this, it is desirable to identify a "burn-in" $(S^{(k)}, R^{(k)})$, $k = 1, \dots K$, which is then discarded. After the burn-in period, the values of the Markov chain still have autocorrelation. A thinning procedure can be applied to generate approximately independent realizations of the conditional distribution of $\boldsymbol{X}_2 \mid \boldsymbol{X}_1$.

5. Inference

Two cases can be distinguished in real-data applications: when (2) is used to directly model observed spatial data, and when (2) provides a copula model describing the dependence structure of an underlying spatial process. The latter is a common framework when we want to specify and fit sub-asymptotic models for extremes (Huser et al., 2025). In such a case, the model for the "observed" process $\{Y(\mathbf{s})\}$ is the model induced by the transformation $Y(\mathbf{s}) = F_Y^{-1}(F_X(X(\mathbf{s})))$. Note that, for notational simplicity, we assume that the cdfs of $X(\mathbf{s})$ and $Y(\mathbf{s})$ (F_X and F_Y, resp.) are invariant with respect to $\mathbf{s}$.

A likelihood-based approach requires the evaluation of the partial derivatives of the cdf of $\boldsymbol{X} = (X(\mathbf{s}_1), \dots, X(\mathbf{s}_m))^\top$ for given m (distinct) spatial locations $\mathbf{s}_1, \dots, \mathbf{s}_m$. We denote by $F_{\boldsymbol{X}}(\cdot; \theta)$ the cdf of $\boldsymbol{X}$. The unknown parameter $\theta = (\theta_W, \theta_{S,R})$, of the distribution of $\boldsymbol{X}$, is split into two parameters: θ_W, the parameter(s) in the covariance matrix $\boldsymbol{\Sigma}_{\theta_W}$ of the standard Gaussian vector $\boldsymbol{W} = (W(\mathbf{s}_1), \dots, W(\mathbf{s}_m))^\top$; $\theta_{S,R}$, the parameter(s) of the distribution of S and R (if any). We obtain the cdf of $\boldsymbol{X}$ by

$$F_{\boldsymbol{X}}(\boldsymbol{x}; \theta) = \int_{-\infty}^{+\infty} \int_0^{+\infty} \Phi_m\left(\frac{x_1 - s}{r}, \dots, \frac{x_m - s}{r}; \boldsymbol{\Sigma}_{\theta_W} \right) f_{S,R}(s, r; \theta_{S,R})\, ds\, dr, \tag{6}$$

where $\boldsymbol{x} = (x_1, \dots, x_m)^\top$, $f_{S,R}(\,\cdot\,; \theta_{S,R})$ is the joint density of R and S and $\Phi_m(\,\cdot\,; \boldsymbol{\Sigma})$ is the m-dimensional cdf of a standard Gaussian vector with correlation matrix $\boldsymbol{\Sigma}$. In few cases, such as the Student's t stochastic process and the Laplace stochastic process (see Section 3.2), expression (6) is available in analytical form. However, this is not true in general, so the evaluation of the likelihood functions is usually performed by numerical integration (Huser et al., 2017; Krupskii et al., 2018). This leads to a high numerical cost, especially when the number of spatial locations m is large, since an m-dimensional Gaussian density function has to be evaluated for each value of r and s within the integrals. The computational cost is even higher when the model is fitted as copula model.

In the following, we propose a two-step estimation procedure that does not rely on specific assumptions about the distributions of R or S, with the aim of reducing the computational burden.

First, we transform vector $\boldsymbol{X}$ to obtain a vector whose distribution depends only on the distribution of $\boldsymbol{W}$ and for which we know the analytical expression. This allows us to avoid numerical integration. We can then use this distribution to estimate the parameter θ_W, denoted by $\widehat{\theta}_W$. In the second step, we reduce the dimension of $\boldsymbol{X}$ to a random variable whose distribution depends on all the parameters, $\theta = (\theta_W, \theta_{S,R})$. We plug the value $\widehat{\theta}_W$ into the value of θ_W to obtain a profile distribution from which to make inferences about the value of $\theta_{S,R}$.

We will begin by describing our proposal in the case that we observe n independent value, $\boldsymbol{x}_i$, $i = 1, \ldots, n$ of $\boldsymbol{X}$, subsections 5.1 and 5.2. Then we describe how the estimation procedure can be modified when using the distribution of $\boldsymbol{X}$ as a copula model (subsection 5.3). We conclude this section with a discussion of how to assess estimation uncertainty.

5.1. Step 1: estimation of θ_W

The estimation method is inspired by the REstricted Maximum Likelihood (REML; Patterson and Thompson, 1971). We identify transformations of the vector $\boldsymbol{X}$ that are independent of S and R, whose analytical distributions can be employed for likelihood-based inference on θ_W. The transformed vector, denoted by $\boldsymbol{Z}$, will have different derivations in the three cases: location, scale, and location-scale mixtures. In all three cases, the restricted likelihood function for θ_W is then constructed as

$$\mathcal{L}(\theta_W;\, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_n) = \prod_{i=1}^{n} f_{\boldsymbol{Z}}\left(\boldsymbol{z}_i;\, \theta_W\right),$$

where $\boldsymbol{z}_i$ are i.i.d. observations of $\boldsymbol{Z}$.

Gaussian location mixtures

We transform $\boldsymbol{X}$ into

$$\boldsymbol{Z} = \left(X(\mathbf{s}_2) - X(\mathbf{s}_1), \ldots, X(\mathbf{s}_m) - X(\mathbf{s}_1)\right)^\top = \left(W(\mathbf{s}_2) - W(\mathbf{s}_1), \ldots, W(\mathbf{s}_m) - W(\mathbf{s}_1)\right)^\top. \tag{7}$$

Its density function (see Appendix C.1) is given by

$$f_{\boldsymbol{Z}}(\boldsymbol{z};\, \theta_W) = (2\pi)^{-(m-1)/2} |\boldsymbol{A}\boldsymbol{\Sigma}_{\theta_W}\boldsymbol{A}^\top|^{-1/2} \exp\left(-\frac{1}{2}\boldsymbol{z}^\top (\boldsymbol{A}\boldsymbol{\Sigma}_{\theta_W}\boldsymbol{A}^\top)^{-1}\boldsymbol{z}\right),$$

where $\boldsymbol{z} = \left(x(\mathbf{s}_2) - x(\mathbf{s}_1), \ldots, x(\mathbf{s}_m) - x(\mathbf{s}_1)\right)^\top$ and $\boldsymbol{A}$ is the $(m-1) \times m$ matrix

$$\boldsymbol{A} = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ -1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ -1 & 0 & 0 & \cdots & 1 \end{bmatrix}. \tag{8}$$

Gaussian scale mixtures

We transform $\boldsymbol{X}$ into

$$\boldsymbol{Z} = \left(\frac{X(\mathbf{s}_2)}{X(\mathbf{s}_1)}, \ldots, \frac{X(\mathbf{s}_m)}{X(\mathbf{s}_1)}\right)^\top = \left(\frac{W(\mathbf{s}_2)}{W(\mathbf{s}_1)}, \ldots, \frac{W(\mathbf{s}_m)}{W(\mathbf{s}_1)}\right)^\top. \tag{9}$$

Its density function (see Appendix C.2), is given by

$$f_{\boldsymbol{Z}}(\boldsymbol{z};\, \theta_W) = \pi^{-m/2} |\boldsymbol{\Sigma}_{\theta_W}|^{-1/2} \Gamma\left(m/2\right) \left(\dot{\boldsymbol{z}}^\top \boldsymbol{\Sigma}_{\theta_W}^{-1} \dot{\boldsymbol{z}}\right)^{-m/2},$$

where $\dot{\boldsymbol{z}} = (1, \boldsymbol{z}^\top)^\top$ and $\boldsymbol{z} = \left(x(\mathbf{s}_2)/x(\mathbf{s}_1), \ldots, x(\mathbf{s}_m)/x_i(\mathbf{s}_1)\right)^\top$.

Gaussian location-scale mixtures

We transform $\boldsymbol{X}$ into

$$\begin{aligned}\boldsymbol{Z} &= \left(\frac{X(\mathbf{s}_3)-X(\mathbf{s}_1)}{X(\mathbf{s}_2)-X(\mathbf{s}_1)},\ldots,\frac{X(\mathbf{s}_m)-X(\mathbf{s}_1)}{X(\mathbf{s}_2)-X(\mathbf{s}_1)}\right)^\top \\ &= \left(\frac{W(\mathbf{s}_3)-W(\mathbf{s}_1)}{W(\mathbf{s}_2)-W(\mathbf{s}_1)},\ldots,\frac{W(\mathbf{s}_m)-W(\mathbf{s}_1)}{W(\mathbf{s}_2)-W(\mathbf{s}_1)}\right)^\top.\end{aligned} \tag{10}$$

Its density function (see Appendix C.3) is

$$f_{\boldsymbol{Z}}(\boldsymbol{z};\theta_W) = \pi^{-(m-1)/2}|\boldsymbol{A}\boldsymbol{\Sigma}_{\theta_W}\boldsymbol{A}^\top|^{-1/2}\Gamma\left(\frac{m-1}{2}\right)\left[\dot{\boldsymbol{z}}^\top(\boldsymbol{A}\boldsymbol{\Sigma}_{\theta_W}\boldsymbol{A}^\top)^{-1}\dot{\boldsymbol{z}}\right]^{-(m-1)/2},$$

where $\dot{\boldsymbol{z}} = (1,\boldsymbol{z}^\top)^\top$ and $\boldsymbol{z} = ([x(\mathbf{s}_3)-x(\mathbf{s}_1)]/[x(\mathbf{s}_2)-x(\mathbf{s}_1)],\ldots,[x(\mathbf{s}_m)-x(\mathbf{s}_1)]/[x(\mathbf{s}_2)-x(\mathbf{s}_1)])^\top$.

Invariance of the estimators to the selection of the reference locations

In the first step, we choose one or two reference locations for the transformed vector $\boldsymbol{Z}$. However, the choice of the reference locations has no impact on the resulting estimates. Indeed, if another location $\mathbf{s}_k$, $k \in \{1,\ldots,m\}$ (or a pair of locations $\mathbf{s}_j, \mathbf{s}_k$, $j,k \in \{1,\ldots,m\}$) is chosen instead of $\mathbf{s}_1$ (or instead of $\mathbf{s}_1, \mathbf{s}_2$), the corresponding likelihood functions would be proportional to those shown here, leading to the same maximum. A proof is provided in the supplementary material, Section S1.

5.2. Step 2: estimation of $\theta_{S,R}$

We consider parameter estimation in the case of a Gaussian location-scale mixture; the adaptation to simpler cases presents no difficulty. The vectors $\boldsymbol{X}$ and $\boldsymbol{W}$ are transformed into the (spatial) average of their m respective elements, $\bar{X}$ and $\bar{W}$, which are linked by the equation

$$\bar{X} = S + R\,\bar{W}. \tag{11}$$

Note that $\bar{W} \sim \mathcal{N}\left(0, \sigma^2_{\bar{W}}\right)$, with $\sigma^2_{\bar{W}} = m^{-2}\,\mathbf{1}^\top\boldsymbol{\Sigma}_{\theta_W}\mathbf{1}$, and that θ_W has been estimated in the first step. Identifiability of $\theta_{S,R}$ from the distribution of $\bar{X}$, when both S and R are non-degenerate random variables, is discussed in Section S2.1 in the supplementary material. A short proof is also provided for the two location-scale mixture models presented in this work: LSM1 and LSM2.

In general, the distribution of $\bar{X}$ is not known and once again, it would be necessary to resort to numerical integration as in (6) to evaluate the cdf

$$F_{\bar{X}}(x;\theta_W,\theta_{S,R}) = \int_{-\infty}^{+\infty}\int_0^{+\infty}\phi\left(\frac{x-s}{r\,\sigma_{\bar{W}}}\right) f_{S,R}(s,r;\,\theta_{S,R})\,ds\,dr,$$

or its density. However, the simulation of $\bar{X}$ does not require the knowledge of $F_{\bar{X}}$ and, unlike the vector case $\boldsymbol{X}$, the simulation is fast. This opens up the possibility of deriving estimates of $\theta_{S,R}$ based on a minimum distance method between distributions (Wolfowitz, 1957; Bolthausen, 1977) by approximating $F_{\bar{X}}$ with a Monte Carlo method.

More precisely, we consider the averages $\bar{x}_i$ from each $\boldsymbol{x}_i$, $i = 1,\ldots,n$, and we denote F_n their empirical cdf. A minimum distance estimator of $\theta_{S,R}$, pretending that θ_W is known, minimizes a

distance $\tau(\theta_{S,R})$ between the empirical cdf $F_n(x)$ and the theoretical cdf $F(x;\,\theta_{S,R})$. In particular, minimization of the L^2 distance

$$\tau(\theta_{S,R}) = \int [F_n(x) - F_{\bar{X}}(x;\,\theta_W, \theta_{S,R})]^2 F_{\bar{X}}(dx;\,\theta_W, \theta_{S,R})$$

with respect to $\theta_{S,R}$ leads to a Cramér-von Mises estimator (Pollard, 1980). Estimators minimizing the L^2 distance are typically strongly consistent and asymptotically normal; they are also more robust to model misspecification than maximum likelihood estimators (Parr, 1981). We consider the sample version of the distance, i.e.

$$T(\theta_{S,R}) = \sum_{i=1}^{n} \left[\frac{i - 1/2}{n} - F_{B,\bar{X}}\left(\bar{x}_{(i)}; \theta_W, \theta_{S,R}\right) \right]^2, \tag{12}$$

where $\bar{x}_{(i)}$, $i = 1, \ldots, n$, are the ordered values of $\bar{x}_i$, $i = 1, \ldots, n$. Note that we approximate $F_{\bar{X}}$ by $F_{B,\bar{X}}$, the empirical cdf of B i.i.d. copies of (11), with $(S, R) \sim F_{S,R}(\cdot;\,\theta_{S,R})$ and $\bar{W} \sim \mathcal{N}(0, m^{-2}\,\mathbf{1}^\top \boldsymbol{\Sigma}_{\widehat{\theta}_W} \mathbf{1})$. A formal proof of consistency of the (simulated) estimator minimizing (12) is given in Section S3 of the supplementary material. We typically set $B = 10^6$ from each distribution, setting the same seeds at each optimization iteration, as shown in the supplementary code. Although this intermediate Monte Carlo approximation must be repeated for each value of $\theta_{S,R}$ in the iterations of the minimization algorithm of (12), in our experience the procedure remains quite fast because it only involves simulating from univariate distributions.

5.3. Copula models

The two-step estimation method described above starts from the premise to observe data from $\boldsymbol{X}$. However, as explained in Section 3.1, in real data applications model (2) may be used as a copula model for the dependence inside of $\mathbf{Y} = (Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_m))^\top$. According to this modelling approach each element of $\boldsymbol{Y}$ is linked to each element of $\boldsymbol{X}$ by $X(\mathbf{s}_i) = F_X^{-1}\left(F_Y(Y(\mathbf{s}_i))\right), \quad i = 1, \ldots, m$, where F_X and F_Y are the cdf of $X(\mathbf{s}_i)$ and $Y(\mathbf{s}_i)$, respectively.

We assume that the analytical expression of F_Y is either known or can be consistently estimated, see Section 7 for an example. In this way, we get

$$\mathbf{U} = (U(\mathbf{s}_1), \ldots, U(\mathbf{s}_m))^\top = (F_Y(Y(\mathbf{s}_1)), \ldots, F_Y(Y(\mathbf{s}_m)))^\top. \tag{13}$$

Then, to transform $\mathbf{U}$ into $\boldsymbol{X}$, one would need to know the marginal distributions of the latter, F_X. As mentioned above, however, analytical expressions for F_X and F_X^{-1}, the quantile function, are rarely available. On the other hand, for a fixed value of $\theta_{S,R}$ an inexpensive Monte Carlo approximation of the quantile function F_X^{-1} can be obtained by generating a sufficiently large number of i.i.d. copies of $X(\mathbf{s}) = S + RW(\mathbf{s})$, with $(S, R) \sim F_{S,R}(\cdot;\,\theta_{S,R})$ and $W \sim \mathcal{N}(0, 1)$. We denote this approximation by $F_X^{-1}(\cdot;\,\theta_{S,R})$, with a slight abuse of notation, and we set

$$\boldsymbol{X}(\theta_{S,R}) = \left(F_X^{-1}(U(\mathbf{s}_1);\,\theta_{S,R}), \ldots, F_X^{-1}(U(\mathbf{s}_m);\,\theta_{S,R})\right)^\top. \tag{14}$$

We now incorporate this additional Monte Carlo approximation into the two-step procedure for estimating the parameters θ_W and $\theta_{S,R}$, as described above, to obtain Algorithm 2. The input of the algorithm are the vectors $\boldsymbol{u}_i$, $i = 1, \ldots, n$, obtained by transforming the n independent observed values of $\boldsymbol{Y}$, $\boldsymbol{y}_i$, using (13).

Algorithm 2 Parameter estimation in the copula case

Input: $\mathbf{u}_i$, $i = 1, \ldots, n$.
Output: Estimated values $\widehat{\theta}_W$ and $\widehat{\theta}_{S,R}$

1. For a current value of $\theta_{S,R}$, compute $\boldsymbol{x}_i(\theta_{S,R})$, $i = 1, \ldots, n$ as in (14);
2. Estimate θ_W with the method described in Section 5.1, obtaining $\widehat{\theta}_W$;
3. Compute the Cramér-von Mises statistic $T(\theta_{S,R})$ as in (12);
4. Iterate points 1–3 by changing the fixed values of $\theta_{S,R}$, to minimize $T(\theta_{S,R})$.

The iterative estimation of $\widehat{\theta}_W$ and $\widehat{\theta}_{S,R}$ increases the computational cost for the computational case. A comparison between the computation time in the copula and non-copula cases can be found in Section 6.2, and in particular in Table 2.

5.4. Confidence intervals

One option for obtaining confidence intervals for the estimates of θ_W and $\theta_{S,R}$ would be to compute the asymptotic variances of the two-step estimators (Gong and Samaniego, 1981; Zhao and Joe, 2005) and use them to construct the confidence intervals.

Here, however, a parametric bootstrap procedure can be employed to approximate these confidence intervals thanks to the ease of simulating from the models. In the non-copula case, once point estimates $\widehat{\theta}_W$ and $\widehat{\theta}_{S,R}$ are obtained as described in Sections 5.1 and 5.2, it is sufficient to simulate, from the Gaussian location-scale mixture model (2), B dataset analogues to the observed one for spatial locations and number of replications. Then, for $b = 1, \ldots, B$, estimates $\widehat{\theta}^b_W$ and $\widehat{\theta}^b_{S,R}$ can be obtained following the same procedures; their distribution can be used to build confidence intervals. Since the estimator $\widehat{\theta}_W$ is obtained by maximizing a correctly specified likelihood, we could alternatively assess its uncertainty by using standard tools for maximum likelihood estimators.

Interval estimation becomes slightly more complicated in the copula case (Section 5.3). In this case, after B datasets from model (2) are simulated using the estimated parameters $\widehat{\theta}_W$ and $\widehat{\theta}_{S,R}$, the data must be transformed by applying the function $F_X(\cdot; \widehat{\theta}_{S,R})$, which leads to B datasets on the uniform marginal scale. Then, two options are available to obtain estimates $\widehat{\theta}^b_W$ and $\widehat{\theta}^b_{S,R}$, for $b = 1, \ldots, B$:

- The standard option is to apply to each dataset the procedure of Section 5.3 ; in some cases, this may be computationally demanding since it involves repeating B times the recursive estimation described in Algorithm 2.
- A faster option is to obtain each estimate $\widehat{\theta}^b_{S,R}$ by fixing θ_W to the value $\widehat{\theta}_W$ estimated on the observed dataset and minimizing (12); then, each $\widehat{\theta}^b_W$ can be estimated, with the method described in Section 5.1, on the dataset transformed by $F_X^{-1}(\cdot\,; \widehat{\theta}^b_{S,R})$. Since uncertainty is not propagated between the estimators of $\theta_{S,R}$ and θ_W, the width of confidence intervals could be (slightly) underestimated.

6. Simulation study

In this section (see also the Supplementary Material), the proposed inferential method is evaluated on simulated data, and its performance is compared to that of some of the existing ones. In

particular, data are simulated from Gaussian location-scale mixtures at locations sampled randomly in a square $[0, 200] \times [0, 200]$, which is similar to the dimension (in km) of the spatial domain of the real data application (see Section 7). The correlation function of the Gaussian process $\{W(\mathbf{s})\}$ is a Matérn correlation function with range parameter $\varphi > 0$ and smoothness parameter $\eta > 0$, such that $\theta_W = (\varphi, \eta)$ and, for $\mathbf{s}_1, \mathbf{s}_2 \in \mathcal{S} \subset \mathbb{R}^2$,

$$\rho(\mathbf{s}_1, \mathbf{s}_2) = \frac{1}{2^{\eta-1}\Gamma(\eta)} \left(\frac{2\sqrt{\eta}\|\mathbf{s}_1 - \mathbf{s}_2\|}{\varphi} \right)^{\eta} K_{\eta} \left(\frac{2\sqrt{\eta}\|\mathbf{s}_1 - \mathbf{s}_2\|}{\varphi} \right),$$

where $\|\cdot\|$ is the Euclidean distance and K_η is the modified Bessel function of the second kind of order η. In the simulations, the true values of these parameters are fixed at $\varphi = 50$ and $\eta = 0.5$. Note that the value $\eta = 0.5$ defines the exponential correlation function; other simulations with $\eta = 1$ were also conducted, leading to very similar results in terms of bias and variance of the estimators, and are not shown here to limit the amount of results presented. The number of spatial locations, m, and the number of independent replications in each spatial location, n, are considered according the following data-poor, data-rich and intermediate configurations: A) $m = 50$, $n = 100$; B) $m = 100$, $n = 500$; C) $m = 200$, $n = 1000$; D) $m = 400$, $n = 2000$. For each replication, the m locations are sampled uniformly within $[0, 200] \times [0, 200]$, increasing the uncertainty in the estimation.

Section 6.1 compares the proposed method with the maximization of the full likelihood, for two models for which the latter is available in closed form; Section S4.1 (in the supplementary material) numerically illustrates the consistency of the proposed method for models belonging to the three classes of Gaussian location, scale and location-scale mixtures; Section S4.2 (in the supplementary material) compares the proposed method with the integration-based one from equation (6); Section S4.3 (in the supplementary material) reproduces the estimations of Sections 6.1 and S4.1 adopting a copula approach; finally, Section 6.2 includes comments on the computation time necessary for parameter estimation in all the above cases.

6.1. Comparison with full likelihood

In general, the finite-dimensional distribution of Gaussian location-scale mixtures (2) is not known in closed form. However, in few cases, such as Laplace and Student's t processes (see Section 3.2) it is available in closed form. In these cases, we can perform inference on the model parameter by maximizing the full likelihood and use it as a benchmark to quantify the loss of statistical efficiency of the proposed inferential solutions. In the following cases, inference is performed on 100 simulated datasets for each of the configurations A, B, C and D (see above).

Figure 2 shows results for the estimation of $\theta_W = (\varphi, \eta)$ for a Laplace process (model SM1 in Table 1). The scaling random variable R has no parameter to estimate in this case, since it has standard Rayleigh distribution. For the parameter θ_W, the proposed estimation method appears unbiased and does not lose much efficiency, compared to the full maximum likelihood.

Figure 3 displays analogue estimation results on a Student's t process (model SM3 in Table 1) with $\nu = 2$ degrees of freedom. The estimators of $\theta_W = (\varphi, \eta)$ show performances similar to the maximum likelihood estimators, as in the previous case. However, the estimator of $\theta_R = \nu$ exhibits clearly higher variability than the one based on the full likelihood, especially in the lower-dimensional configurations. This is due to the relative statistical inefficiency of the minimum

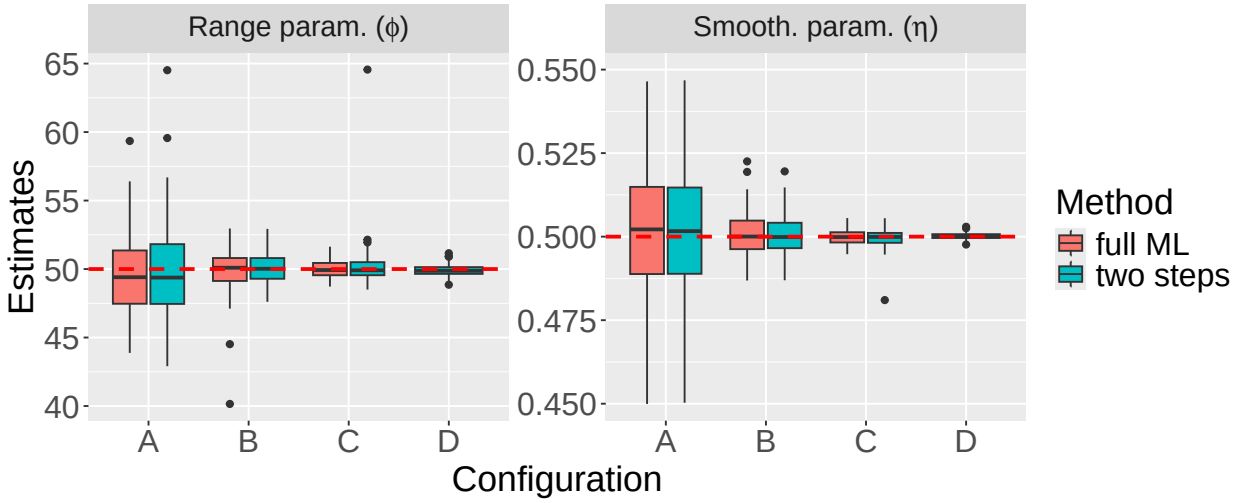


Figure 2: Estimation of the parameters of a Laplace process on 100 simulated datasets, for each configuration, with true values $\varphi = 50$ and $\eta = 0.5$ (red lines). Orange boxplots report full maximum likelihood estimates (not computed for Configuration D for computational reasons), while blue boxplots show the estimates based on the likelihood of ratios explained in Section 5.1.

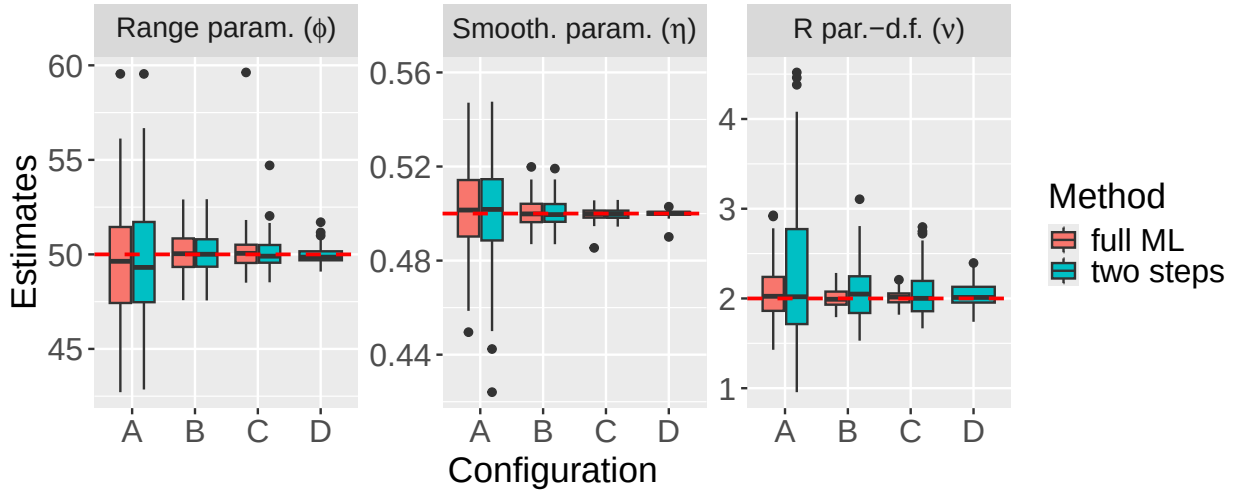


Figure 3: Estimation of the parameters of a Student's t process on 100 simulated datasets for each configuration, with true values (red lines) $\varphi = 50$, $\eta = 0.5$ and $\nu = 2$. Orange boxplots report full maximum likelihood estimates (not computed for Configuration D for computational reasons), while blue boxplots show the two-step solution for inference explained in Section 5.

distance estimator used in Step 2 (see Section 5.2), compared to the maximum likelihood estimator. The efficiency of the two-step estimator could be improved by replacing the second step with the maximization of the likelihood based on the distribution of $\bar{X}$ in (11), in particular when it is available in closed form and in the non-copula cases. However, for the cases presented here, we opted to use the same estimation method, which is valid for every distribution of S and $R > 0$, and for all copula cases.

To further confirm consistent behaviour of the two-step estimation procedure, in particular for

Table 2: Average computation time (rounded to at most 2 significant digits), in seconds, for parameter estimation on a single dataset in the simulation studies of Section 6 and Section S4, for the cases with 2, 3 and 4 parameters, without and with copula, and for configurations A ($m = 50, n = 100$), B ($m = 100, n = 500$), C ($m = 200, n = 1000$) and D ($m = 400, n = 2000$).

Number of parameters	Method	No copula				Copula			
		A	B	C	D	A	B	C	D
2	two steps	0.3	1	5	32	1	2	8	-
	full likelihood	2.6	70	550	-	3	50	540	-
	integration	450	8600	-	-	-	-	-	-
3	two steps	45	50	55	85	70	110	250	-
	full likelihood	15	130	800	-	130	300	1200	-
	integration	470	5300	-	-	-	-	-	-
4	two steps	290	310	320	370	-	-	-	-

the parameter $\theta_{S,R}$ of the random scale and location of the mixture models, another simulation study is provided in Section S4.1 in the supplementary material, for some of the models presented in Section 3.2.

6.2. Computation time

In addition to consistency and statistical efficiency, an important feature of estimation methods for Gaussian location-scale mixtures is computational efficiency. In many cases, the computation times may be inflated by the presence of latent variables in the models, with additional parameters to estimate. Table 2 reports the average computation time, in seconds, for the simulations of Section 6 and Section S4 in the supplementary material. These averages are computed over the 100 datasets of each simulation study and over all the models belonging to each setting, with the exception of the integration-based method, for which 50 datasets are simulated. In particular, the case with 2 parameters refers to the Laplace process (model SM1, Figures 2, S6 and S8); the case with 3 parameters covers the Student's t process (model SM3, Figures 3 and S9), as well as model LM1 (Figures S2, S7 and S10), model SM4 (Figures S3 and S11) and model LSM1 (Figures S4 and S12); finally, the case with 4 parameters refers to model LM2 (Figure S5). The code for the simulation studies is written in the R programming language and executed on a 2.3 GHz machine with eight cores and 12 GB of memory. We use the Nelder-Mead method for multivariate optimization and the Brent method for univariate optimization.

The computation times shown in Table 2 indicate that the two-step approach is much faster than the maximization of the closed-form full likelihood, with much better scalability as the number of locations m and replications n is increased, such as in Configurations C and D. This may be due to the separation of the functions to be optimized in two steps, and to the simpler form of these functions compared to that of some likelihood functions. The third method, based on the numerical integration of (6), is much more computationally demanding and becomes prohibitive for Configurations C and D. The cases reported here refer only to single integrals, as explained in Section S4.2. A general case of Gaussian location-scale mixture (with both random location S and random scale R) would involve a double integral, making the computation time even longer.

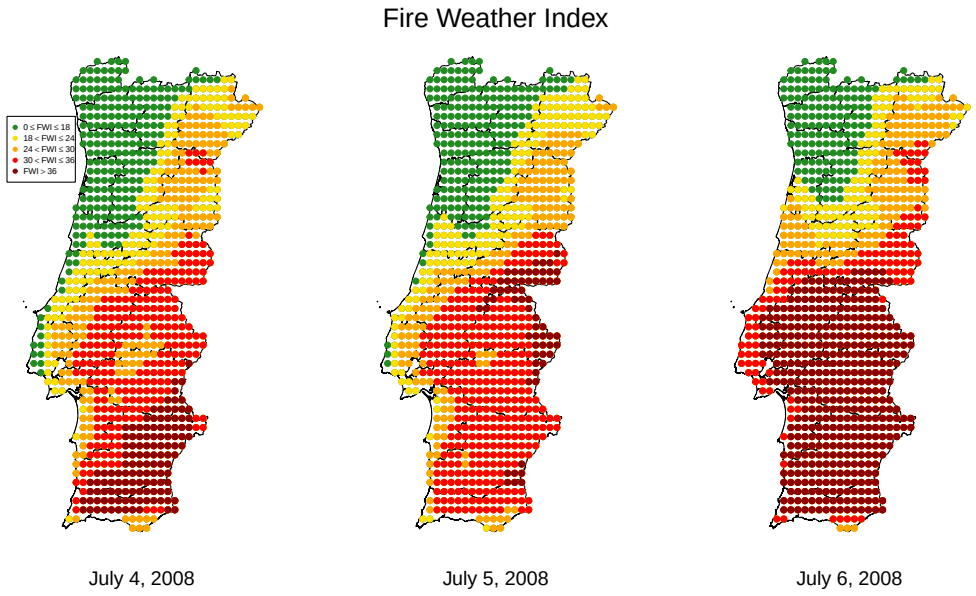


Figure 4: Three consecutive days of observed FWI over Portugal.

On the other hand, the proposed method requires similar computation time for the three cases of Gaussian location, scale and location-scale mixtures.

7. Real data application to fire weather data in Portugal

7.1. Data description

ERA5-Land is a large hourly global dataset of various weather variables on a longitude-latitude grid of size $0.1° \times 0.1°$. It is obtained by assimilating observation data (e.g., weather stations, radar, satellite) into a physical climate model. Data are available for the period from 1940 to today, and are routinely used as a realistic approximation of the true weather of the past. Following common practice, we here refer to them as "observations", even though producing them has required using a model. We apply the models we propose to a dataset of daily Fire Weather Index (FWI), a continuous random variable measuring the weather-related risk of wildfires, computed from ERA5-Land data on the whole domain of continental Portugal for recent years. The original dataset includes 23 years (from 2000 to 2022) of daily observations on 947 spatial locations. For illustration of the data, the maps in Figure 4 show FWI values observed for three consecutive days in summer 2008. Although the FWI, which is based on weather information, appears to take higher values in the south of Portugal, the area with the highest historical susceptibility to wildfires is the north of the country, one of Europe's wildfire hotspots, due to the land-use and land-cover configurations(Verde and Zêzere, 2010; Neves et al., 2023; see also Moghli et al., 2025). Therefore, we split the geographical domain into two subsets, as shown in Figure S13 in the supplementary material, and our analysis focuses only on the northern region. This reduces the dimensionality of the dataset to $m = 491$ spatial locations. Moreover, to ensure approximate temporal stationarity, we restrict the

analysis to the two main wildfire months of July and August, leading to 62 values at each location for each of the 23 years, for a total of $n = 1426$ replications.

7.2. Marginal modelling

As explained in Section 3.1, Gaussian location-scale mixtures (2) may be used as copula models for the dependence structure of the data and specifying a marginal distribution of the observed data. A typical approach in extreme value analysis, based on the peaks over thresholds theory, is to fix some high threshold, which can also depend on the spatial location $\mathbf{s}$, and model the positive excesses over the threshold with a generalized Pareto distribution (GPD).

Various extensions of the GPD have been proposed to avoid the explicit selection of a threshold by modelling instead all the data, while trying to maintain a good fit on the tails (Naveau and Segers, 2024). These extensions reduce the uncertainty brought into the model by the selection of the threshold, and allow to model the entire marginal distribution of the data.

We use the Extended Generalized Pareto Distribution (EGPD) Naveau et al. (2016) as a model for F_Y that complies with extreme value theory in both the upper and lower tails. Its cdf is defined as

$$F_Y(y) = B\left(H_\xi\left(y/\sigma\right)\right), \tag{15}$$

where $H_\xi(\cdot) = 1-(1+\xi x)^{-1/\xi}$, for $\xi \neq 0$, is the cdf of a GPD, with shape parameter ξ. Here $\sigma > 0$ is a scale parameter and $B(\cdot)$ is a continuous cdf on $[0,1]$ satisfying the following three conditions for three positive constants a, b and c: $\lim_{u\to 0} \bar{B}(1-u)/u = a$; $\lim_{u\to 0} B(u\,w(u))/B(u) = b$ where $w(\cdot)$ is any positive function such that $w(u) = 1 + o(u)$ as $u \to 0$; $\lim_{u\to 0} B(u)/u^\kappa = c$ with $\kappa \geq 0$.

We choose the parametric specification $B(u) = pu^{\kappa_1} + (1-p)u^{\kappa_2}$, with $p \in [0,1]$ and $\kappa_2 \geq \kappa_1 > 0$ (Naveau et al., 2016). In this case, the lower tail behaviour is controlled by κ_1, while κ_2 modifies the shape of the density in its central part. This specification permits possible bimodality in the marginal distribution of the data, as illustrated in Figure 5. This bimodality is likely due to the way the FWI is constructed; high-intensity rainfall events cause the index value to approach zero.

The marginal model presented above has five parameters, σ, ξ, p, κ_1 and κ_2. We account for the spatial non-stationarity of the data, visible in Figure 4, by letting parameters σ, ξ and p in (15) vary according to smooth functions of the coordinate $\mathbf{s}$, leading to a cdf $F_Y^{\mathbf{s}}$ of $Y(\mathbf{s})$ varying with $\mathbf{s}$. The estimation of these functions using a penalized likelihood approach based on regression splines is implemented in the experimental version of the R package `evgam` (Youngman, 2022), available at github.com/byoungman/evgam. The extra flexibility seems able to capture the non-stationary behaviour of the data, as shown in Figure 5 for three spatial locations that are far from each other, and also in Figure S14 in the supplementary material, where we compare empirical and fitted quantiles of order $u \in \{0.5, 0.75, 0.9, 0.95\}$ for all the locations in the north of Portugal. Therefore we use the estimates of F_Y^s to transform the data as in (13), and then use the resulting vectors $\boldsymbol{u}_i$, $i = 1, \ldots, n$, in Algorithm 2 to fit the dependence model parameters.

7.3. Dependence modelling

We consider several candidate models in the class of Gaussian location-scale mixtures: the Gaussian location mixture 1 (LM1), the Gaussian scale mixture 5 (SM5), and the Gaussian location-scale mixture 2 (LSM2). The parameter estimation process took approximately 15 minutes for the

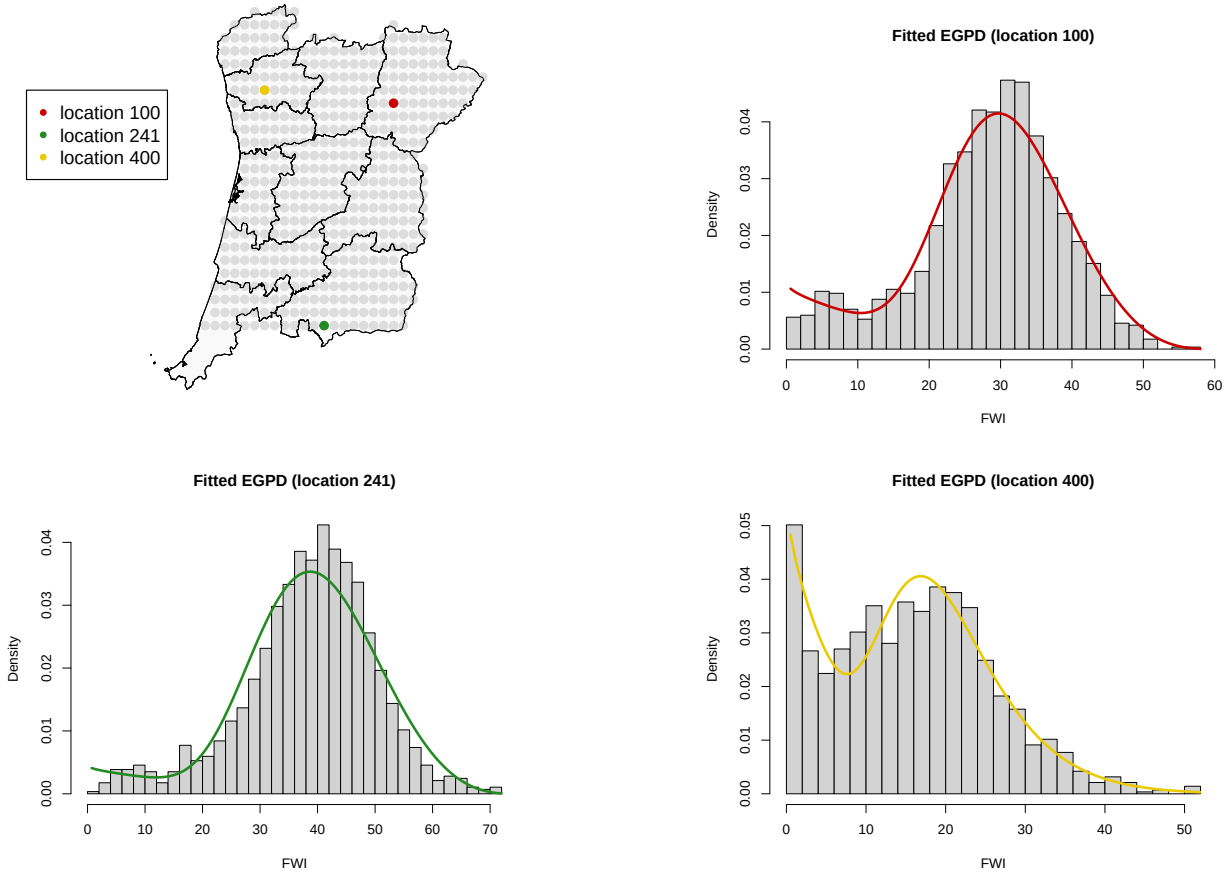


Figure 5: Three locations in the north of Portugal and the corresponding FWI histograms and fitted EGPD densities, with spatially-varying parameters σ, ξ and p.

three-parameter models LM1 and SM5, and three hours for the four-parameter model LSM2. This can be compared with the copula cases in Table 2). The estimated parameter values for these models and for a Gaussian process are reported in Table 3. Note that these results indicate, for the upper tail of the data distribution, compatibility with both asymptotic dependence and asymptotic independence. Indeed, the confidence interval for γ in model SM5, which includes 0, and that for λ_1 in model LSM2, which includes 1, do not show a significant prevalence of one extremal dependence class.

Figure 6 shows the empirical coefficients $\chi(u)$ for $u = 0.5, 0.75, 0.9, 0.95, 0.975, 0.99$ and for different distances in space (data are summarized by boxplots for bins of size 25 km), and compares them with the values of the coefficient obtained by simulation from the fitted models. The standard Gaussian process underestimates the risk of bivariate occurrences of extreme values, while models LM1 and SM5 seem to be more appropriate for describing the behaviour of data at high quantiles, i.e. for $u \geq 0.9$. Model LSM2, which shows the best fit for central quantiles, corresponding to $u = 0.5$ and $u = 0.75$, overestimates the coefficient χ for higher quantiles. If the analysis focuses on wildfire-prone conditions corresponding to relatively extreme fire-weather index values, LM1 and SM5 are the most appropriate models for representing spatial clustering of extreme events.

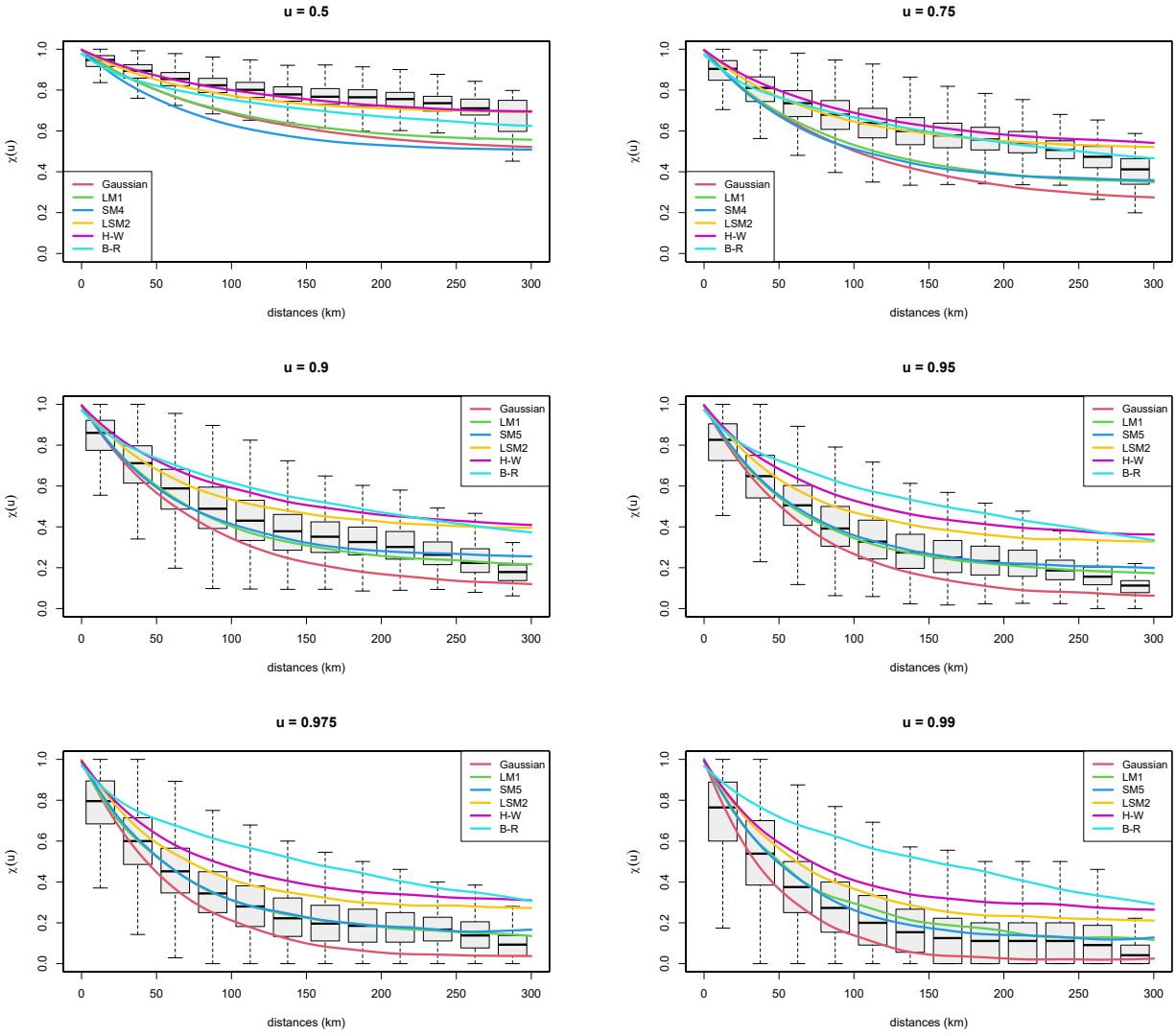


Figure 6: Empirical bivariate $\chi(u)$ for pairs at different distances and thresholds $u \in \{0.5, 0.75, 0.9, 0.95, 0.975, 0.99\}$ and fitted models: Gaussian process (red lines), Gaussian location mixture 1 (green lines), Gaussian scale mixture 5 (blue lines), Gaussian location-scale mixture 2 (yellow lines), Huser-Wadsworth model (purple lines) and Brown-Resnick process (light blue lines).

Table 3: Estimated parameter values (and corresponding 0.95 bootstrap confidence intervals) for a Gaussian process and models LM1, SM5, LSM2.

Model	$\widehat{\theta}_W$	$\widehat{\theta}_{S,R}$
Gaussian	$\widehat{\varphi} = 117.51$ $(116.87, 118.27)$, $\widehat{\eta} = 1.20$ $(1.20, 1.21)$	-
LM1	$\widehat{\varphi} = 106.46$ $(105.90, 107.49)$, $\widehat{\eta} = 1.20$ $(1.19, 1.20)$	$\widehat{\lambda} = 2.34$ $(2.14, 2.64)$
SM5	$\widehat{\varphi} = 86.22$ $(85.24, 87.35)$, $\widehat{\eta} = 1.31$ $(1.31, 1.32)$	$\widehat{\gamma} = 0.02$ $(-0.13, 0.17)$
LSM2	$\widehat{\varphi} = 94.73$ $(93.95, 107.21)$, $\widehat{\eta} = 1.28$ $(1.26, 1.28)$	$\widehat{\lambda}_1 = 1.10$ $(0.99, 1.24)$, $\widehat{\lambda}_2 = 0.85$ $(0.79, 0.94)$

Table 4: Negative log-likelihood of the observed threshold exceedance counts $N_i(p)$, for the six models shown in Figure 6.

Model	Gaussian	LM1	SM5	LSM2	H-W	B-R
nll ($p = 0.90$)	4546.3	4498.7	**4420.0**	4432.6	4457.9	4455.3
nll ($p = 0.95$)	3154.2	3114.9	**3087.8**	3099.4	3101.5	3129.4
nll ($p = 0.99$)	1184.3	**1168.7**	1170.9	1175.3	1171.1	1217.1

In this context, model selection should depend on the applied purpose of the analysis, as well as considerations about data features and the size of the dataset. For example, if the data exhibits tail asymmetry with slower-than-Gaussian joint tail decay, this can only be captured by Gaussian location-scale mixtures such as LSM2, rather than LM1 or SM5. Conversely, if only upper-tail dependence is of interest and/or a smaller dataset is available, more parsimonious models such as LM1 and SM5 may be preferred.

For comparison, two additional copula models were fitted to the data and the resulting $\chi(u)$ displayed in Figure 6. These models are a Huser-Wadsworth random-scale model (H-W; Huser and Wadsworth, 2019) and a max-stable Brown-Resnick process (B-R; Brown and Resnick, 1977). Both models overestimate the dependence at high quantiles, and the latter seems too rigid to capture the decay of $\chi(u)$ values at increasing spatial distances. Section S5.1 of the supplementary material provides details on the adopted inferential methods and the estimated parameter values.

In addition to the bivariate dependence diagnostics shown in Figure 6, we can evaluate how well the models reproduce the spatial extent of extreme events. Given that we have $n = 1426$ daily observations across $m = 491$ spatial locations, we compute $N_i(p)$, the number of locations at which values exceeding the $p^{(th)}$ quantile are observed on day i, for $i = 1, \ldots, n$. We assume that the $N_i(p)$ are i.i.d. copies of the $N(p)$ a discrete random variable taking values in $\{0, 1, \ldots, m\}$. The distribution of this variable differs for the estimated models and for p and it can be approximated by Monte Carlo. Table 4 shows the values of the negative log-likelihood values for different model specification for $p \in \{0.90, 0.95, 0.99\}$. Models LM1 and SM5 produce the best results, confirming their ability to capture the spatial extent of extremal dependence, as shown in Figure 6.

7.4. Conditional simulation

The conditional simulation algorithm presented in Section 4 is applied to the FWI data in the north of Portugal, as illustrated in Figure 7. In particular, data from the 1st of July, 2000, are analysed. The Viseu district is assumed to be unobserved, and its observations are left out from the dataset; therefore, in the following $\boldsymbol{X}_2$ and $\mathbf{Y}_2$ refer to the Viseu district, while $\boldsymbol{X}_1$ and $\mathbf{Y}_1$ refer to the other districts in the north of Portugal. The model assumed for the dependence structure

of the data is Gaussian location mixture 1, with the parameter values estimated in Section 7.3. The Metropolis-Hastings simulation scheme (see 1. in Algorithm 1) involved $k = 3 \cdot 10^5$ simulated values of $S \mid \boldsymbol{X}_1$, with a thinning rate of 100. Since multiple values of S are generated in this way, the entire conditional distribution of $\boldsymbol{X}_2 \mid \boldsymbol{X}_1$ is simulated. The corresponding distribution of $\mathbf{Y}_2 \mid \mathbf{Y}_1$, on the FWI scale, is obtained by applying $\widehat{F}_Y^{\mathbf{s}\,-1}\{\widehat{F}_X(\cdot)\}$ to it, with the spatially-varying $\widehat{F}_Y^{\mathbf{s}}$ estimated in Section 7.2 and $\widehat{F}_X$ estimated in Section 7.3. To obtain point predictions, these distributions are summarized by their medians. In particular, the predicted values of $\mathbf{Y}_{2,i}$, where i refers to the day (1st of July, 2000), are compared to the observed ones in the Viseu district, showing a high degree of similarity. Since this may be due to the spatially varying marginal distribution, Figure 7 also shows the same comparison on the uniform scale, that is, on the scale of $\widehat{F}_X(\boldsymbol{X}_2)$. The simulated data on this scale are independent of the marginal distribution, but the predicted values are still very close to the observed ones.

By repeating the above procedure for each of the $n = 1426$ daily observations, and taking the median value of the conditionally simulated distribution of $\mathbf{Y}_2 \mid \mathbf{Y}_1$ as the point prediction at each location in the Viseu district, we obtained an RMSE of 1.09. For comparison, predictions obtained using ordinary Kriging based on an estimated variogram fitted with a Matérn covariance model yielded an RMSE of 1.74. The use of an intermediate method, which involves the estimated marginal EGPD being used to transform the data into a uniform scale and then into a standard Gaussian scale, and the application of ordinary Kriging to the transformed data, results in an RMSE of 1.54.

8. Conclusions

The objective of this work was threefold. First, a review of the key properties of Gaussian location-scale mixtures was conducted. These mixtures constitute a general class that includes the known subclasses of Gaussian location mixtures and Gaussian scale mixtures. The capacity to define both a random scale and a random location introduces enhanced flexibility into the model's formulation, particularly concerning extremal dependence in both the upper and lower tail. In this framework, 13 models are proposed and studied with respect to their extremal dependence structure and, when available, their marginal and multivariate distributions. A subset of these models have been previously examined in the existing literature; however, others constitute novel propositions. In the latter case, the presented results are supported by analytical proofs. For this class of models, a conditional simulation algorithm is proposed.

Secondly, we propose a method for inference that is both general, being valid for any specified distribution of the random scale and the random location, and fast and stable, not involving numerical integrations. Our inferential approach could also be useful for multivariate (not spatial) models that make use of location and scale mixtures of multivariate Gaussian distributions.

Thirdly, we leverage the flexibility of these models, in conjunction with an inferential method that does not involve any censoring, to propose an extended modelling of both the tails and the bulk of the data distribution. Consequently, an extended modelling approach is adopted for marginal distributions as well, based on the extended generalized Pareto distribution.

A possible extension would be to model also temporal dependence, defining Gaussian location-scale mixtures as space-time models. However, the construction of these models would not be straightforward. A promising idea is to adapt the proposed inference technique to space-time location-mixture models as $X(\mathbf{s}, t) = S(t) + W(\mathbf{s}, t)$, for $\mathbf{s} \in \mathcal{S},\ t \in \mathcal{T}$, where $S(t)$ is a time process

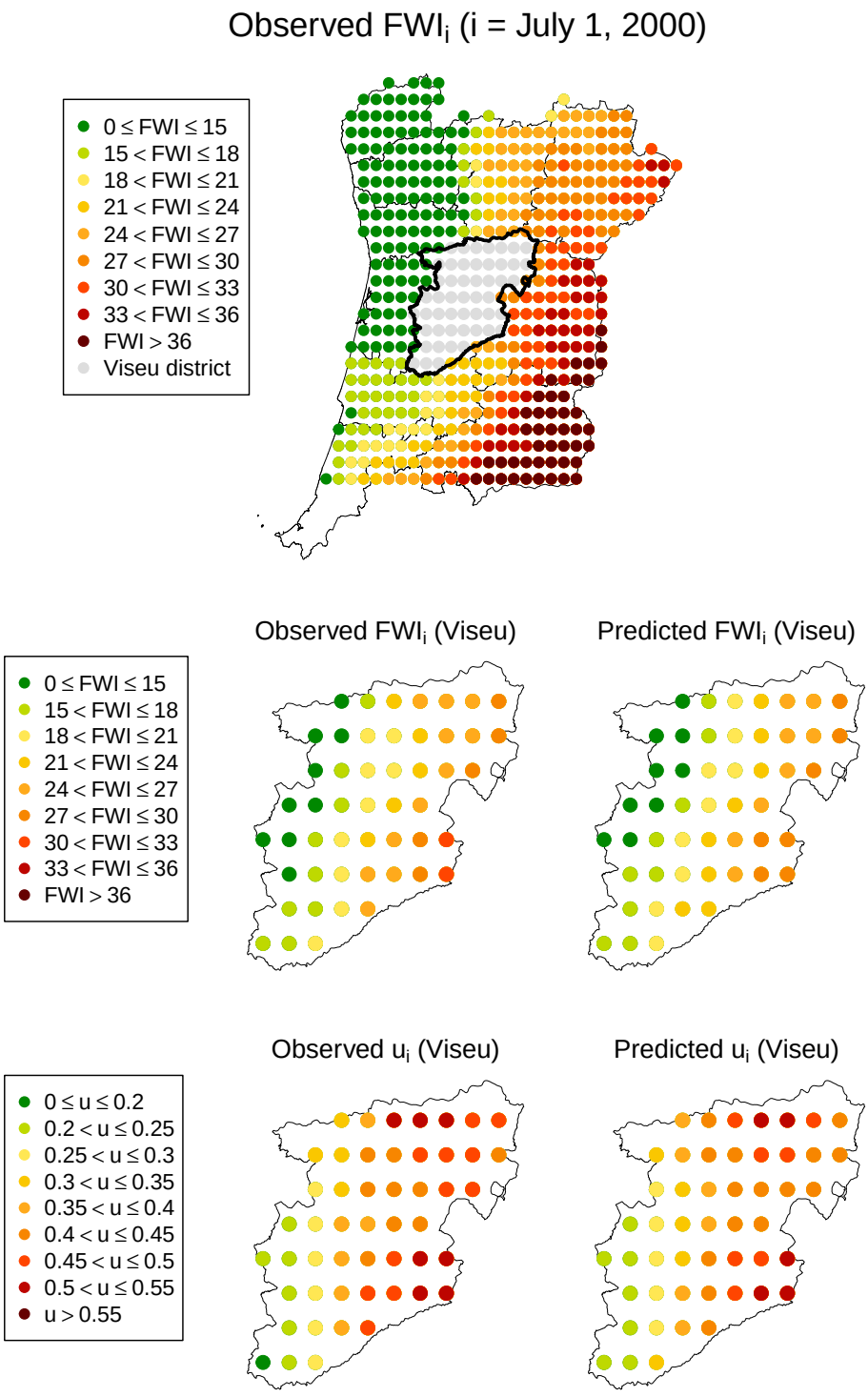


Figure 7: Conditional simulation described in Section 7.4. The top figure displays the observed FWI values on July 1st, 2000, in the north of Portugal, excluding the Viseu district. The central and bottom figures show the observed and predicted values in the Viseu district on the FWI and uniform scale, respectively.

independent of the space-time Gaussian process $W(\mathbf{s}, t)$. Existing inference approaches are either prohibitive for this setting or come with an extremely high computational cost. By adapting our transformation trick, we can consider $\{X(\mathbf{s}, t) - X(\mathbf{s}_1, t)\} = \{W(\mathbf{s}, t) - W(\mathbf{s}_1, t)\}$, for $\mathbf{s}_1 \in \mathcal{S}$, which is a Gaussian process. The specification of the process $S(t)$ is not obvious because it must consider inferential aspects while maintaining the realism of the model when applied to real data. Moreover, it would be necessary to develop different techniques for the cases of Gaussian scale and location-scale mixtures, since the corresponding transformations would not lead to processes with closed-form likelihoods.

Acknowledgments

The authors thank Raphaël Huser and Likun Zhang for their comments on a previous version of this work.

Appendix A. Extremal behaviour of SM4, LSM1

Appendix A.1. Proof for SM4 model

Let $X(\mathbf{s}) = RW(\mathbf{s})$ with $R = \sqrt{E}/G$, where $E \sim \mathrm{E}(1/2)$ and $G \sim \mathrm{Gamma}(1/\gamma, 1/\gamma)$. The density function of the inverse Gamma random variable $1/G$ is regularly varying with index $-(1/\gamma + 1)$, so, by Karamata's theorem (see Bingham et al., 1989), its distribution function is regularly varying with index $-1/\gamma$ (see Section 2). Then, since $1/G$ is heavier tailed than $\sqrt{E}$, which follows a Rayleigh distribution, by Breiman's lemma (Breiman, 1965) R is also regularly varying with index $1/\gamma$. Following Section 3.1 of Huser et al. (2017), we have that $\{X(\mathbf{s})\}$ is asymptotically dependent with

$$\chi = 2T_{1/\gamma+1}\left[-\sqrt{(1/\gamma+1)(1-\rho)/(1+\rho)}\right].$$

□

Appendix A.2. Proof for LSM1 model

Let $X(\mathbf{s}) = S + RW(\mathbf{s})$ with $S \sim \mathrm{E}(\lambda)$, and $R = \sqrt{E}$, with $E \sim \mathrm{E}(1/2)$.
For $\lambda < 1$, we need to prove that $\{X(\mathbf{s})\}$ is AD, with

$$\chi = \mathbb{E}\left[\min\left\{\frac{Z_1^\lambda}{\mathbb{E}(Z_1^\lambda)}, \frac{Z_2^\lambda}{\mathbb{E}(Z_2^\lambda)}\right\}\right] = \mathbb{E}\left[\min\left\{\frac{Z_1^\lambda}{(1-\lambda^2)^{-1}}, \frac{Z_2^\lambda}{(1-\lambda^2)^{-1}}\right\}\right] = (1-\lambda^2)\mathbb{E}\left[\min\{Z_1, Z_2\}^\lambda\right], \tag{A.1}$$

where $Z_i = \exp\{R\, W(\mathbf{s}_i)\} \sim \mathrm{LogLaplace}(1)$, $i = 1, 2$ and $\bar{\chi} = 1$. To prove this result, consider the exponential transformation

$$\exp\{X(\mathbf{s})\} = \exp\{S\}\exp\{R\, W(\mathbf{s})\},\ \mathbf{s} \in \mathcal{S},$$

that defines a scale mixture with scaling variable $\exp\{S\} \sim \mathrm{Pareto}(\lambda)$ and $\exp\{RW(\mathbf{s})\} \sim \mathrm{LogLaplace}(1)$. Therefore, Proposition 1 of Engelke et al. (2019) provides the expression (A.1), since the χ index is invariant to monotonically increasing transformations of the random variable, such as the logarithmic one.
For $\lambda > 1$, we need to prove that $\{X(\mathbf{s})\}$ is AI, i.e. $\chi = 0$, with

$$\bar{\chi} = \max\left\{\frac{2}{\lambda} - 1,\ \rho\right\},$$

where $\rho = \text{cor}(W(\mathbf{s}_1), W(\mathbf{s}_2))$. To prove this result, consider again the multiplicative representation $\exp\{X(\mathbf{s})\}$ and see Engelke et al. (2019), Proposition 5.
For $\lambda = 1$, $\{X(\mathbf{s})\}$ is asymptotically independent, with $\chi = 0$ and $\bar{\chi} = 1$, as is possible to prove by considering the same multiplicative representation and following Engelke et al. (2019), Proposition 6, case 3(c). □

Appendix B. Details of the Metropolis-Hastings algorithm for conditional simulation

The acceptance probability in (5), for each step of the Metropolis-Hastings algorithm, is

$$\rho(s', r', s^{(k)}, r^{(k)}) = \min\left\{ \frac{f_{S,R}(s', r')\phi_m((\boldsymbol{x}_1 - s')/r'; \boldsymbol{\Sigma}_{1;1}) r'^{-m} g(r^{(k)}, s^{(k)} | r', s')}{f_{S,R}(s^{(k)}, r^{(k)})\phi_m((\boldsymbol{x}_1 - s^{(k)})/r^{(k)}; \boldsymbol{\Sigma}_{1;1}) r^{(k)-m} g(r', s' | r^{(k)}, s^{(k)})}, 1 \right\},$$

where $g(a, b \,|\, c, d)$ is the sampling density from (c, d) to (a, b); note that

$$g(r^{(k)}, s^{(k)} | r', s') = g(r^{(k)} | r') g(s^{(k)} | s') = \phi(\log(r^{(k)}) - \log(r')) \frac{1}{r^{(k)}} \phi(s^{(k)} - s'),$$

$$g(r', s' | r^{(k)}, s^{(k)}) = g(r' | r^{(k)}) g(s' | s^{(k)}) = \phi(\log(r') - \log(r^{(k)})) \frac{1}{r'} \phi(s' - s^{(k)}),$$

and, due to the symmetry of the Gaussian density function of the random walk proposal, all terms but the Jacobian $1/r^{(k)}$ and $1/r'$ simplify, leading to:

$$\rho(s', r', s^{(k)}, r^{(k)}) = \min\left\{ \frac{f_{S,R}(s', r')\phi_m((\boldsymbol{x}_1 - s')/r'; \boldsymbol{\Sigma}_{1;1}) r'^{-m+1}}{f_{S,R}(s^{(k)}, r^{(k)})\phi_m((\boldsymbol{x}_1 - s^{(k)})/r^{(k)}; \boldsymbol{\Sigma}_{1;1}) r^{(k)-m+1}}, 1 \right\}.$$

Appendix C. Derivation of density functions for the vectors $\boldsymbol{Z}$

Appendix C.1. Gaussian location mixtures

The vector (7) can be written as

$$\boldsymbol{Z} = (X(\mathbf{s}_2) - X(\mathbf{s}_1), \ldots, X(\mathbf{s}_m) - X(\mathbf{s}_1))^\top = (W(\mathbf{s}_2) - W(\mathbf{s}_1), \ldots, W(\mathbf{s}_m) - W(\mathbf{s}_1))^\top.$$

Note that this is a linear combination of the components of a multivariate Gaussian, so $\boldsymbol{Z} \sim \mathcal{N}_{m-1}(\mathbf{0}, \boldsymbol{A}\boldsymbol{\Sigma}_{\theta_W}\boldsymbol{A}^\top)$, where $\boldsymbol{A}$ is the $(m-1) \times m$ matrix in (8), and its density function, for $\boldsymbol{z} \in \mathbb{R}^{m-1}$, is

$$f_{\boldsymbol{Z}}(\boldsymbol{z}; \theta_W) = \phi_{m-1}(\boldsymbol{z}; \boldsymbol{A}\boldsymbol{\Sigma}_{\theta_W}\boldsymbol{A}^\top) = (2\pi)^{-(m-1)/2} |\boldsymbol{A}\boldsymbol{\Sigma}_{\theta_W}\boldsymbol{A}^\top|^{-1/2} \exp\left(-\frac{1}{2} \boldsymbol{z}^\top (\boldsymbol{A}\boldsymbol{\Sigma}_{\theta_W}\boldsymbol{A}^\top)^{-1} \boldsymbol{z} \right).$$

Appendix C.2. Gaussian scale mixtures

The vector (9) can be written as

$$\boldsymbol{Z} = \left(\frac{X(\mathbf{s}_2)}{X(\mathbf{s}_1)}, \ldots, \frac{X(\mathbf{s}_m)}{X(\mathbf{s}_1)} \right)^\top = \left(\frac{W(\mathbf{s}_2)}{W(\mathbf{s}_1)}, \ldots, \frac{W(\mathbf{s}_m)}{W(\mathbf{s}_1)} \right)^\top.$$

To obtain the density function of $\boldsymbol{Z}$, we can first define another vector

$$\widetilde{\boldsymbol{Z}} = \left(W(\mathbf{s}_1), \frac{W(\mathbf{s}_2)}{W(\mathbf{s}_1)}, \ldots, \frac{W(\mathbf{s}_m)}{W(\mathbf{s}_1)} \right)^\top.$$

The density function of $\widetilde{\boldsymbol{Z}}$ is $f_{\widetilde{\boldsymbol{Z}}}(\widetilde{\boldsymbol{z}};\,\theta_W) = \phi_m(\widetilde{z}_1, \widetilde{z}_1\widetilde{z}_2, \ldots, \widetilde{z}_1\widetilde{z}_m; \boldsymbol{\Sigma}_{\theta_W})\,|\widetilde{z}_1|^{m-1}$, where $\widetilde{z}_j$, $j = 1, \ldots, m$, is the j-th component of $\widetilde{\boldsymbol{z}} = (\widetilde{z}_1, \ldots, \widetilde{z}_m)^\top$. Defining also $\boldsymbol{z} = (\widetilde{z}_2, \ldots, \widetilde{z}_m)^\top$ and $\dot{\boldsymbol{z}} = (1, \widetilde{z}_2, \ldots, \widetilde{z}_m)^\top = (1, \boldsymbol{z}^\top)^\top$, we can obtain the density function of $\boldsymbol{Z}$ as

$$
\begin{aligned}
f_{\boldsymbol{Z}}(\boldsymbol{z};\,\theta_W) &= \int_{-\infty}^{\infty} f_{\widetilde{\boldsymbol{Z}}}(\widetilde{\boldsymbol{z}};\,\theta_W) d\widetilde{z}_1 = \int_{-\infty}^{\infty} \phi_m(\widetilde{z}_1, \widetilde{z}_1\widetilde{z}_2, \ldots, \widetilde{z}_1\widetilde{z}_m; \boldsymbol{\Sigma}_{\theta_W})\; |\widetilde{z}_1|^{m-1} d\widetilde{z}_1 \\
&= \int_{-\infty}^{\infty} (2\pi)^{-m/2} |\boldsymbol{\Sigma}_{\theta_W}|^{-1/2} \exp\left(-\frac{1}{2}\widetilde{z}_1^2\; \dot{\boldsymbol{z}}^\top \boldsymbol{\Sigma}_{\theta_W}^{-1} \dot{\boldsymbol{z}}\right) |\widetilde{z}_1|^{m-1} d\widetilde{z}_1 \\
&= 2\int_{0}^{\infty} (2\pi)^{-m/2} |\boldsymbol{\Sigma}_{\theta_W}|^{-1/2} \exp\left(-\frac{1}{2} y\; \dot{\boldsymbol{z}}^\top \boldsymbol{\Sigma}_{\theta_W}^{-1} \dot{\boldsymbol{z}}\right) y^{(m-1)/2} \frac{1}{2} y^{-1/2} dy \qquad [\text{setting } y = \widetilde{z}_1^2] \\
&= (2\pi)^{-m/2} |\boldsymbol{\Sigma}_{\theta_W}|^{-1/2} \int_{0}^{\infty} \exp\left(-y\; \dot{\boldsymbol{z}}^\top \boldsymbol{\Sigma}_{\theta_W}^{-1} \dot{\boldsymbol{z}}/2\right) y^{m/2-1} dy \\
&= (2\pi)^{-m/2} |\boldsymbol{\Sigma}_{\theta_W}|^{-1/2} \frac{\Gamma\left(m/2\right)}{\left(\dot{\boldsymbol{z}}^\top \boldsymbol{\Sigma}_{\theta_W}^{-1} \dot{\boldsymbol{z}}/2\right)^{m/2}} \int_{0}^{\infty} \frac{\left(\dot{\boldsymbol{z}}^\top \boldsymbol{\Sigma}_{\theta_W}^{-1} \dot{\boldsymbol{z}}/2\right)^{m/2}}{\Gamma\left(m/2\right)} \exp\left(-y\; \dot{\boldsymbol{z}}^\top \boldsymbol{\Sigma}_{\theta_W}^{-1} \dot{\boldsymbol{z}}/2\right) y^{m/2-1} dy \\
&= \pi^{-m/2} |\boldsymbol{\Sigma}_{\theta_W}|^{-1/2} \Gamma\left(m/2\right) \left(\dot{\boldsymbol{z}}^\top \boldsymbol{\Sigma}_{\theta_W}^{-1} \dot{\boldsymbol{z}}\right)^{-m/2},
\end{aligned}
$$

where the last integration for y comes from the gamma density.

Appendix C.3. Gaussian location-scale mixtures

The vector (10) can be written as

$$
\boldsymbol{Z} = \left(\frac{X(\mathbf{s}_3) - X(\mathbf{s}_1)}{X(\mathbf{s}_2) - X(\mathbf{s}_1)}, \ldots, \frac{X(\mathbf{s}_m) - X(\mathbf{s}_1)}{X(\mathbf{s}_2) - X(\mathbf{s}_1)}\right)^\top = \left(\frac{W(\mathbf{s}_3) - W(\mathbf{s}_1)}{W(\mathbf{s}_2) - W(\mathbf{s}_1)}, \ldots, \frac{W(\mathbf{s}_m) - W(\mathbf{s}_1)}{W(\mathbf{s}_2) - W(\mathbf{s}_1)}\right)^\top.
$$

Note that, to compute the density function of $\boldsymbol{Z}$, we can combine the formulas of the two previous cases, since this vector can be seen as a vector of ratios of differences of the elements of $\boldsymbol{W}$. From the location mixture case above, we know that $(W(\mathbf{s}_2) - W(\mathbf{s}_1), \ldots, W(\mathbf{s}_m) - W(\mathbf{s}_1)) \sim \mathcal{N}_{m-1}(\mathbf{0}, \boldsymbol{A}\boldsymbol{\Sigma}_{\theta_W}\boldsymbol{A}^\top)$, where $\boldsymbol{A}$ is the $(m-1) \times m$ matrix in (8). Then, we can use the results on the ratio of components of a multivariate Gaussian, shown in the scale mixture case, to obtain

$$
f_{\boldsymbol{Z}}(\boldsymbol{z};\,\theta_W) = \pi^{-(m-1)/2} |\boldsymbol{A}\boldsymbol{\Sigma}_{\theta_W}\boldsymbol{A}^\top|^{-1/2} \Gamma\left(\frac{m-1}{2}\right) \left[\dot{\boldsymbol{z}}^\top (\boldsymbol{A}\boldsymbol{\Sigma}_{\theta_W}\boldsymbol{A}^\top)^{-1} \dot{\boldsymbol{z}}\right]^{-(m-1)/2},
$$

where $\boldsymbol{z} \in \mathbb{R}^{m-2}$ and $\dot{\boldsymbol{z}} = (1, \boldsymbol{z}^\top)^\top \in \mathbb{R}^{m-1}$.

References

Andrews, D. F. and Mallows, C. L. (1974). Scale mixtures of normal distributions. Journal of the Royal Statistical Society, Series B (Statistical Methodology), 36(1):99–102.

Arellano-Valle, R. B. and Azzalini, A. (2021). A formulation for continuous mixtures of multivariate normal distributions. Journal of Multivariate Analysis, 185:104780.

Armagan, A., Dunson, D. B., and Lee, J. (2013). Generalized double Pareto shrinkage. Statistica Sinica, 23(1):119.

Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. Journal of the Royal Statistical Society, Series B (Statistical Methodology), 65(2):367–389.

Azzalini, A. and Capitanio, A. (2014). The Skew-Normal and Related Families. Cambridge University Press, Cambridge.

Azzalini, A. and Dalla Valle, A. (1996). A class of multivariate skew-normal distributions. Scandinavian Journal of Statistics, 23(4):419–430.

Bacro, J. N., Gaetan, C., and Toulemonde, G. (2016). A flexible dependence model for spatial extremes. Journal of Statistical Planning and Inference, 172:36–52.

Barndorff-Nielsen, O., Kent, J., and Sørensen, M. (1982). Normal variance-mean mixtures and z distributions. International Statistical Review/Revue Internationale de Statistique, 50(2):145–159.

Barndorff-Nielsen, O. E. (1977). Infinitely Divisible Distributions. Springer, New York.

Bingham, N. H., Goldie, C. M., and Teugels, J. L. (1989). Regular variation. Cambridge University Press, Cambridge.

Blei, D. M., Kucukelbir, A., and McAuliffe, J. D. (2017). Variational inference: A review for statisticians. Journal of the American Statistical Association, 112(518):859–877.

Bolthausen, E. (1977). Convergence in distribution of minimum-distance estimators. Metrika, 24(1):215–227.

Breiman, L. (1965). On some limit theorems similar to the arc-sin law. Theory of Probability & Its Applications, 10:323–331.

Brown, B. and Resnick, S. (1977). Extremes values of independent stochastic processes. Journal of Applied Probability, 14:732–739.

Coles, S., Heffernan, J., and Tawn, J. (1999). Dependence measures for extreme value analyses. Extremes, 2:339–365.

Davison, A. C. (2003). Statistical Models. Cambridge university press, Cambridge.

de Fondeville, R. and Davison, A. C. (2018). High-dimensional peaks-over-threshold inference. Biometrika, 105(3):575–592.

de Haan, L. (1984). A spectral representation for max-stable processes. The Annals of Probability, 12:1194–1204.

Dell'Oro, L. and Gaetan, C. (2025). Flexible space–time models for extreme data. Spatial Statistics, 68:100916.

Demarta, S. and McNeil, A. J. (2005). The t copula and related copulas. International Statistical Review, 73(1):111–129.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 39(1):1–22.

Engelke, S., Opitz, T., and Wadsworth, J. L. (2019). Extremal dependence of random scale constructions. Extremes, 22:623–666.

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

Fischer, A., Gaunt, R. E., and Sarantsev, A. (2025). The variance-gamma distribution: a review. Statistical Science, 40(2):235–258.

Frühwirth-Schnatter, S. (2006). Finite Mixture and Markov Switching Models. Springer, New York.

Fuentes, M. (2002). Spectral methods for nonstationary spatial processes. Biometrika, 89(1):197–210.

Gong, G. and Samaniego, F. J. (1981). Pseudo maximum likelihood estimation: theory and applications. The Annals of Statistics, 9(4):861–869.

Gong, Y. and Huser, R. (2022a). Asymmetric tail dependence modeling, with application to cryptocurrency market data. The Annals of Applied Statistics, 16(3):1822–1847.

Gong, Y. and Huser, R. (2022b). Flexible modeling of multivariate spatial extremes. Spatial Statistics, 52:100713.

Huser, R., Opitz, T., and Thibaud, E. (2017). Bridging asymptotic independence and dependence in spatial extremes using Gaussian scale mixtures. Spatial Statistics, 21:166–186.

Huser, R., Opitz, T., and Wadsworth, J. L. (2025). Modeling of spatial extremes in environmental data science: time to move away from max-stable processes. Environmental Data Science, 4:e3.

Huser, R. and Wadsworth, J. L. (2019). Modeling spatial processes with unknown extremal dependence class. Journal of the American Statistical Association, 114:434–444.

Joe, H. (1997). Multivariate Models and Multivariate Dependence Concepts. Chapman and Hall/CRC, New York.

Kabluchko, Z., Schlather, M., and de Haan, L. (2009). Stationary max-stable fields associated to negative definite functions. The Annals of Probability, 37:2042–2065.

Kotz, S., Kozubowski, T., and Podgórski, K. (2001). The Laplace Distribution and Generalizations. Birkäuser, Boston.

Krupskii, P. (2026). A class of skew-multivariate distributions for spatial data. arXiv preprint arXiv:2601.19049.

Krupskii, P., Huser, R., and Genton, M. G. (2018). Factor copula models for replicated spatial data. Journal of the American Statistical Association, 113(521):467–479.

Lee, S. X. and McLachlan, G. J. (2021). On mean and/or variance mixtures of normal distributions. In Statistical Learning and Modeling in Data Analysis, pages 117–127. Springer, Cham.

Madan, D. B. and Seneta, E. (1990). The variance gamma (VG) model for share market returns. Journal of Business, 63(4):511–524.

McLachlan, G. J. and Peel, D. (2000). Finite Mixture Models. Wiley, New York.

Moghli, A., Opitz, T., Ruffault, J., Pimont, F., Brotons, L., Dupuy, J. L., and Duane, A. (2025). Future landscapes scenarios, alternatives and EWE evolution. Deliverable D1.8 FIRE-RES project. DOI: 10.5281/zenodo.16602986.

Morris, S. A., Reich, B. J., Thibaud, E., and Cooley, D. (2017). A space-time skew-t model for threshold exceedances. Biometrics, 73:749–758.

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

Naveau, P. and Segers, J. (2024). Multivariate extreme value theory. arXiv preprint arXiv:2412.18477. To appear in Handbook on Statistics of Extremes, Chapman & Hall/CRC.

Neves, A. K., Campagnolo, M. L., Silva, J. M., and Pereira, J. M. (2023). A Landsat-based atlas of monthly burned area for Portugal, 1984–2021. International Journal of Applied Earth Observation and Geoinformation, 119:103321.

Opitz, T. (2013). Extremal t processes: Elliptical domain of attraction and a spectral representation. Journal of Multivariate Analysis, 122:409–413.

Opitz, T. (2016). Modeling asymptotically independent spatial extremes based on Laplace random fields. Spatial Statistics, 16:1–18.

Parr, W. C. (1981). Minimum distance estimation: a bibliography. Communications in Statistics - Theory and Methods, 10(12):1205–1224.

Patterson, H. D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. Biometrika, 58(3):545–554.

Pollard, D. (1980). The minimum distance method of testing. Metrika, 27(1):43–70.

Reich, B. J., Fuentes, M., and Dunson, D. B. (2012). A flexible spatial mixture model for extreme precipitation. Journal of the American Statistical Association, 107(497):167–179.

Ribatet, M. (2022). SpatialExtremes: Modelling Spatial Extremes. R package version 2.1-0.

Robert, C. P. and Casella, G. (2004). Monte Carlo Statistical Methods. Springer, New York, 2nd edition.

Røislien, J. and Omre, H. (2006). T-distributed random fields: A parametric model for heavy-tailed well-log data. Mathematical Geology, 38:821–849.

Smith, R. L. (1990). Max-stable processes and spatial extremes. Unpublished manuscript, University of North Carolina.

Verde, J. C. and Zêzere, J. L. (2010). Assessment and validation of wildfire susceptibility and hazard in Portugal. Natural Hazards and Earth System Sciences, 10(3):485–497.

Wadsworth, J. L. and Tawn, J. A. (2012). Dependence modelling for spatial extremes. Biometrika, 99(2):253–272.

Wolfowitz, J. (1957). The minimum distance method. The Annals of Mathematical Statistics, 28(1):75–88.

Youngman, B. D. (2022). evgam: An R package for generalized additive extreme value models. Journal of Statistical Software, 103(3):1–26.

Zammit-Mangion, A., Sainsbury-Dale, M., and Huser, R. (2025). Neural methods for amortized inference. Annual Review of Statistics and Its Application, 12:311–335.

Zhang, L., Shaby, B. A., and Wadsworth, J. L. (2022a). Hierarchical transformed scale mixtures for flexible modeling of spatial extremes on datasets with many locations. Journal of the American Statistical Association, 117:1357–1369.

Zhang, Z., Huser, R., Opitz, T., and Wadsworth, J. L. (2022b). Modeling spatial extremes using normal mean-variance mixtures. Extremes, 25(2):175–197.

Zhao, Y. and Joe, H. (2005). Composite likelihood estimation in multivariate data analysis. Canadian Journal of Statistics, 33(3):335–356.