Environmetrics

WILEY

RESEARCH ARTICLE OPEN ACCESS

Spatial Modeling of Extremes and an Angular Component

G. Tamagny | M. Ribatet

Laboratoire de Mathématiques Jean Leray, LMJL, UMR 6629, Nantes Université, École Centrale Nantes, CNRS, Nantes, France

Correspondence: G. Tamagny (gaspard.tamagny@ec-nantes.fr)

Received: 20 April 2024 | **Revised:** 27 June 2025 | **Accepted:** 14 July 2025

Funding: The authors received no specific funding for this work.

Keywords: Bayesian hierarchical models | circular statistics | environmental application | extreme value theory | Markov Chain Monte Carlo | spatial extremes

ABSTRACT

Many environmental processes, such as rainfall, wind, or snowfall, are inherently spatial, and the modeling of extremes has to take into account that feature. In addition, such processes may be associated with a nonextremal feature, for example, wind speed and direction or extreme snowfall and time of occurrence in a year. This article proposes a Bayesian hierarchical model with a conditional independence assumption that aims at modeling simultaneously spatial extremes and an angular component. The proposed model relies on the extreme value theory as well as recent developments for handling directional statistics over a continuous domain. Working within a Bayesian setting, a Gibbs sampler is introduced whose performances are analysed through a simulation study. The paper ends with an application to extreme wind speed in France. Results show that extreme wind events in France are mainly coming from the West, apart from the Mediterranean part of France and the Alps.

1 | Introduction

The modeling of environmental extremes, such as floods, heat waves, or extreme wind events, is of paramount importance as a potential cause of severe damage to infrastructures or ecosystems. To better understand these climatic events and provide relevant guidelines to policy makers, the last decades have seen many theoretical and methodological developments for the statistical modeling of spatial extremes. For extreme-value processes, many approaches have been suggested ranging from Bayesian hierarchical models (Cooley et al. 2007; Ribatet et al. 2012), copula modeling (Nelsen 2006; Embrechts et al. 2001; Gudendorf and Segers 2010; Ribatet and Sedki 2013), max-stable, and Pareto processes (Schlather 2002; Padoan et al. 2010; Dombry and Ribatet 2015; Thibaud and Opitz 2015; de Haan 1984; de Haan and Ferreira 2014) or asymptotic independent models (Davison et al. 2013; Huser and Wadsworth 2022). Contrary to all the other aforementioned approaches, Bayesian hierarchical models (with the so- called conditional independence assumption) stand apart as they cannot model areal quantities as cumulative rainfall amount over a spatial domain. Although this might appear as a severe limitation, such models have been found to be very competitive and widely applicable as opposed, for instance, to max-stable or Pareto processes for which inference is difficult, and pointwise predictions are less accurate (Davison et al. 2012; Dombry et al. 2017; Huser et al. 2019). Davison et al. (2012, 2013) and Huser and Wadsworth (2022) give a comprehensive overview of available approaches and discuss the pros and cons of each approaches.

Despite all these significant advances, emphasis has often been put on extreme event magnitude only, that is, estimation of return levels, but, depending on the situation, this might be too limited. For instance, one may have to work in a multivariate setting to allow for the joint modeling of extreme wave heights and surge levels (Dixon et al. 1998). In this context, Genton et al. (2015)

and Hashorva and Kume (2021) propose a framework to model multivariate spatial extremes but as it fully complies with the extreme value theory, this framework may not be relevant whenever at least one component of the multivariate process cannot be treated as extreme. Still there exists case studies where one needs to model a multivariate process where some components are extremes and some are not. As an illustration, impacting their structural integrity and durability (Holmes 2015), the design of buildings or power lines must take into account for both magnitude and direction of extreme wind events. Clearly in this context, the wind speed is the extremal component while the wind direction is nonextremal. Similarly, the design of wind turbines and the layout of wind farms require careful consideration of extreme wind direction changes, as these can lead to significant load fluctuations on turbine blades, potentially causing mechanical failure or reducing efficiency (Sang et al. 2017). In France, the government has set ambitious targets for wind energy production as part of its transition to renewable energy sources (SFEC 2023) and it is therefore essential to understand and accurately model extreme wind phenomena, including both speed and direction.

As emphasized with the two previous examples, this work focuses on the special case where the nonextremal component is an angle. Since the pioneering review of Mardia (1972), the theoretical framework for directional data is now well established, see Mardia and Jupp (2009) for a comprehensive review. Still, the spatial modelling of an angular process is challenging as the classical circular distributions, for example, Von Mises or wrapped distributions do not easily extend to the spatial setting. Fortunately, and following the works of Gelfand and Wang (2013, 2014), the family of projected circular distributions appears to be especially appealing in this context as, as we will see later, one can work with bivariate Gaussian processes within an augmented data framework. Consequently, and although other approaches might be used (Casson and Coles 1998; Breckling 1989; Jona-Lasinio et al. 2012), throughout this paper, we will restrict our attention to projected Gaussian circular distributions.

The paper is organized as follows: Section 2 introduces two separate spatial models, one for the angular component of the data and the other for its extremal part. Section 3 explains how these two previous models can be merged into a single one to enable the joint modeling of the intensity and the angular component of spatial extreme events. Section 4 gives results of a simulation study, while Section 5 applies the proposed methodology to extreme wind speeds in France. The paper ends with a discussion. Specific details on how the proposed model is fitted from MCMC algorithms are deferred to the appendix.

2 | Independent Modeling of Spatial Extremes and Angles

Recall that our primary objective is to propose a model for a bivariate stochastic process $\{(\eta(s), \theta(s))^\top : s \in \mathcal{X}$, where the first component $\{\eta(s) : s \in \mathcal{X}\}$ is the extremal component, for example, extreme wind speed, and the second $\{\theta(s) : s \in \mathcal{X}\}$ the angular one, for example, associated wind direction. In this section, we will introduce successively a sensible model for each of these components.

2.1 | Latent Variable Model for Extremes

As stated within the introduction, although many approaches exist for the spatial modeling of extreme event magnitude, we restrict our attention to a single approach: the Bayesian hierarchical model of Cooley et al. (2007) that has been found very competitive in many situations. Although working with exceedances above a threshold would be possible, in this paper, we will restrict our attention to (pointwise) block maxima data, for example, annual maxima wind speed. With this setting, the univariate extreme value theory (Coles 2001; de Haan and Fereira 2006) states that it is sensible to assume that the univariate marginal distributions of $\{\eta(s) : s \in \mathcal{X}\}$ belong to the Generalized Extreme Value (GEV) family (Jenkinson 1955), that is, for all $s \in \mathcal{X}$, we have

$$\Pr\{\eta(s) \le z\} = \exp\left[-\left\{1 + \xi(s)\frac{z-\mu(s)}{\sigma(s)}\right\}^{-1/\xi(s)}\right], \quad 1 + \xi(s)\frac{z-\mu(s)}{\sigma(s)} > 0 \tag{1}$$

where $\mu(s) \in \mathbb{R}, \sigma(s) \in (0, \infty), \xi(s) \in \mathbb{R}$ are, respectively, the location, scale, and shape parameters of the GEV distribution.

As suggested with the above notations, it is not unusual that the process $\{\eta(s) : s \in \mathcal{X}\}$ is nonstationary and, that, in particular, one has to assume that the GEV parameters $\{\mu(s), \sigma(s), \xi(s) : s \in \mathcal{X}\}$ vary smoothly over the spatial domain $\mathcal{X}$ according to some (trivariate) stochastic process. To avoid unnecessarily complicated structures, it is common practice (Casson and Coles 1999; Cooley et al. 2007) to assume that each GEV parameter is mutually independent and follows univariate Gaussian processes. For example, with the French wind speed dataset of Section 5, we will assume that the latent GEV location parameter process $\{\mu(s) : s \in \mathcal{X}\}$ follows a Gaussian process with an isotropic exponential covariance family, that is, for all $s_1, s_2 \in \mathcal{X}$,

$$\gamma_\mu(h) = \mathrm{Cov}\{\mu(s_1), \mu(s_2)\} = \tau_\mu \exp\left(-\frac{h}{\lambda_\mu}\right), \qquad h = \|s_1 - s_2\|$$

where $\boldsymbol{\psi}_\mu = (\tau_\mu, \lambda_\mu)^\top \in (0, \infty)^2$ are respectively the sill and range parameters, and mean function

$$m_\mu(s) = \beta_{0,\mu} + \beta_{1,\mu}\mathrm{alt}(s)$$

where $\mathrm{alt}(s)$ denotes the elevation at location $s \in \mathcal{X}$ and $\boldsymbol{\beta}_\mu = (\beta_{0,\mu}, \beta_{1,\mu})^\top$ are unknown regression parameters. Similar formulations are used for the scale and shape GEV parameter processes.

Then we further impose a conditional independence assumption, that is, conditionally on these three Gaussian processes, the block maxima are assumed to be independent:

$$\begin{aligned} \eta(s)|\{\mu(s), \sigma(s), \xi(s)\} &\overset{\text{ind}}{\sim} \mathrm{GEV}\{\mu(s), \sigma(s), \xi(s)\}, \qquad s \in \mathcal{X} \\ \mu(\cdot) &\sim \mathrm{GP}(\boldsymbol{\beta}_\mu; \boldsymbol{\psi}_\mu) \\ \sigma(\cdot) &\sim \mathrm{GP}(\boldsymbol{\beta}_\sigma; \boldsymbol{\psi}_\sigma) \\ \xi(\cdot) &\sim \mathrm{GP}(\boldsymbol{\beta}_\xi; \boldsymbol{\psi}_\xi) \end{aligned} \tag{2}$$

where GP($\boldsymbol{\beta}; \boldsymbol{\psi}$) denotes a univariate Gaussian process whose mean function depends on the regression parameter vector $\boldsymbol{\beta}$ and whose covariance function depends on the spatial dependence parameter vector $\boldsymbol{\psi}$. Note that, throughout this paper, $\boldsymbol{\beta}$ will refer to the parameter vector of mean functions while $\boldsymbol{\psi}$ to the spatial dependence parameter vector of (cross)-covariance functions.

2.2 | The Projected Gaussian Process Model

We now focus on the angular component $\{\theta(s) : s \in \mathcal{X}\}$ and restrict our attention to a single approach: the projected Gaussian model. Focusing only on this model is motivated by the fact that it extends naturally to the spatial setting. However, before working in a spatial context, we first recall how one can get univariate angular distributions from projections. The (univariate) projected Gaussian distribution (Gelfand and Wang 2013) is the distribution of a random angle θ obtained via the projection of a bivariate Gaussian vector $\mathbf{X} = (X_1, X_2)^\top$ onto the unit circle $\mathbb{S}^1 = \{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\|_2 = 1\}$. More formally, we have $\theta = \arctan^*(X_2/X_1)$, where $\arctan^*$ is a quadrant-specific tangent inverse function (Jammalamadaka and SenGupta 2001).

Extending the above univariate circular distribution to the spatial setting amounts to switch the bivariate random vector $\mathbf{X}$ for a bivariate Gaussian process $\{\mathbf{X}(s) = (X_1(s), X_2(s))^\top : s \in \mathcal{X}\}$ and gives the projected Gaussian process model (Gelfand and Wang 2014). The model is then completely characterized by specifying the mean and cross–covariance functions, that is, for all $s, s_1, s_2 \in \mathcal{X}$,

$$m_\theta : s \mapsto m_\theta(s) = (m_\theta^{(1)}(s), m_\theta^{(2)}(s))^\top = \mathbb{E}[\mathbf{X}(s)],$$
$$\gamma_\theta : (s_1, s_2) \mapsto \gamma_\theta(s_1, s_2) = \mathbb{E}[\{\mathbf{X}(s_1) - m_\theta(s_1)\}\{\mathbf{X}(s_2) - m_\theta(s_2)\}^\top]$$

In practice, one must assume some relevant parametric structures. For example, for the French wind speed data set of Section 5, Model 0 assumes that the mean function satisfies

$$m_\theta^{(1)}(s) = \beta_{0,\theta}^{(1)} + \beta_{1,\theta}^{(1)}\text{lon}(s) + \beta_{2,\theta}^{(1)}\text{lat}(s) + \beta_{3,\theta}^{(1)}\text{alt}(s),$$
$$m_\theta^{(2)}(s) = \beta_{0,\theta}^{(2)} + \beta_{1,\theta}^{(2)}\text{lon}(s) + \beta_{2,\theta}^{(2)}\text{lat}(s) + \beta_{3,\theta}^{(2)}\text{alt}(s)$$

where lon(s), lat(s), and alt(s) denote longitude, latitude, and elevation, respectively, at location $s \in \mathcal{X}$ and $\boldsymbol{\beta}_\theta = \{\beta_{i,\theta}^{(j)} : i = 0, \ldots, 3, j = 1, 2\}$ are unknown regression parameters.

Similarly, one can assume a stationary, isotropic, and separable cross- covariance function, that is, for all $s_1, s_2 \in \mathcal{X}$, we have

$$\gamma_\theta(s_1, s_2) = T \otimes \rho(h), \qquad T = \begin{bmatrix} \tau_\theta & \rho_\theta\sqrt{\tau_\theta} \\ \rho_\theta\sqrt{\tau_\theta} & 1 \end{bmatrix}$$
$$\tau_\theta > 0, \quad \rho_\theta \in (-1, 1) \tag{3}$$

where $h = \|s_1 - s_2\|_2$, $\otimes$ denotes the Kronecker product, ρ is any parametric correlation function, for example, exponential, $\tau_\theta = \text{Var}\{X_1(s)\}$ and $\rho_\theta = \text{Cor}\{X_1(s), X_2(s)\}$ is the cross–correlation parameter for all $s \in \mathcal{X}$. As pointed out by Gelfand and Wang (2013), one must set $\text{Var}\{X_2(s)\} = 1$ to ensure identifiability of the parameters. Figure 1 plots one realization from the projected Gaussian process model and exhibits how it can handle a large variety of behaviors, for example, multimodality and asymmetry.

Working with this model is however challenging as no closed form exists for its likelihood. Fortunately, this hurdle can be bypassed by working within an augmented data framework and defining a radial random process $\{R(s) : s \in \mathcal{X}\}$ so that, for any $s \in \mathcal{X}$, one can easily switch from polar coordinates $\{R(s), \theta(s)\}$

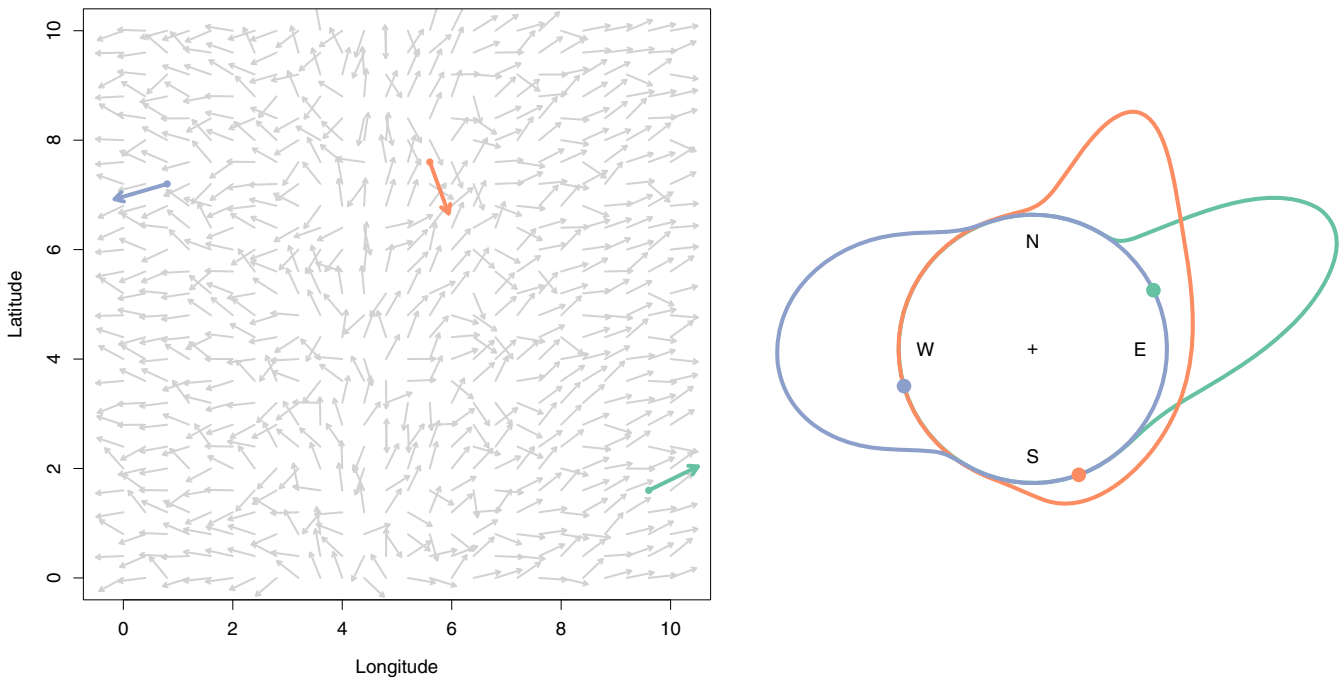

FIGURE 1 | Illustration of the projected Gaussian process. Left: Realization of a projected Gaussian process with three highlighted locations (colored arrows). Right: projected Gaussian densities at the three highlighted locations with corresponding observed direction (colored circles).

to Cartesian ones $\mathbf{X}(s) = \{R(s)\cos\theta(s), R(s)\sin\theta(s)\}^\top$. In practice, inference is therefore based on the induced bivariate Gaussian process $\{\mathbf{X}(s) : s \in \mathcal{X}\}$ and it is easily seen that, for any $\mathbf{s} = (s_1, \ldots, s_k)^\top \in \mathcal{X}^k$, $k \geq 1$, the joint density of the random vector $\{(R(s_j), \theta(s_j)) : j = 1, \ldots, k\}$ is nothing but a multivariate Gaussian density written in polar coordinates, that is, for any $\mathbf{r} \in (0,\infty)^k$, $\mathbf{t} \in [0, 2\pi)^k$ we have

$$f_{\mathbf{s}}(\mathbf{r}, \mathbf{t}) = (2\pi)^{-k/2}|\gamma_\theta(\mathbf{s})|^{-1/2} \exp\left(-\frac{\left\{\begin{pmatrix}\mathbf{r}\cos\mathbf{t}\\ \mathbf{r}\sin\mathbf{t}\end{pmatrix} - m_\theta(\mathbf{s})\right\}^\top \gamma_\theta(\mathbf{s})^{-1}\left\{\begin{pmatrix}\mathbf{r}\cos\mathbf{t}\\ \mathbf{r}\sin\mathbf{t}\end{pmatrix} - m_\theta(\mathbf{s})\right\}}{2}\right)\prod_{i=1}^k r_i \tag{4}$$

where $m_\theta(\mathbf{s}) = \mathbb{E}[\mathbf{X}(\mathbf{s})]$ and $\gamma_\theta(\mathbf{s})$ is a block matrix of dimension $2k \times 2k$ whose blocks (each of dimension $k \times k$) are $\mathrm{Cov}\{X_i(\mathbf{s}), X_j(\mathbf{s})\}$, $i, j = 1, 2$. Note that in (4), the multiplications $\mathbf{r}\cos\mathbf{t}$ and $\mathbf{r}\sin\mathbf{t}$ are done component-wise.

3 | Extreme-Angular Bayesian Hierarchical Model

In this section, we detail how one can merge models of Sections 2.1 and 2.2 to define what we shall call an extreme-angular Bayesian hierarchical model. Clearly, to allow for some dependence between the extremal process $\{\eta(s) : s \in \mathcal{X}\}$ and the angular one $\{\theta(s) : s \in \mathcal{X}\}$, the two aforementioned models should interact together, and, in this paper, we propose that the mean function $m_\theta(\cdot)$ of the projected Gaussian process may also depend on the GEV parameters latent processes $\{\mu(s) : s \in \mathcal{X}\}$, $\{\sigma(s) : s \in \mathcal{X}\}$ and $\{\xi(s) : s \in \mathcal{X}\}$. The motivation for this choice is that it may happen that the most severe extreme events are associated with a very specific angle. For example, for the French wind speed dataset of Section 5, given that we have observed a very large extreme wind speed somewhere along the Atlantic coast, it is likely that the associated wind direction is West or South West. The converse, however, is not necessarily true as the knowledge of a wind direction may not give any indication that the wind speed is extremely severe or not.

Our proposed model is

$$\begin{aligned}
\eta(s)|\{\mu(s),\sigma(s),\xi(s)\} &\overset{\text{ind}}{\sim} \mathrm{GEV}\{\mu(s),\sigma(s),\xi(s)\},\ s \in \mathcal{X}\\
\begin{pmatrix}R(\cdot)\cos\theta(\cdot)\\ R(\cdot)\sin\theta(\cdot)\end{pmatrix}|\{\mu(\cdot),\sigma(\cdot),\xi(\cdot)\} &\sim \mathrm{GP}_{2,\mathrm{polar}}(\boldsymbol{\beta}_\theta;\boldsymbol{\psi}_\theta)\\
\mu(\cdot) &\sim \mathrm{GP}(\boldsymbol{\beta}_\mu;\boldsymbol{\psi}_\mu)\\
\sigma(\cdot) &\sim \mathrm{GP}(\boldsymbol{\beta}_\sigma;\boldsymbol{\psi}_\sigma)\\
\xi(\cdot) &\sim \mathrm{GP}(\boldsymbol{\beta}_\xi;\boldsymbol{\psi}_\xi)
\end{aligned} \tag{5}$$

where $\mathrm{GP}_{2,\mathrm{polar}}(\boldsymbol{\beta}_\theta;\boldsymbol{\psi}_\theta)$ denotes a bivariate Gaussian process whose likelihood is expressed in polar coordinates as in (4), mean function $m_\theta(\cdot)$ depends on the regression parameter vector $\boldsymbol{\beta}_\theta$ and $\boldsymbol{\psi}_\theta = (\tau_\theta, \rho_\theta, \lambda_\theta)$ are the spatial dependence parameters appearing in the cross-covariance function (3)—λ_θ being the range parameter of the exponential correlation function $\rho(h) = \exp(-h/\lambda_\theta)$.

It is important to emphasize that the mean function m_θ may depends not only on spatial covariates such as longitude or latitude, the GEV parameters $\mu(\cdot)$, $\sigma(\cdot)$, and $\xi(\cdot)$ but also on any suitable transformation of these GEV parameters. For instance, for the French wind speed dataset of Section 5, we will assume that the mean function is

$$\begin{aligned}
m_\theta^{(1)}(s) &= \beta_{0,\theta}^{(1)} + \beta_{1,\theta}^{(1)} q_{0.95}(s) + \beta_{2,\theta}^{(1)} q_{0.99}(s)\\
m_\theta^{(2)}(s) &= \beta_{0,\theta}^{(2)} + \beta_{1,\theta}^{(2)}\mathrm{lon}(s) + \beta_{2,\theta}^{(2)}\mathrm{lat}(s) + \beta_{3,\theta}^{(2)}\mathrm{alt}(s)
\end{aligned}$$

where $q_p(s)$ is the quantile of order $p \in (0,1)$ of the GEV distribution at location s, that is, with location, scale, and shape parameters equal to $\mu(s)$, $\sigma(s)$, and $\xi(s)$ respectively.

Inference from Model (5) is challenging as, due to the presence of the latent processes $\mu(\cdot)$, $\sigma(\cdot)$, $\xi(\cdot)$ and $R(\cdot)$, the likelihood has an intractable integral representation. Although other approaches are possible, for example, E.-M. type approaches (Dempster et al. 2018), in this paper we suggest to work within a Bayesian framework and perform numerical integration using MCMC techniques (Hastings 1970). Figure 2 gives the directed acyclic graph of Model (5) and shows where prior distributions are placed and how the latent radial process $\{R(s) : s \in \mathcal{X}\}$ comes into view. In practice, a Gibbs sampler (Casella and George 1992) is used whose explicit expressions and implementation details can be found in Appendix A. Note that numerical integration is performed by sampling from the full conditional distributions of the latent variables, that is, Steps 1 and 2 of Appendix A.

Spatial models often aim at making predictions at unobserved locations $s_* \in \mathcal{X}$ such as the values of the GEV parameters, return levels or the circular mode of the posterior distribution of $\theta(s^*)$. Within a Bayesian framework, it is typically done through the posterior predictive distribution

$$\int \pi(\eta(s_*), \theta(s_*)|\boldsymbol{\Upsilon}, \mathcal{D}_n)\pi(\boldsymbol{\Upsilon}|\mathcal{D}_n)\mathrm{d}\boldsymbol{\Upsilon}$$

where $\mathcal{D}_n = \{(\eta_i(s_j), \theta_i(s_j)) : i = 1, \ldots, n, j = 1, \ldots, k\}$ is the dataset of values of the processes over k sites and n replicates and $\boldsymbol{\Upsilon}$ is the vector containing all the parameters of the model. As expected, the above integral representation has no closed form and, similarly to the Gibbs sampler introduced above, one has to resort to numerical integration where each state of the Markov chain generated during the fitting stage is browsed to generate a sample from this distribution, see Algorithm 1. Algorithm 1 can be used in two different ways to get what one may call unconditional or conditional predictions. Unconditional predictions consist in sampling from the predictive posterior distribution exactly as described in the algorithm. Conditional sampling is somewhat similar to kriging and imposes that the angular process satisfies $\theta(\tilde{s}_j) = \theta_j$ for some conditioning spatial locations $\tilde{s}_j \in \mathcal{X}$ and conditioning values $\tilde{\theta}_j \in [0, 2\pi)$, $j = 1, \ldots, \tilde{k}$, $\tilde{k} \geq 1$. Often, but not invariably, the conditioning locations $\tilde{s}_j$ will be identical to those where we have some data, that is, $\tilde{s}_j = s_j$, $j = 1, \ldots, k$. In this situation, line 4 of Algorithm 1 is substituted by the sampling of latent radius $R_I(\cdot)$ at each conditioning locations $\tilde{s}_j$ and then proceed with the rest of the algorithm as usual.

ALGORITHM 1 | Pointwise predictive posterior predictions from the extreme-angular model.

input : A Markov chain $\{\boldsymbol{\Upsilon}_t : t = 1, \dots, T\}$ sampledfrom the Gibbs sampler introduced above and a new location $s_* \in \mathcal{X}$.

output: Predictive posterior predictions of an unknown quantity $U(s_*)$.

for $t = 1, \dots, T$ **do**

`/* Conditional sampling of the GEV parameters */`

Sample from $\mu_t(s_*) \mid \mu(\mathbf{s}), \boldsymbol{\Upsilon}_t$, that is, aconditional Normal distribution;

Sample $\sigma_t(s_*)$ and $\xi_t(s_*)$ in the same way;

`/* Compute the Cartesian coordinates */`

Pickup a random replicate $I \sim \text{Unif}\{1, \dots, n\}$;

Compute the Cartesian coordinates

$$\mathbf{X}(s_j) = \{R_I(s_j)\cos\theta_I(s_j), R_I(s_j)\sin\theta_I(s_j)\}^\top, \qquad j = 1, \dots, k$$

`/* Conditional sampling of the bivariate Gaussian process */`

Sample from $\mathbf{X}(s_*) \mid \mathbf{X}(\mathbf{s})$, that is,a conditional bivariate Normal distribution;

`/* Retrieve the angular component */`

Compute $\theta_t(s_*) = \arctan^*\{X_1(s_*)/X_2(s_*)\}$;

`/* Generate a realization of the quantity of interest */`

Compute (or sample) $U_t(s_*)$ based on the simulated values $\mu_t(s_*), \sigma_t(s_*), \xi_t(s_*)$ and $\theta_t(s_*)$;

end

Return the sample mean, median, and mode based on $\{U_t(s_*) : t = 1, \dots, T\}$;

4 | Simulation Study

In order to assess the performance of our MCMC sampler, we run a simulation study with a varying sample size n, number of locations k, and three different dependence configurations as shown in Table 1. In Configuration I, the angular mean function m_θ is independent of the GEV parameters and, consequently, angles and extremes are independent. With this setting, the Gibbs sampler introduced in Section 3 is equivalent to the use of two independent samplers: one for the GEV component and one for the angular component. Configuration II provides a linear dependence between angles and extremes through the shape parameter of the GEV distribution. However, since the Gaussian process for the shape parameter has a constant mean and low variance, the shape parameter is roughly constant over $\mathcal{X}$ and, as so, dependence between angles and extremes magnitudes is limited. On the contrary, in Configuration III, the angular mean function depends on the GEV location parameter, which varies significantly over $\mathcal{X}$ yielding to a strong dependence between angles and extremes magnitudes.

Working within a fixed domain framework, that is, the spatial domain $\mathcal{X}$ is fixed, we numerically analyze infill and sample size asymptotics, that is., when the number of locations $k \to \infty$, the sample size n being fixed, and conversely. For each dependence setting of Table 1, we consider the cases where $k = 10, 25, 50$, and $n = 20, 50, 100$, leading to 27 configurations overall.

Figure 3 shows the evolution of the mean squared error for some Bayesian pointwise estimators as k and n increase. The top row of Figure 3 focuses on the Gaussian process of the GEV location parameter. More precisely, results are shown for $\beta_{2,\mu}$, a regression parameter of the mean function m_μ, the range parameter λ_μ of the covariance function $\gamma_\mu(\cdot)$ as well as the ratio $\tau_\mu/(1+\lambda_\mu)$, where τ_μ is the sill parameter of $\gamma_\mu(\cdot)$. As expected, the mean squared error (MSE) for $\beta_{2,\mu}$ decreases as both k and n increase. Indeed, as k increases, the number of GEV parameters increases and the linear structure $\mathbb{E}[\mu(s)] = x(s)^\top \boldsymbol{\beta}_\mu$ becomes more apparent. Similarly, as the number of replicates n increases, the GEV parameter estimates become increasingly more accurate, and the above linear structure has less noise. Results (not shown) for other mean function regression parameters show similar patterns across all dependence configurations.

Interestingly, the MSE for λ_μ has a completely different behavior and no convergence is visible. A similar pattern (not shown) can be seen for the sill parameter τ_μ. As shown in Zhang (2004), there is no consistency in the estimation of both the sill and range parameters of a Gaussian process with a single realization. This result applies to the GEV part of our model as each location has a single set of GEV parameters and, consequently, a single realization of the Gaussian processes.

The bottom row of Figure 3 is similar to the top row with an emphasis on the angular component, namely $\beta^{(2)}_{2,\theta}$, ρ_θ and λ_θ. As expected and using the same arguments as the ones stated previously, the evolution of the MSE for $\beta^{(2)}_{2,\theta}$ is similar to that for $\beta_{2,\mu}$. There is, however, a subtle difference in the estimation of the parameters of the bivariate Gaussian process $\{\boldsymbol{X}(s) : s \in \mathcal{X}\}$ compared to that for the GEV parameters, for example, $\{\mu(s) : s \in \mathcal{X}\}$. Contrary to the latter case, parameters are now estimated from $n > 1$ independent replicates, and no consistency issues exist for the joint estimation of the sill and range parameters.

At first sight, the above lack of consistency may cause problems in the estimation of the GEV parameters or related quantities such as quantiles. Fortunately, Zhang (2004) have shown that the ratio of the sill and range parameters can be consistently estimated (see top right panel of Figure 3) and, more importantly, that there is no impact on the predictions of the GEV parameters. Figure 4 compares the true GEV parameters from that estimated from our MCMC sampler. As expected, one can see that the predicted GEV parameters match the theoretical ones. The shape parameter being usually harder to estimate, slightly worse performances can be seen. Overall, it corroborates that the consistency issue has no impact on GEV parameter estimation.

Figure 5 compares the estimated circular densities to the theoretical ones. Again, one can see that there is a perfect match, indicating that, although there is a lack of consistency in the estimation of the sill and range parameters of GEV Gaussian processes, it has no impact on the prediction performances for the angular component.

We now assess the predictive performance of the proposed model and, more specifically, that associated to the prediction of some angular quantity over the spatial domain $\mathcal{X}$. As the projected

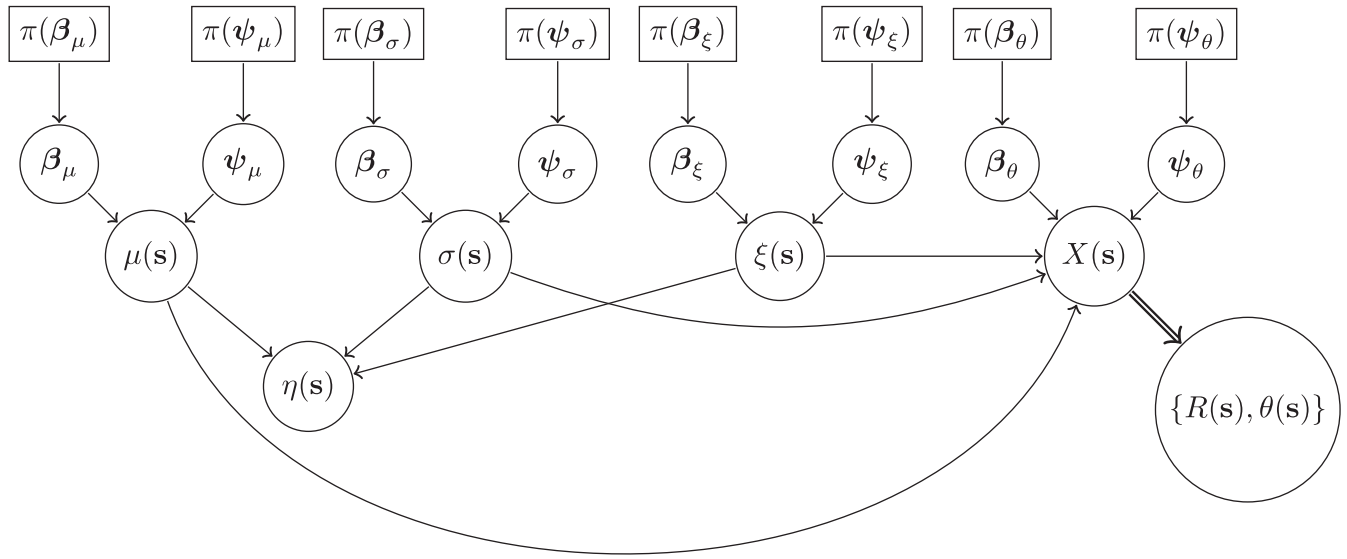

FIGURE 2 | Directed acyclic graph for the extreme-angular Bayesian hierarchical model. A double arrow indicates a deterministic relationship between the two nodes. Squared nodes denote prior distributions.

TABLE 1 | Configuration settings for the simulation study. The angular mean function is set to $m_\theta(s) = \{m_\theta^{(1)}(s), m_\theta^{(2)}(s)\}^\top$. For each configuration, the latent GEV parameter processes $\mu(\cdot)$, $\sigma(\cdot)$, $\xi(\cdot)$ were held fixed to Gaussian processes with mean functions $m_\mu(s) = 2 - 3\text{lon}(s) - 2\text{lat}(s)$, $m_\sigma(s) = 2 + \text{lat}(s)$, $m_\xi(s) = 0.05$, sill parameters 0.1,0.5,0.05 and range parameters 1,2,3 respectively.

	σ^2	ρ	$m_{\theta,1}(s)$	$m_{\theta,2}(s)$	**Modality**
I: Independent	0.4	0.3	0.5	0	Multimodal
II: Light	0.4	0.3	0.5	$2\xi(s)$	Multimodal
III: Strong	1.0	0	$10 + 0.5\mu(s)$	0	Unimodal

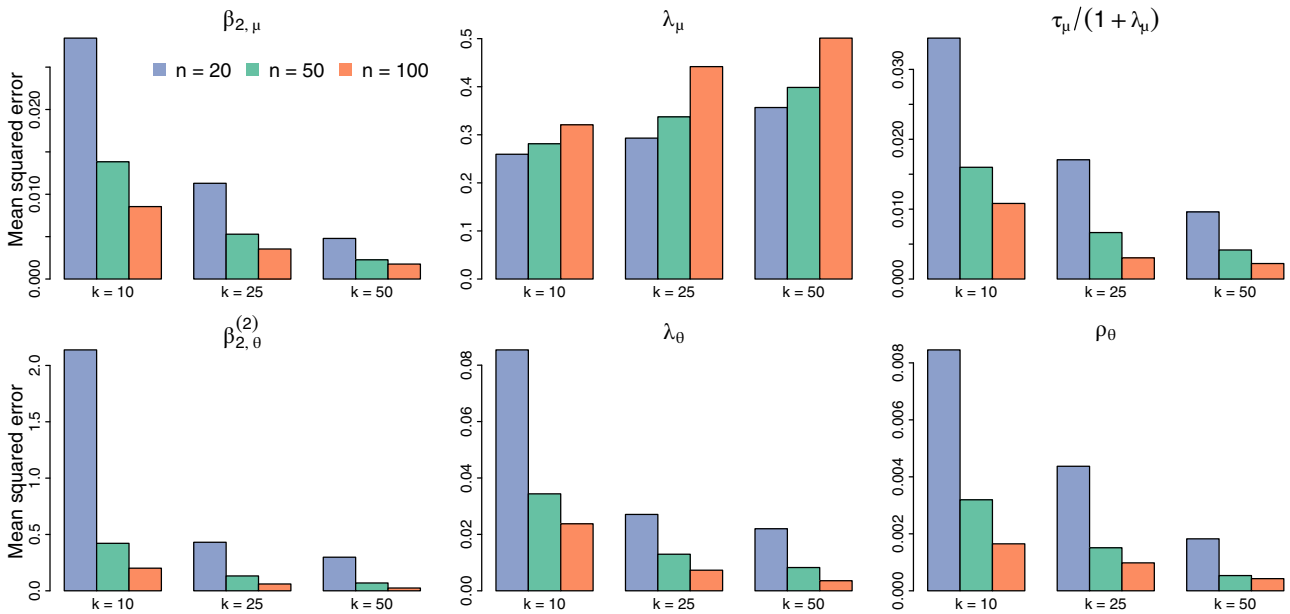

FIGURE 3 | Evolution of the mean squared error (MSE) with varying number of observations n and locations k for Configuration II. Top: MSE for the regression (left), range (middle) parameters and quotient of scale and range (right) of the covariance function of the Gaussian process $\mu(\cdot)$. Bottom: MSE for regression parameter (left), range parameter (middle) and correlation parameter (right) of the projected Gaussian process $\theta(\cdot)$.

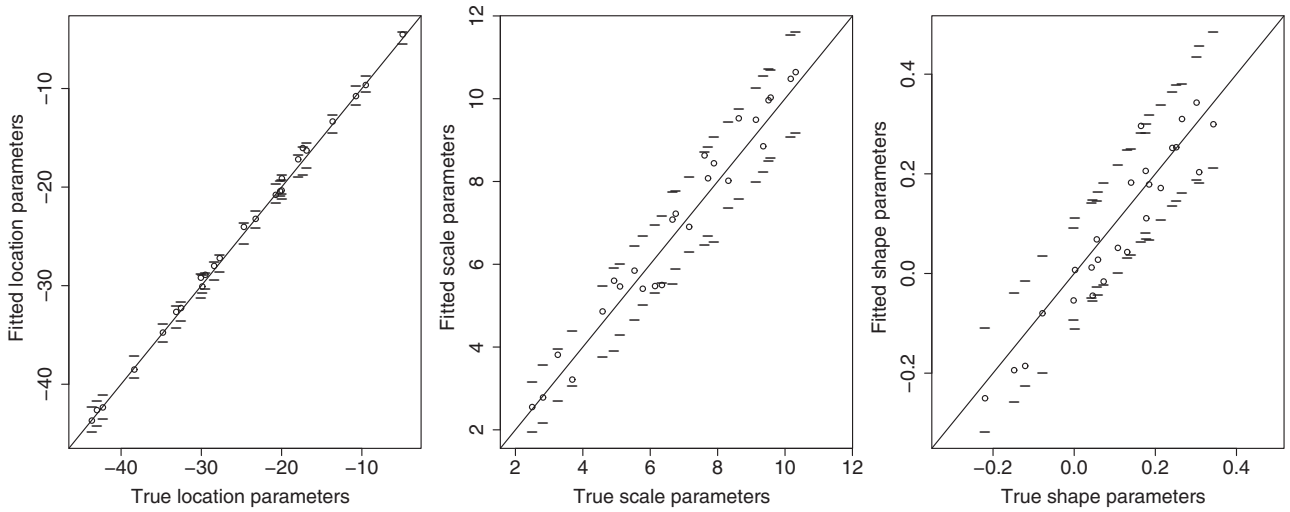

FIGURE 4 | Model checking: extremal component. Comparison between the true GEV parameters and the posterior medians derived from the MCMC sampler with $k = 25$ locations and $n = 50$ observations for Configuration II. 95% credible intervals are also reported.

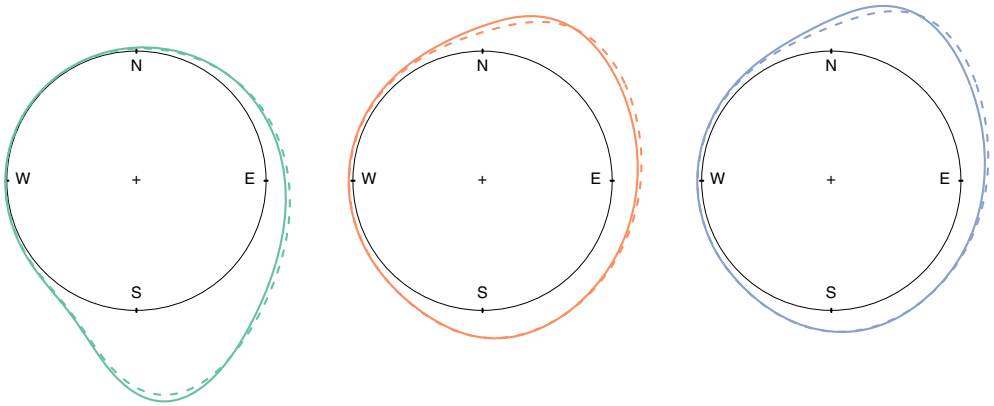

FIGURE 5 | Model checking: angular component. Estimated (solid line) and theoretical (dashed line) circular distribution for locations with a low (left), medium (middle) and large (right) value of the GEV shape parameter ξ. Results are obtained from Configuration II, with $k = 25$ locations and $n = 50$ observations.

Gaussian process may be multimodal, some care is needed in defining the Bayesian point estimate, and, in the sequel, we will focus on the (main) posterior mode for the angle. Figure 6 shows prediction maps for Configuration II. With this setting, the predictive posterior distribution is bimodal and may introduce spatial discontinuities in the posterior modes. To be more specific, there is a cutoff value for the shape parameter for which if exceeded, the modal direction is top-right and bottom-right otherwise. The left panel of Figure 6 illustrates the relationship between predicted angles and the shape parameter as well as the aforementioned cutoff behavior. The middle panel is similar to the previous one except that predicted quantiles of order 0.95 are now overlaid. Since GEV quantiles are a function of the GEV location, scale, and shape parameters, overall, the same behavior can be seen. However, due to the analytic expression for GEV quantiles, there is no cut-off behavior as the one stated previously. For instance, estimates in the outermost bottom-right sub-region are lower due to the contribution of the location parameter as opposed to the general behavior "bottom−right direction for small return levels."

The right panel of Figure 6 shows the relationship between predicted angles and the angular dispersion $s = \sqrt{1 - \|\overline{x}_n\|_2}$, with $\overline{x}_n$ the empirical (Cartesian) mean of the angular data, that is, $\{x_i = (\cos\theta_i, \sin\theta_i) : i = 1, \ldots, n\}$. The angular dispersion satisfies $s \in [0, 1]$, where 0 corresponds to a degenerate distribution, that is, the angular distribution puts all its mass on a specific angle, and 1 corresponds to the uniform distribution on $[0, 2\pi)$. One can see that the two modes have different impact. More precisely, one can see that the mode associated to the top-right direction is the preponderant one with typically large dispersion, whereas that associated to the bottom-right direction only dominates after a rare cut-off exceedance and has generally less dispersion.

5 | Application

The data, freely available from Météo-France, consist of annual maxima of wind speeds and their associated wind directions observed at a height of 10m for $k = 110$ weather stations in

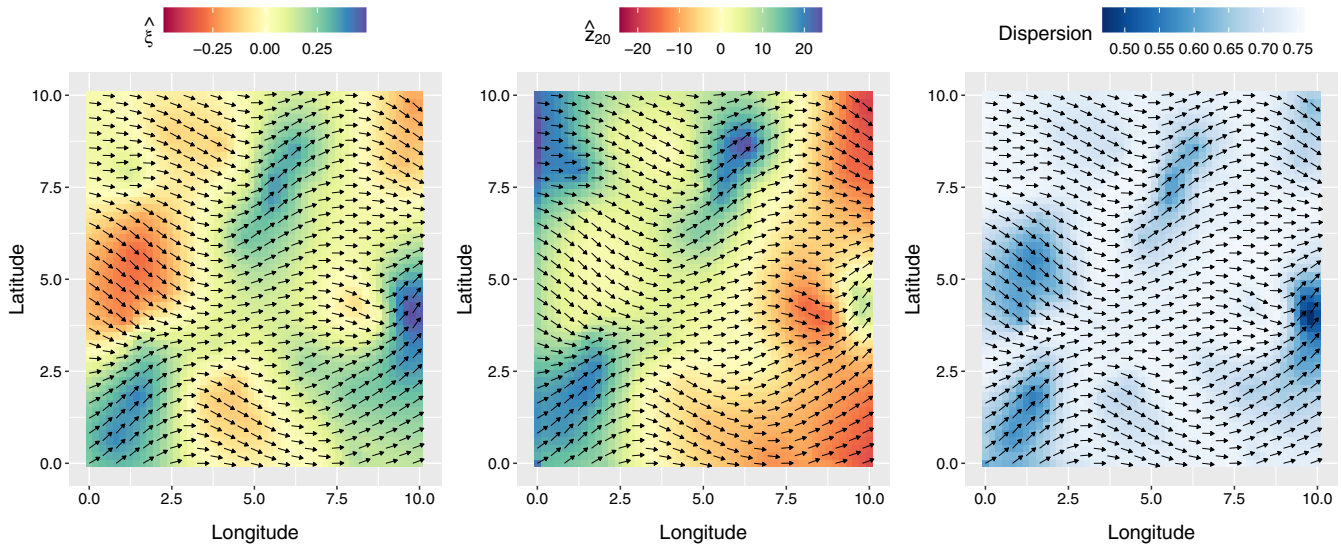

FIGURE 6 | Prediction maps for Configuration II with $n = 100$ observations and $k = 50$ locations. All predictions are derived from the predictive posterior distribution (median for extreme related quantities and angular mode for the main direction). Left: Prediction of the main direction and GEV shape parameter. Middle: Prediction of the main direction and GEV quantiles of order 0.95. Right: prediction of the main direction and circular standard deviation.

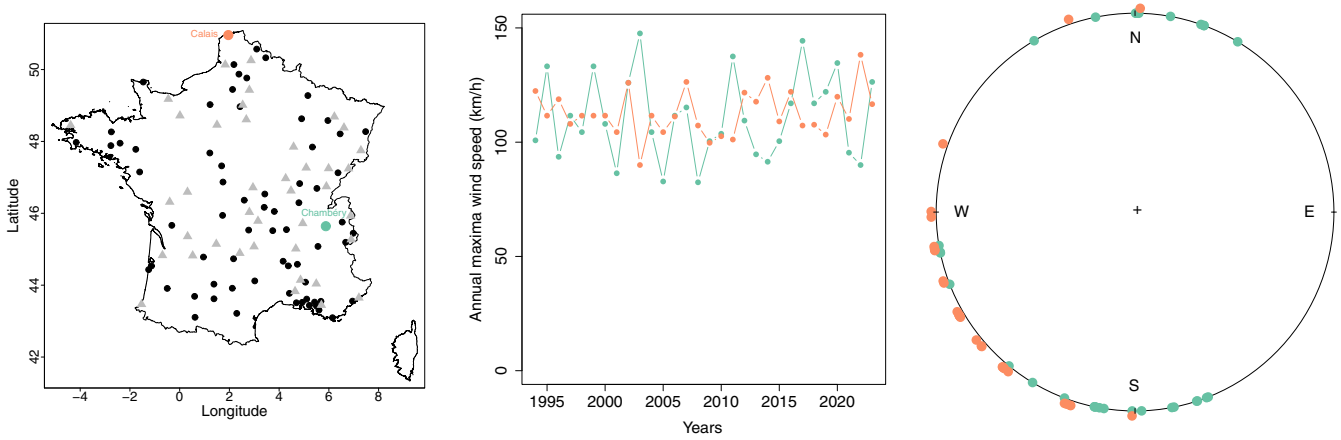

FIGURE 7 | Wind data in continental France. From left to right: study region and locations of the weather stations (train set: black circle; test set: gray triangle); annual maxima wind speed time series and empirical distributions of the direction of extreme wind speed at two selected stations: Calais (orange) and Chambéry (green).

France and recorded from 1994 to 2023. To assess the performance of the model, the dataset is split into a train and a test set. Figure 7 plots the spatial distribution of the weather stations, the annual maxima wind speed time series, and the empirical distributions of wind directions associated to the annual maxima for two selected stations. One can see that annual wind speed maxima at Calais have a lower amplitude than in Chambéry and appear to have a single main direction (South-West), while that at Chambéry may originate from two different directions (North and South-South-West). As expected, the spatial distribution of extreme wind speeds is not stationary over France; neither is that for wind directions. Modeling such nonstationary behavior is challenging, but it is hoped that thanks to its flexibility, the use of Model (5) will be able to cope with those two different types of nonstationarity.

To assess prediction performance, a sensible performance metric is the Widely Applicable Information Criterion (WAIC) (Watanabe 2013) that, similarly to Akaike Information Criterion (Akaike 1974), penalizes model complexity, but is more relevant when performing model selection in a Bayesian hierarchical setting. Table 2 displays performance scores, computed on the train set, for several competitive models with a varying degree of complexity. Note that in addition to geophysical covariates such as longitude, latitude, and elevation, some models may also consider extreme wind speed quantiles for the modeling of wind speed direction. Again, the rationale for this type of dependence is that it may be sensible to assert that the largest extreme wind speeds are attached with a very specific wind direction.

TABLE 2 | Widely Applicable Information Criterion (WAIC) for selected models (computed on the train set only). All models have $m_{\xi}(s) = \beta_0$. The quantity $q_p(s)$ denotes the quantile of order p of the GEV distribution at location s.

		WAIC_θ	WAIC_η	WAIC
Model 0	$m_\mu(s) = \beta_0 + \beta_1 \text{alt}(s)$ $m_\sigma(s) = \beta_0 + \beta_1 \text{lon}(s) + \beta_2 \text{lat}(s) + \beta_3 \text{alt}(s)$ $m_\theta^{(1)}(s) = \beta_0 + \beta_1 \text{lon}(s) + \beta_2 \text{lat}(s) + \beta_3 \text{alt}(s)$ $m_\theta^{(2)}(s) = \beta_0 + \beta_1 \text{lon}(s) + \beta_2 \text{lat}(s) + \beta_3 \text{alt}(s)$	13,693	**16,343**	30,036
Model 1	Model 0 with $m_\theta^{(1)}(s) = \beta_0 + \beta_1 \text{lon}(s) + \beta_2 \text{lat}(s) + \beta_3 \text{alt}(s) + \beta_4 \mu(s) + \beta_5 \sigma(s)$ $m_\theta^{(2)}(s) = \beta_0 + \beta_1 \text{lon}(s) + \beta_2 \text{lat}(s) + \beta_3 \text{alt}(s) + \beta_4 \mu(s) + \beta_5 q_{0.95}(s)$	13,791	16,581	30,372
Model 2	Model 0 with $m_\theta^{(1)}(s) = \beta_0 + \beta_1 \text{lon}(s) + \beta_2 \text{lat}(s) + \beta_3 \text{alt}(s) + \beta_4 q_{0.95}(s) + \beta_5 q_{0.99}(s)$	12,925	16,412	29,337
Model 3	Model 2 with $m_\theta^{(1)}(s) = \beta_0 + \beta_1 q_{0.95}(s) + \beta_2 q_{0.99}(s)$	12,923	16,352	**29,275**
Model 4	Model 2 with $m_\mu(s) = \beta_0$ $m_\sigma(s) = \beta_0 + \beta_1 \text{lat}(s) + \beta_2 \text{alt}(s)$	12,913	16,364	29,277
Model 5	Model 3 with $m_\mu(s) = \beta_0$ $m_\sigma(s) = \beta_0 + \beta_1 \text{lat}(s) + \beta_2 \text{alt}(s)$	12,948	16,418	29,366
Model 6	Model 3 with $m_\theta^{(1)}(s) = \beta_0 + \beta_1 q_{0.5}(s) + \beta_2 q_{0.95}(s) + \beta_3 q_{0.99}(s)$	**12,898**	16,378	29,277
Model 7	Model 3 with $m_\theta^{(2)}(s) = \beta_0 + \beta_1 \text{lon}(s) + \beta_2 \text{lat}(s) + \beta_3 \text{alt}(s) + \beta_4 q_{0.5}(s)$	13,108	16,387	29,495

Although a large number of models have been considered, and to save space, results are only presented for a limited number of configurations. Model 0 assumes independence between the extreme value process $\{\eta(s) : s \in \mathcal{X}\}$ and the angular process $\{\theta(s) : s \in \mathcal{X}\}$ while all the other models assume dependence. One can see that the independent model performs best in predicting extreme wind speeds but is surprisingly poor in predicting wind directions. Hence, and as already mentioned previously, it seems that the knowledge of extreme wind speeds may bring relevant information for the prediction of extreme wind direction. Overall, Model 3 appears to be the most accurate as it is almost as efficient as the independent model in predicting wind speeds while ensuring good predictive performances for wind directions. This model makes use of the 20- and 100-year return levels.

A specificity of any Bayesian analysis is that results may be strongly influenced by the definition of the prior distributions. This is even more true in our context since, as suggested by Banerjee et al. (2014) and as a result of the lack of consistency previously mentioned, informative priors should be used for the sill and range parameters of the covariance functions in order to yield non-degenerate marginal posterior distributions. Figure 8 shows how estimates changes with the prior definition — ranging from uninformative up to very informative priors. For our application, the results are conclusive: estimates are almost not impacted by the prior choice. Indeed, the left panel shows that the marginal posterior for the cross-correlation parameter ρ_θ is almost identical whatever is the prior distribution. Similar results were found for the other parameters. The right panel of Figure 8 further confirms this statement as the posterior predictive distribution for the 20-year return level remains almost unchanged regardless of the prior distribution. We can therefore be confident in saying that, for our application, the posterior distribution is mainly driven by the data and not by the prior choices we made.

Figure 9 assesses the goodness of fit for Model 3. The two leftmost panels compare predicted values and observed wind speeds for the two highlighted stations of Figure 7. Although some bias is apparent, predictions are rather good. Better performances are likely to be obtained by using additional geophysical covariates, but unfortunately, such additional covariates were not provided in the data set, which was limited to longitude, latitude, and altitude only. The right panel compares the kernel density estimates of extreme wind speed directions and the fitted angular densities for these two locations. Predictions appear to be accurate, and one can see that our model is able to cope with both unimodal (Calais)

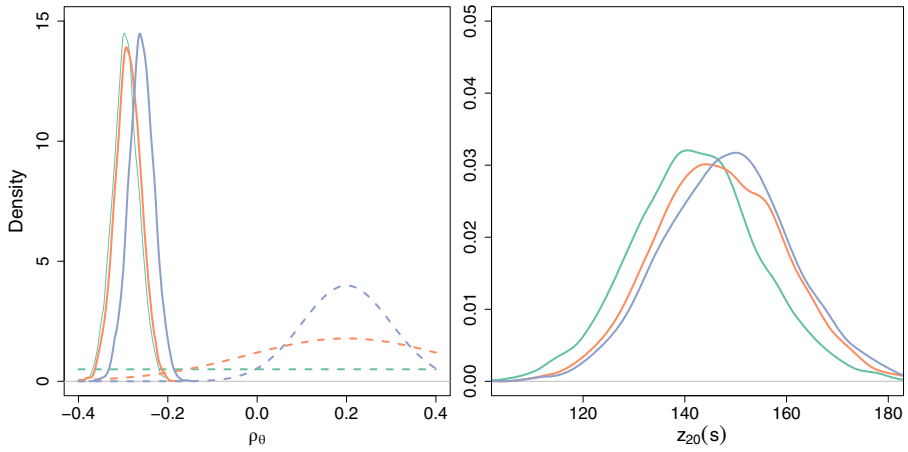

FIGURE 8 | Impact of the prior distribution (dashed lines) on the posterior and predictive posterior distributions (solid lines). The prior distributions are taken to be uninformative (green), moderately informative (orange) or strongly informative (blue). Left: Marginal prior and posterior distributions for the covariance parameter ρ_θ. Right: predictive posterior distribution of the 20-year return level.

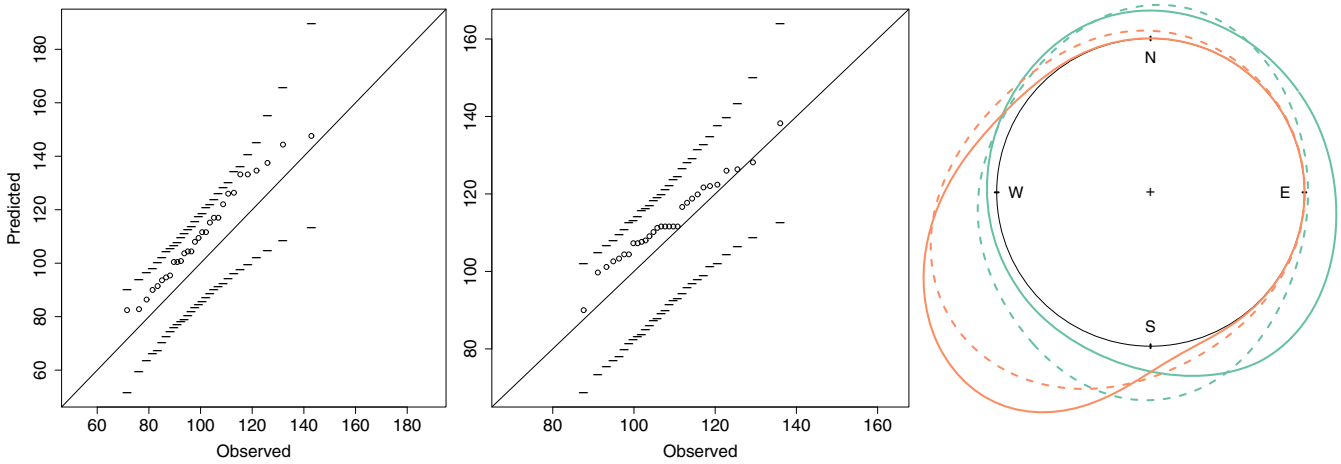

FIGURE 9 | Model checking for the French wind speed dataset based on unconditional predictions. The left and middle panels compare the observed and predicted maxima at Chambéry and Calais, respectively (with 95% credible envelopes). The right panel compares the predicted angular distributions (solid lines) and the angular kernel density estimates (dashed lines) at Chambéry (green) and Calais (orange) stations.

and bi-modal (Chambéry) distributions, even if the South mode of the latter is slightly shifted toward the East. Similar results were obtained for all the other stations.

Using Algorithm 1, we evaluate the conditional and unconditional predictive performances of Model 3 on both the train and test sets. Performance is assessed using the root mean squared error for wind speeds and the mean angular distance for wind directions. The angular distance is defined as $\alpha(\theta_1, \theta_2) = 1 - \cos(\theta_1 - \theta_2)$ (Grimit et al. 2006). Table 3 summarizes the empirical distribution of these two performance metrics. For the test set, predictions were made conditionally on the observed data at the training stations for each year, leading to one prediction per year for the angular component. This contrasts with unconditional predictions that are time-independent, that is, for each station, we use a single point estimate, for example, posterior median/mode, whatever the year is. Although not directly comparable, this difference explains why there is a performance gap between unconditional and conditional predictions for the angular component and no significant changes for the extremal part. Unsurprisingly, unconditional predictions shows worse results than conditional ones. We can see that 75% of the mean angular distances values are below 0.79 and 0.36 for the train and test sets respectively. This result can be considered as rather good since a uniform prediction on $[0, 2\pi)$ gives a mean absolute distance of 1. Similarly, only 25% of the RMSE values exceed 12.4 km/h and 10 km/h for the train and test sets respectively. However, the 97.5% quantiles and the maximum values for both metrics indicate that there are some stations where predictions are poor. Maybe not surprisingly, the worse performances were obtained in the South–East part of France where the spatial distribution is expected to change abruptly.

TABLE 3 | Summary statistics for the distribution of Root Mean Squared Errors (RMSE) and Mean Angular Distance (MAD) between pointwise conditional prediction and observed values computed on the train and test sets. Q_p denotes the empirical quantile of order $p \in [0, 1]$.

	Minimum	$Q_{0.025}$	$Q_{0.25}$	$Q_{0.5}$	Mean	$Q_{0.75}$	$Q_{0.975}$	Maximum
Train set (unconditional predictions)								
MAD	0.10	0.11	0.18	0.26	0.46	0.79	1.35	1.45
RMSE	2.1	2.1	4.9	7.2	9.4	12.4	26.0	28.6
Test set (conditional predictions)								
MAD	0.05	0.08	0.14	0.20	0.33	0.36	1.16	1.25
RMSE	2.2	2.4	4.6	6.4	8.9	10.5	25.7	42.6

TABLE 4 | Posterior medians and 95% credible intervals (in parenthesis) for Model 3.

Generalized extreme value layer

	β_0	β_{lon}	β_{lat}	β_{alt}	τ	λ	
μ	$95_{(93,97)}$	—	—	$2.4_{(-2.4,7.3)}$	$78_{(60,102)}$	$8_{(1,20)}$	
σ	$2_{(-13,18)}$	$-0.3_{(-0.7,0.1)}$	$0.19_{(-0.15,0.53)}$	$2.9_{(1.2,4.9)}$	$6.5_{(4,11)}$	$104_{(57,184)}$	
ξ	$-0.07_{(-0.1,0)}$	—	—	—	$0.01_{(0.00,0.01)}$	$127_{(75,225)}$	
Angular layer							
$m_\theta^{(1)}$	β_0	$\beta_{\text{q}(0.95)}$	$\beta_{\text{q}(0.99)}$		τ	λ	ρ_θ
	$-1.7_{(-3,-0.4)}$	$-0.20_{(-0.26,-0.15)}$	$0.18_{(0.14,0.23)}$		$0.6_{(0.5,0.7)}$	$45_{(41,49)}$	$-0.3_{(-0.4,-0.2)}$
$m_\theta^{(2)}$	β_0	β_{lon}	β_{lat}	β_{alt}			
	$8.4_{(7.1,9.7)}$	$0.03_{(0.01,0.05)}$	$-0.19_{(-0.22,-0.16)}$	$-0.2_{(-0.3,0.0)}$			

Table 4 presents the parameter estimates for Model 3 using uninformative priors across all parameters. Interestingly, the sum between the parameter estimates associated to the 20- and 100-year return levels varies around 0, that is, $\beta_{q(0.95)} + \beta_{q(0.99)} \approx 0$. As $q(p_1) - q(p_2) = \sigma c(p_1, p_2, \xi)$, where $q(p)$ is the quantile function of a GEV(μ, σ, ξ) and $c(p_1, p_2, \xi)$ only depends on p_1, p_2 and ξ, extreme wind speed direction appear to depend mainly on the spread, skewness, and higher order moments of extreme wind speeds rather than their central tendency. This behavior was shared with all the models we considered and whose mean function of the projected Gaussian process makes use of at least two return levels.

The top row of Figure 10 gives predictions maps from Model 3. One can see that the shape parameter appears to be rather constant over France apart from the Atlantic coastline and the South–East part of France. The 100-year return levels are larger in the most Western part, that is, Western part of Brittany and the Landes, and in mountainous regions, that is, the South-ast part of France and especially the French Alps. Overall, the wind direction dispersion is rather limited. One exception is the South-East part of France where the angular dispersion is much larger indicating that wind speed directions may vary roughly over space. Further, as expected, one can see that extreme wind event appears to come from West but tends to change smoothly to the South-West direction for the Northern part of France. Again the South-East part of France and the Alps show a completely different behavior. For the former, extreme wind speed events appear to come from West-North and is in agreement with the observational study of Obermann et al. (2018). The behavior in the French Alps is more erratic and may be a consequence that the wind trajectory has to follow the valley in mountainous regions.

The bottom row of Figure 10 is similar to the top row except that results are obtained from Model 0, which assumes independence between extreme magnitudes and angles. The aim of such plots is to emphasize the impact of assuming dependence between wind direction and wind speeds. Clearly, the major differences happen in the South-East part of France where Model 3 tends to predict larger wind speeds than Model 0. As far as wind direction is of concern, Model 0 produces unrealistically smooth surfaces in this region since the wind direction should be highly influenced with the valley orientations. Model 3 in this regard seems more realistic and confirms the results obtained in Table 2.

6 | Conclusion

Since many environmental processes are associated with an angular component, this paper proposed a Bayesian hierarchical model to handle such features by allowing the joint modeling of extremes and angular components. An inferential framework has been proposed whose performances were assessed through a simulation study. As an aside, the lack of consistency for some Gaussian processes dependence parameters already found in the literature has been numerically illustrated, and it has been shown that, fortunately, this issue has no impact on the prediction of return levels. An application to extreme wind speeds and direction in continental France was conducted. Results show that extreme wind speeds in the Atlantic part of France are mainly dominated by West directions. A different behavior is seen for the

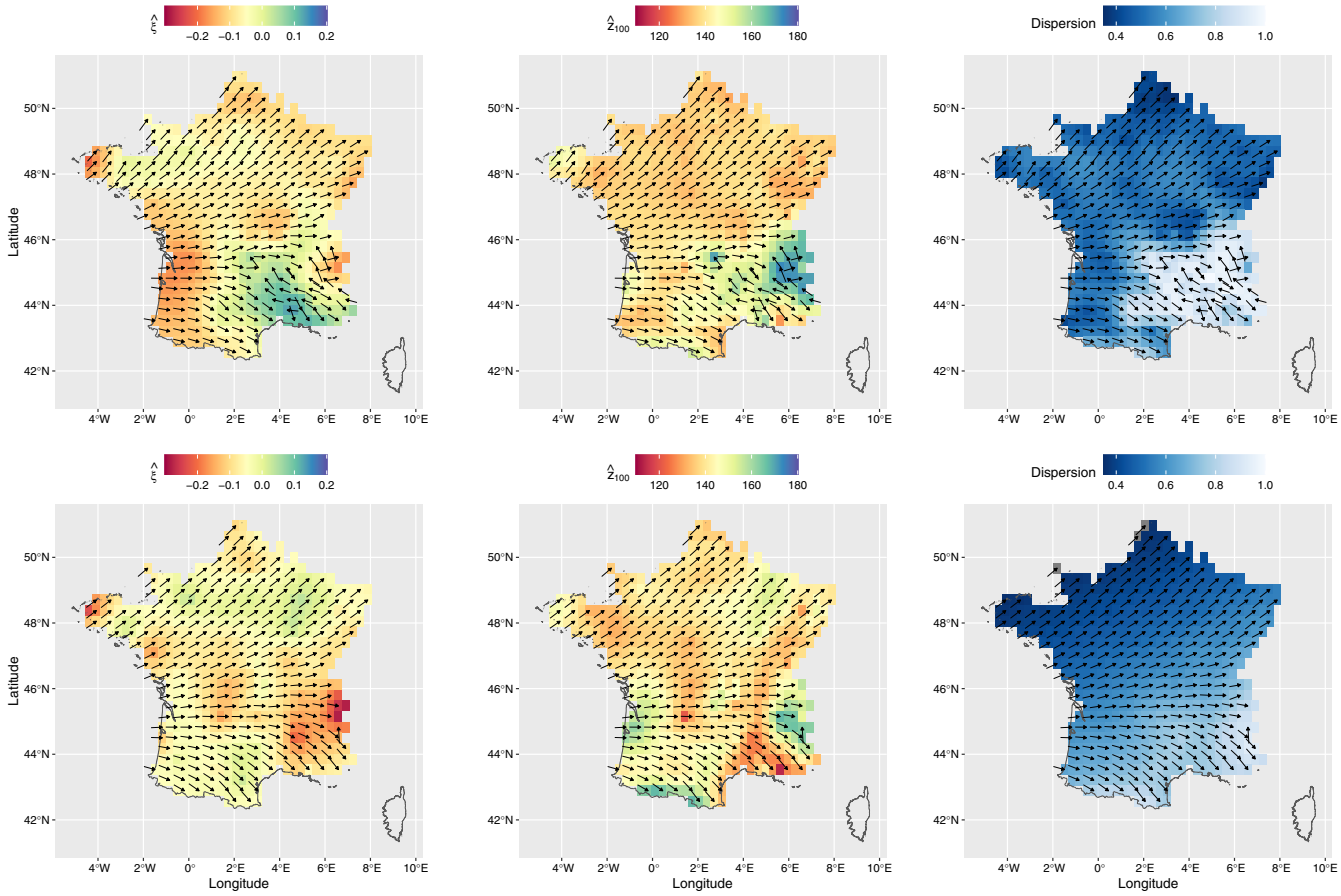

FIGURE 10 | Model 3 (top) and Model 0 (bottom) pointwise prediction maps of the wind direction and the GEV shape parameter (left), 100-year return level (middle) and the wind direction dispersion (right). For each plot, the arrows points to the (major) mode of the predictive posterior distribution.

Mediterranean part of France where extreme winds arise from multiple winds patterns, that is, North and South directions. This leads to a multimodal angular distribution, whose low-scale spatial variations are harder to recover from a large-scale model. Probably accurate modeling in this region would require a much more spatially dense data set.

Although this paper illustrates the usefulness of the proposed model, it has some limitations. First, our model applies to pointwise block-maxima data while practitioners may want to use exceedances above a threshold as in the original work of Cooley et al. (2007). Extending the proposed model to peaks over threshold should not be too complicated and will require to switch the GEV distribution for the Generalized Pareto distribution and to define a spatial model for the probability of exceeding the spatial threshold $\{u(s) : s \in \mathcal{X}\}$. Second, the proposed model is time stationary and cannot handle situations where there is a clear impact of climate change. An extension to the nonstationary case may be considered using nonstationary covariance structures, time-dependent covariates, or even time-dependent model parameters. Second, as a consequence of the conditional independence assumption, areal quantities such as the total amount of rainfall in a subdomain are not possible. To overcome this problem, the conditional independence assumption should be relaxed and other extreme-value models should be considered, such as max-stable (de Haan and Fereira 2006; Schlather 2002) or r-Pareto processes (Thibaud and Opitz 2015), max-mixture (Wadsworth and Tawn 2012; Engelke et al. 2019; Krupskii et al. 2018) or max-infinitely divisible processes (Padoan 2013)—see Huser and Wadsworth (2022) for a comparison of those models. The use of those processes is, however, challenging since the likelihood associated to these processes is extremely CPU demanding and one has to develop an elegant framework to overcome this computational burden.

Data Availability Statement

The data that support the findings of this study are made available by Météo-France. These data were derived from the following resources available in the public domain: "Observations du réseau sol de France" https://donneespubliques.meteofrance.fr/?fond=produit&id_produit=93&id_rubrique=32.

References

Akaike, H. 1974. "A New Look at the Statistical Model Identification." *IEEE Transactions on Automatic Control* 19, no. 6: 716–723.

Banerjee, S., B. Carlin, and A. Gelfand. 2014. *Hierarchical Modeling and Analysis for Spatial Data*. 2nd ed. Chapman and Hall/CRC.

Breckling, J. 1989. *The Analysis of Directional Time Series: Applications to Wind Speed and Direction*. Vol. 61. Springer Science & Business Media.

Casella, G., and E. I. George 1992. "Explaining the Gibbs Sampler." *American Statistician* 46, no. 3: 167–174.

Casson, E., and S. Coles. 1998. "Extreme Hurricane Wind Speeds: Estimation, Extrapolation and Spatial Smoothing." *Journal of Wind Engineering and Industrial Aerodynamics* 74: 131–140.

Casson, E., and S. Coles. 1999. "Spatial Regression Models for Extremes." *Extremes* 1, no. 4: 449–468.

Coles, S. 2001. *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics. Springer.

Cooley, D., D. Nychka, and P. Naveau. 2007. "Bayesian Spatial Modeling of Extreme Precipitation Return Levels." *Journal of the American Statistical Association* 102, no. 479: 824–840.

Davison, A., R. Huser, and E. Thibaud. 2013. "Geostatistics of Dependent and Asymptotically Independent Extremes." *Mathematical Geosciences* 45, no. 5: 511–529.

Davison, A., S. Padoan, and M. Ribatet. 2012. "Statistical Modelling of Spatial Extremes." *Statistical Science* 7, no. 2: 161–186.

de Haan, L. 1984. "A Spectral Representation for Max-Stable Processes." *Annals of Probability* 12, no. 4: 1194–1204.

de Haan, L., and A. Fereira. 2006. *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering.

de Haan, L., and A. Ferreira. 2014. "The Generalized Pareto Process; With Application." *Bernoulli* 20, no. 4: 1717–1737.

Dempster, A., N. Laird, and D. Rubin. 2018. "Maximum Likelihood From Incomplete Data via the Em Algorithm." *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 39, no. 1: 1–22.

Dixon, M., J. Tawn, and J. Vassie. 1998. "Spatial Modelling of Extreme Sea-Levels." *Environmetrics* 9, no. 3: 283–301.

Dombry, C., S. Engelke, and M. Oesting. 2017. "Bayesian Inference for Multivariate Extreme Value Distributions." *Electronic Journal of Statistics* 11, no. 2: 4813–4844.

Dombry, C., and M. Ribatet. 2015. "Functional Regular Variations, Pareto Processes and Peaks Over Threshold." *Statistics and Its Interface* 8, no. 1: 9–17.

Embrechts, P., F. Lindskog, and E. Mcneil. 2001. "Modelling Dependence With Copulas and Applications to Risk Management." In *Handbook of Heavy Tailed Distributions in Finance*, vol. 8.

Engelke, S., T. Opitz, and J. Wadsworth. 2019. "Extremal Dependence of Random Scale Constructions." *Extremes* 22: 623–666.

Gelfand, A., and A. Wang. 2013. "Directional Data Analysis Under the General Projected Normal Distribution." *Statistical Methodology* 10, no. 1: 113–127.

Gelfand, A., and A. Wang. 2014. "Modeling Space and Space-Time Directional Data Using Projected Gaussian Processes." *Journal of the American Statistical Association* 109: 1565–1580.

Genton, M., S. Padoan, and H. Sang. 2015. "Multivariate max-stable spatial processes." *Biometrika* 102, no. 1: 215–230.

Grimit, E., T. Gneiting, V. Berrocal, and N. Johnson. 2006. "The Continuous Ranked Probability Score for Circular Variables and Its Application to Mesoscale Forecast Ensemble Verification." *Quarterly Journal of the Royal Meteorological Society* 132, no. 621C: 2925–2942.

Gudendorf, G., and J. Segers. 2010. "Extreme-Value Copulas." In *Copula Theory and Its Applications*, edited by P. Jaworski, F. Durante, W. K. Härdle, and T. Rychlik, 127–145. Springer.

Hashorva, E., and A. Kume. 2021. "Multivariate Max-Stable Processes and Homogeneous Functionals." *Statistics & Probability Letters* 173: 109066.

Hastings, W. 1970. "Monte Carlo Sampling Methods Using Markov Chains and Their Applications." *Biometrika* 57: 97–109.

Holmes, J. D. 2015. *Wind Loading of Structures*. 3rd ed. CRC Press.

Huser, R., C. Dombry, M. Ribatet, and M. Genton. 2019. "Full-Likelihood Inference for Max-Stable Processes." *Stat* 8, no. 1: e218.

Huser, R., and J. Wadsworth. 2022. "Advances in Statistical Modeling of Spatial Extremes." *WIREs Computational Statistics* 14, no. 1: e1537.

Jammalamadaka, S., and A. SenGupta. 2001. *Topics in Circular Statistics*. World Scientific.

Jenkinson, A. F. 1955. "The Frequency Distribution of the Annual Maximum (Or Minimum) Values of Meteorological Elements." *Quarterly Journal of the Royal Meteorological Society* 81, no. 348: 158–171.

Jona-Lasinio, G., A. Gelfand, and M. Jona-Lasinio. 2012. "Spatial Analysis of Wave Direction Data Using Wrapped Gaussian Processes." *Annals of Applied Statistics* 6, no. 4: 1478–1498.

Krupskii, P., R. Huser, and M. Genton. 2018. "Factor Copula Models for Replicated Spatial Data." *Journal of the American Statistical Association* 113, no. 521: 467–479.

Mardia, K. 1972. *Statistics of Directional Data*. Academix Press.

Mardia, K., and P. Jupp. 2009. *Directional Statistics*. John Wiley and Sons.

Nelsen, R. 2006. *An Introduction to Copulas*. Springer Series in Statistics.

Obermann, A., S. Bastin, S. Belamari, et al. 2018. "Mistral and Tramontane Wind Speed and Wind Direction Patterns in Regional Climate Simulations." *Climate Dynamics* 51: 1059–1076.

Padoan, S. 2013. "Extreme Dependence Models Based on Event Magnitude." *Journal of Multivariate Analysis* 122: 1–19.

Padoan, S., M. Ribatet, and S. Sisson. 2010. "Likelihood-Based Inference for Max-Stable Processes." *Journal of the American Statistical Association* 105, no. 489: 263–277.

Ribatet, M., D. Cooley, and A. Davison. 2012. "Bayesian Inference From Composite Likelihoods, With an Application to Spatial Extremes." *Statistica Sinica* 22: 813–845.

Ribatet, M., and M. Sedki. 2013. "Extreme Value Copulas and Max-Stable Processes." *Journal de la Société Française de Statistique* 154, no. 1: 138–150.

Sang, L., J. Murata, M. Morimoto, Y. Kamada, T. Maeda, and Q. Li. 2017. "Experimental Investigation of Load Fluctuation on Horizontal Axis Wind Turbine for Extreme Wind Direction Change." *Journal of Fluid Science and Technology* 12, no. 1: JFST0005.

Schlather, M. 2002. "Models for Stationary Max-Stable Random Fields." *Extremes* 5, no. 1: 33–44.

SFEC. 2023. Stratégie française pour l'énergie et le climat. Technical report, Ministere de la transition énergétique.

Thibaud, E., and T. Opitz. 2015. "Efficient Inference and Simulation for Elliptical Pareto Processes." *Biometrika* 102, no. 4: 855–870.

Wadsworth, J., and J. Tawn. 2012. "Dependence Modelling for Spatial Extremes." *Biometrika* 99, no. 2: 253–272.

Watanabe, S. 2013. "A Widely Applicable Bayesian Information Criterion." *Journal of Machine Learning Research* 14, no. 27: 867–897.

Zhang, H. 2004. "Inconsistent Estimation and Asymptotically Equal Interpolations in Model-Based Geostatistics." *Journal of the American Statistical Association* 99, no. 465: 250–261.

Appendix A

Gibbs Sampler

Inference for our latent extreme–angular Bayesian hierarchical model may be performed using a Gibbs sampler, whose steps we now

describe. To ease notations, we define $\boldsymbol{R}_t = \{R_t(s_1), \ldots, R_t(s_k)\}$ and $\boldsymbol{\mu}_t = \{\mu_t(s_1), \ldots, \mu_t(s_k)\}$ with similar notations for the GEV scale and shape parameters. Given a current value of the Markov chain

$$\boldsymbol{\Upsilon}_t = (\boldsymbol{R}_t, \boldsymbol{\mu}_t, \boldsymbol{\sigma}_t, \boldsymbol{\xi}_t, \rho_{\theta,t}, \tau_{\mu,t}, \tau_{\theta,t}, \tau_{\sigma,t}, \tau_{\xi,t}, \lambda_{\theta,t}, \lambda_{\mu,t}, \lambda_{\sigma,t}, \lambda_{\xi,t}, \boldsymbol{\beta}_{\theta,t}, \boldsymbol{\beta}_{\mu,t}, \boldsymbol{\beta}_{\sigma,t}, \boldsymbol{\beta}_{\xi,t})$$

the next state $\boldsymbol{\Upsilon}_{t+1}$ of the chain is obtained as follows:

Step 1: Updating the GEV parameters at each site

Each component of $\mu(\mathbf{s}) = \{\mu(s_1), \ldots, \mu(s_k)\}$ is updated singly according to the following scheme. Generate a proposal $\mu_p(\mathbf{s})$ from a symmetric random walk and compute the acceptance probability

$$\alpha\{\mu(\mathbf{s}), \mu_p(\mathbf{s})\} = \min\left\{1, r_1\{\mu(\mathbf{s}), \mu_p(\mathbf{s})\} r_2\{\mu(\mathbf{s}), \mu_p(\mathbf{s})\} r_3\{\mu(\mathbf{s}), \mu_p(\mathbf{s})\}\right\}$$

with

$$r_1\{\mu(\mathbf{s}), \mu_p(\mathbf{s})\} = \prod_{j=1}^{k} \frac{\pi\{\eta(s_j)|\mu_p(s_j), \sigma(s_j), \xi(s_j)\}}{\pi\{\eta(s_j)|\mu(s_j), \sigma(s_j), \xi(s_j)\}}$$

$$r_2\{\mu(\mathbf{s}), \mu_p(\mathbf{s})\} = \prod_{i=1}^{n} \frac{\pi\{R_i(\mathbf{s}), \theta_i(\mathbf{s})|\mu_p(\mathbf{s}), \sigma(\mathbf{s}), \xi(\mathbf{s}), \boldsymbol{\beta}_\theta, \rho_\theta, \tau_\theta\}}{\pi\{R_i(\mathbf{s}), \theta_i(\mathbf{s})|\mu(\mathbf{s}), \sigma(\mathbf{s}), \xi(\mathbf{s}), \boldsymbol{\beta}_\theta, \rho_\theta, \tau_\theta\}}$$

$$r_3\{\mu(\mathbf{s}), \mu_p(\mathbf{s})\} = \frac{\pi(\boldsymbol{\mu}_p|\boldsymbol{\beta}_\mu, \tau_\mu, \lambda_\mu)}{\pi(\boldsymbol{\mu}|\boldsymbol{\beta}_\mu, \tau_\mu, \lambda_\mu)}$$

where r_1 is a ratio of GEV likelihoods, r_2 is a ratio of projected Gaussian likelihoods, and r_3 is a ratio of multivariate Gaussian likelihoods. With probability $\alpha\{\mu(\mathbf{s}), \mu_p(\mathbf{s})\}$, the $\mu(\mathbf{s})$ component of $\boldsymbol{\Upsilon}_{t+1}$ is set to $\mu_p(\mathbf{s})$; otherwise, it remains at $\mu(\mathbf{s})$. The scale and shape parameters are updated similarly.

Due to possible components related to μ, σ, or ξ, the design matrix D_θ (related to the regression parameter β_θ) needs to be updated each time one of those parameters is changed.

Step 2: Updating the radius at each site and replicating

Components of $\mathbf{R}_i = \{R_i(s_1), \ldots, R_i(s_k)\}$, $i = 1, \ldots, n$, are updated one by one according to the following scheme. Generate a proposal $R_{p,i}(s_j)$ from a log-normal distribution and accept with probability

$$\alpha\{R_i(s_j), R_{p,i}(s_j)\} = \min\left\{1, \frac{\pi(\mathbf{R}_{p,i}, \boldsymbol{\theta}_i|\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\xi}, \boldsymbol{\beta}_\theta, \tau_\theta, \rho_\theta, \lambda_\theta)}{\pi(\mathbf{R}_i, \boldsymbol{\theta}_i|\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\xi}, \boldsymbol{\beta}_\theta, \tau_\theta, \rho_\theta, \lambda_\theta)}\right\}$$

that is, a ratio of radial Gaussian likelihoods, based on (4).

Step 3: Updating the angle regression parameters

Due to the use of conjugate priors, $\boldsymbol{\beta}_\theta$ is drawn directly from a multivariate Gaussian distribution having covariance matrix and mean vector

$$\{(\Sigma_\theta^*)^{-1} + nD_\theta^T\Sigma_\theta^{-1}D_\theta\}^{-1}, \quad \{(\Sigma_\theta^*)^{-1} + D_\theta^T\Sigma_\theta^{-1}D_\theta\}^{-1} \times \left\{(\Sigma_\theta^*)^{-1}\mu_\theta^* + D_\theta^T\Sigma_\theta^{-1}\sum_{i=1}^{n}\mathbf{X}_i\right\}$$

where μ_θ^* and Σ_θ^* are the mean vector and covariance matrix of the prior distribution of $\boldsymbol{\beta}_\mu$, D_θ is the design matrix related to the regression coefficients $\boldsymbol{\beta}_\theta$, $\mathbf{X}_i = \left(\mathbf{R}_i \cos\boldsymbol{\theta}_i \;\; \mathbf{R}_i \sin\boldsymbol{\theta}_i\right)^\top$ and Σ_θ the covariance matrix of $\mathbf{X}_i$.

Step 4: Updating the GEV regression parameters

Due to the use of conjugate priors, $\boldsymbol{\beta}_\mu$ is drawn directly from a multivariate Gaussian distribution having covariance matrix and mean vector

$$\{(\Sigma_\mu^*)^{-1} + D_\mu^T\Sigma_\mu^{-1}D_\mu\}^{-1}, \quad \{(\Sigma_\mu^*)^{-1} + D_\mu^T\Sigma_\mu^{-1}D_\mu\}^{-1}\{(\Sigma_\mu^*)^{-1}\mu_\mu^* + D_\mu^T\Sigma_\mu^{-1}\boldsymbol{\mu}\}$$

where μ_μ^* and Σ_μ^* are the mean vector and covariance matrix of the prior distribution for $\boldsymbol{\beta}_\mu$, D_μ is the design matrix related to the regression coefficients $\boldsymbol{\beta}_\mu$ and Σ_μ the covariance matrix of $\boldsymbol{\mu}$. Again the regression parameters for the GEV scale and shape parameters are updated similarly.

Step 5: Updating the sill parameters of the covariance function

Due to the use of conjugate priors, τ_μ is drawn directly from an inverse Gamma distribution whose shape and rate parameters are

$$\frac{k}{2} + \kappa_{\tau_\mu}^*, \quad \theta_{\tau_\mu}^* + \frac{1}{2}\tau_\mu(\boldsymbol{\mu} - D_\mu\boldsymbol{\beta}_\mu)^T\Sigma_\mu^{-1}(\boldsymbol{\mu} - D_\mu\beta_\mu)$$

where $\kappa_{\tau_\mu}^*$ and $\theta_{\tau_\mu}^*$ are, respectively, the shape and scale parameters of the inverse Gamma prior distribution. The sill parameters of the covariance function for the GEV scale and shape parameters are updated similarly.

Step 6: Updating the projected Gaussian parameters

To update the parameter τ_θ, we generate a proposal $\tau_{\theta,p}$ from a log-normal distribution and compute the acceptance probability

$$\alpha(\tau_\theta, \tau_{\theta,p}) = \min\left\{1, \frac{\pi(\mathbf{R}, \boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\xi}, \boldsymbol{\beta}_\theta, \tau_{\theta,p}, \rho, \lambda_\theta)\pi(\tau_{\theta,p}|\kappa_{\tau_\theta}^*, \theta_{\lambda_\theta}^*)}{\pi(\mathbf{R}, \boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\xi}, \boldsymbol{\beta}_\theta, \tau_{\theta,t}, \rho, \lambda_\theta)\pi(\tau_{\theta,t}|\kappa_{\tau_\theta}^*, \theta_{\lambda_\theta}^*)} \frac{\tau_{\theta,p}}{\tau_\theta}\right\}$$

a ratio of projected Gaussian likelihood times the ratio of the prior densities with a correction due to the use of nonsymmetric proposal distribution and where $\kappa_{\tau_\theta}^*$ and $\theta_{\lambda_\theta}^*$ are respectively the shape and scale parameters of the Gamma prior distribution. With probability $\alpha(\tau_\theta, \tau_{\theta,p})$, the τ_θ component of $\boldsymbol{\Upsilon}_{t+1}$ is set to $\tau_{\theta,p}$; otherwise, it remains at τ_θ. The parameter λ_θ is updated similarly as well as the parameter ρ_θ except that, for the latter parameter, we use the following symmetric proposal distribution: $\rho_p \sim \rho_\theta + U(-\epsilon_\rho, \epsilon_\rho)$ and consequently no correction like ρ_p/ρ_θ is required.

Step 7: Updating the range parameters of the covariance function

To update the parameter λ_μ, we generate a proposal $\lambda_{\mu,p}$ from a log-normal distribution and compute the acceptance probability

$$\alpha(\lambda_\mu, \lambda_{\mu,p}) = \min\left\{1, \frac{\pi(\boldsymbol{\mu}|\tau_\mu, \lambda_{\mu,p}, \boldsymbol{\beta}_\mu)}{\pi(\boldsymbol{\mu}|\tau_\mu, \lambda_\mu, \boldsymbol{\beta}_\mu)}\left(\frac{\lambda_{\mu,p}}{\lambda_\mu}\right)^{k_{\lambda_\mu}^*-1} \exp\left(\frac{\lambda_\mu - \lambda_{\mu,p}}{\theta_{\lambda_\mu}^*}\right) \times \frac{\lambda_{\mu,p}}{\lambda_\mu}\right\}$$

a ratio of multivariate Normal densities times the ratio of the prior densities and that of the proposal densities and where $\kappa_{\lambda_\mu}^*$ and $\theta_{\lambda_\mu}^*$ are respectively the shape and the scale parameters of the Gamma prior distribution. With probability $\alpha(\lambda_\mu, \lambda_{\mu,p})$, the λ_η component of $\boldsymbol{\Upsilon}_{t+1}$ is set to $\lambda_{\mu,p}$; otherwise it remains at λ_μ. The range parameters related to the scale and shape GEV parameters are updated similarly. If the covariance family has a shape parameter like the powered exponential or the Whittle–Matérn covariance functions, this is updated in the same way.