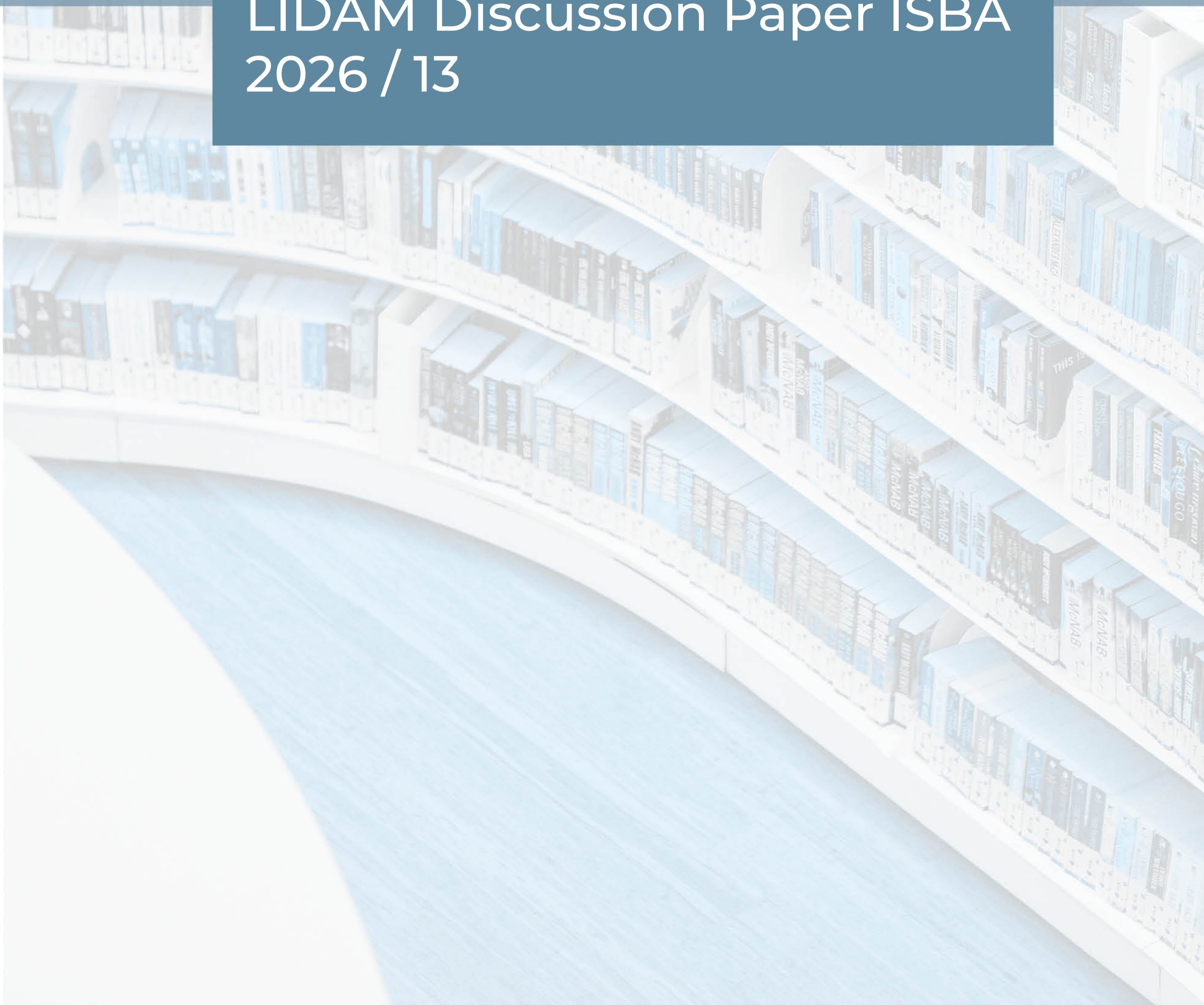
A SPATIO-TEMPORAL STATISTICAL FRAMEWORK FOR HEATWAVE ATTRIBUTION UNDER CLIMATE CHANGE

Kamal Gasser, Johan Segers, Francesco Ragone

LIDAM Discussion Paper ISBA
2026 / 13



UCLouvain

LIDAM
ISBA | Louvain Institute of Data Analysis and Modeling in economics and statistics

ISBA
Voie du Roman Pays 20 - L1.04.01
B-1348 Louvain-la-Neuve
Email : lidam-library@uclouvain.be
https://uclouvain.be/en/research-institutes/lidam/isba/publication

A spatio-temporal statistical framework for heatwave attribution under climate change

Kamal Gasser[1*], Johan Segers[1,2], Francesco Ragone[3]

[1] LIDAM/ISBA, UCLouvain, Belgium. [2]Dept. of Mathematics, KU Leuven, Belgium.

[3]School of Computing and Mathematical Sciences, University of Leicester, UK.

[*]Lead and corresponding author. Email: kamal.gasser@uclouvain.be

Abstract

We develop a unified statistical framework for attributing heatwaves as spatio-temporal phenomena under climate change. We quantify the impact of anthropogenic forcing on the probability and persistence of heatwaves not captured by standard marginal extreme-value approaches. Our methodology constructs a generative model for daily temperature fields that separates marginal nonstationarity from spatio-temporal dependence. We combine three components: a Bayesian spatial quantile regression model for the bulk of the data; a nonstationary spatial generalized extreme value model for tail behavior; and a copula-based model capturing both asymptotic dependence and independence in the extremes. The framework is applied to the CMIP6 MRI-ESM2 climate model, contrasting factual and counterfactual scenarios for probabilistic attribution. Our results show that the approach captures key heatwave characteristics inaccessible to traditional methods, enabling direct estimation of event-level attribution metrics. Overall, it provides a flexible basis for analyzing and attributing complex climate extremes as space-time objects.

Funding: This work was supported by the EXALT project funded by the Projet d'Actions de Recherche Concertées No. 24/29-146 of the Communauté française de Belgique.

Code availability: https://github.com/kamalgasser/heatwave-attribution-paper

1 Introduction

In recent decades, the global climate has undergone significant warming due primarily to anthropogenic activity, with global mean surface temperature (GMST) rising by approximately 1.1°C since the pre-industrial period (Intergovernmental Panel on Climate Change (IPCC), 2021). The IPCC Sixth Assessment Report emphasizes that even modest shifts in mean temperature can induce substantial increases in the frequency, intensity, and duration of extreme events (Seneviratne et al., 2021). Among these, heatwaves pose severe risks to human health, agriculture, and ecosystems (Perkins, 2015; Risser et al., 2025). Europe has been particularly affected: the 2003 heatwave led to more than 70 000 excess deaths (Zaitchik et al., 2006), and more recent events in 2019 and 2022 have broken temperature records by unprecedented margins (Ballester et al., 2023; Vautard et al., 2020).

From a statistical perspective, a key challenge is to quantify how anthropogenic forcing alters the probability of such events. However, heatwaves are inherently spatio-temporal phenomena, characterized jointly by high temperatures, temporal persistence over consecutive days, and spatial coherence across regions. Many quantities of interest such as the probability, duration, or spatial extent of a regional heatwave are therefore defined at the level of the entire space-time field, rather than at a single location or time point. Standard approaches based on marginal extreme-value theory are not designed to estimate such quantities, as they discard temporal persistence and spatial dependence.

The statistical literature on extreme event attribution is currently dominated by two main approaches. First, index-based methods model spatially aggregated quantities, such as regional mean temperature or predefined heatwave indices, using parametric extreme value theory (EVT) models with covariates such as GMST anomalies (van Oldenborgh et al., 2022; Wake, 2025). While computationally convenient, these approaches collapse the multi-dimensional structure of heatwaves into a one-dimensional summary and do not provide a probabilistic description of their spatial extent or temporal evolution. Second, methods based on pointwise or spatially smoothed extreme-value models estimate marginal distributions at individual locations (Auld et al., 2023), but treat spatial and temporal dependence only indirectly. As a result, they cannot directly quantify the probability of events defined by simultaneous exceedances across space and time.

In the spatial extremes literature, dependence is often modeled using max-stable processes, which arise as asymptotic limits of normalized maxima (Davison et al., 2019; Genton et al., 2015; Huser & Davison, 2014). While theoretically appealing, these models are restricted to asymptotic dependence and can be too rigid for environmental applications, where data frequently exhibit weakening dependence at extreme levels (Huser & Wadsworth, 2022). This limitation has motivated the development of sub-asymptotic models capable of capturing both asymptotic dependence and independence (Dell'Oro & Gaetan, 2024; Huser & Wadsworth, 2019; Huser et al., 2017; Simpson & Wadsworth, 2021; Wadsworth & Tawn, 2018). In the context of heatwaves, accurately representing this dependence structure is crucial, as it directly affects the probability of spatially extended and temporally persistent events (Li & Castro-Camilo, 2025).

In this paper, we propose a statistical framework for modeling heatwaves as coherent spatio-temporal events under nonstationary climate conditions. The key idea is to separate marginal nonstationarity from spatio-temporal dependence, allowing each component to be modeled flexibly while yielding a coherent joint representation of daily temperature fields. This enables the estimation of event-level quantities, such as the maximum intensity and duration of regional heatwaves, which are not directly accessible using standard marginal approaches.

Our approach combines three components. First, a Bayesian spatial quantile regression (BSQR) model (Reich et al., 2011) is used to capture the nonstationarity in the bulk of the temperature distribution and to transform observations to a uniform scale. Second, a nonstationary spatial generalized extreme value (GEV) model is employed to represent the marginal tail behavior of seasonal maxima (Cooley & Sain, 2010; Dyrrdal et al., 2015; Sang & Gelfand, 2009). Third, a flexible sub-asymptotic spatio-temporal dependence model (Dell'Oro & Gaetan, 2024; Huser & Wadsworth, 2019) is used to capture the joint behavior of daily temperatures across space and time. Together, these components define a generative model for daily temperature fields that preserves both marginal nonstationarity and realistic dependence structures.

The contribution of this work is the development of a coherent statistical framework that enables the analysis of heatwaves as spatio-temporal objects through the combination of several model components. By combining marginal modeling with flexible dependence structures, the proposed approach allows the estimation of quantities relevant for extreme event attribution that cannot be obtained from standard marginal EVT methods. In particular, it provides a probabilistic characterization of heatwave occurrence, persistence, and spatial extent under both factual and counterfactual climate scenarios.

The remainder of the paper is organized as follows. Section 2 describes the climate model data and introduces the spatio-temporal definition of a heatwave. Section 3 presents the statistical methodology. Section 4 applies the framework to CMIP6 simulations to quantify the influence of anthropogenic forcing on heatwave characteristics. Section 5 concludes with a discussion of the results and implications for extreme event attribution.

2 Climate Model Data and Heatwave Definition

2.1 Climate Model Data and Temperature Anomalies

For the attribution analysis, we use climate model simulations from the Coupled Model Intercomparison Project Phase 6 (CMIP6). Among the available CMIP6 models, we focus exclusively on the Meteorological Research Institute Earth System Model version 2 (MRI-ESM2) (Yukimoto et al., 2019), and which has been widely used in recent detection and attribution analyses (Bône et al., 2023; Engdaw et al., 2023). The MRI-ESM2 model provides daily maximum near-surface air temperature (TASMAX) under multiple forcing scenarios. In particular, we use:

- the historical experiment ("hist"), which includes both anthropogenic and natural forcings and represents the *factual* world;
- the natural-forcing-only experiment ("hist-nat"), which excludes anthropogenic forcings and represents the *counterfactual* world.

Both experiments cover the period 1850–2014, with the hist-nat simulation extending to 2020. To extend the factual world beyond 2014, we append the Shared Socioeconomic Pathway 2-4.5 (SSP245) scenario to the historical simulation which is a common practice in attribution study (Engdaw et al., 2023; Philip et al., 2022; Quilcaille et al., 2025). SSP245 corresponds to a "middle-of-the-road" socioeconomic pathway and a radiative forcing level of $4.5\,\mathrm{W\,m^{-2}}$ by 2100, leading to an expected global mean temperature increase of approximately $2\,^{\circ}\mathrm{C}$ to $3\,^{\circ}\mathrm{C}$ above preindustrial levels by the end of the century (Tebaldi et al., 2021). This extension allows the factual simulations to cover the period up to 2020. In summary, the factual world is represented by the MRI-ESM2 historical and SSP245 simulations, while the counterfactual world is represented by the hist-nat simulation. Comparing these two ensembles enables the quantification of the influence of anthropogenic forcing on extreme temperature events.

Let $Y^{\mathrm{F}}(\mathbf{s}, d, t)$ and $Y^{\mathrm{C}}(\mathbf{s}, d, t)$ denote daily maximum summer (June–July–August, JJA) temperatures in the factual and counterfactual worlds, respectively. Here, $\mathbf{s} = (s_x, s_y)$ denotes the longitude and latitude of a grid-cell center, with $\mathbf{s} \in \{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$, while $d \in \{1, \ldots, D\}$ indexes the day of the summer season with $D = 92$ and $t \in \{1, \ldots, T\}$ indexes the year with $T = 171$. For each grid cells we have a total of $DT = 15\,732$ recorded values. To remove the mean seasonal cycle and focus on extremes relative to a preindustrial baseline, temperature values are converted into daily anomalies by using the $N = 100$ first years:

$$X(\mathbf{s}, d, t) = Y(\mathbf{s}, d, t) - \frac{1}{N} \sum_{t'=1}^{N} Y(\mathbf{s}, d, t'). \tag{1}$$

The transformation is applied to both factual and counterfactual simulations. The baseline period in both cases is 1850–1949, which corresponds to an early-industrial climate with minimal anthropogenic influence.

2.2 Definition of heatwaves

Heatwaves do not have a universally agreed definition and are typically characterized by both extreme temperature and persistence. In the literature, they are commonly defined using high percentile thresholds relative to local climatology combined with a minimum duration criterion (Fischer & Schär, 2010; Perkins, 2015; Vautard et al., 2013). In this work, we adopt a percentile-based and duration-based definition aligned with the climate extremes literature (Perkins, 2015; Stefanon et al., 2012). The definition below is intended as an operational indicator for attribution and simulation purposes rather than as a new conceptual definition of heatwaves. A grid cell at location $\mathbf{s}$ is defined as experiencing a heatwave on summer day d in year t if the following indicator equals one:

$$H(\mathbf{s}, d, t) = \mathbb{1}\left\{X(\mathbf{s}, d-2, t) \geq \vartheta_{\mathbf{s}}^{\mathrm{thres}},\ X(\mathbf{s}, d-1, t) \geq \vartheta_{\mathbf{s}}^{\mathrm{thres}},\ X(\mathbf{s}, d, t) \geq \vartheta_{\mathbf{s}}^{\mathrm{thres}}\right\}, \tag{2}$$

that is, if daily temperature anomalies exceed a high threshold for at least three consecutive days. The threshold $\vartheta_{\mathbf{s}}^{\mathrm{thres}}$ is defined as the 95th percentile of the counterfactual anomaly distribution, which is assumed to be stationary both within and across years:

$$\vartheta_{\mathbf{s}}^{\text{thres}} = \inf\left\{x \in \mathbb{R} : \mathsf{P}\left(X^{\mathrm{C}}(\mathbf{s}, d, t) \leq x\right) \geq 0.95\right\}. \tag{3}$$

In practice, we use the empirical counterpart of the threshold, defined as

$$\hat{\vartheta}_{\mathbf{s}}^{\text{thres}} = \inf\left\{x \in \mathbb{R} : \frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T}\mathbb{1}\left\{X^{\mathrm{C}}(\mathbf{s}, d, t) \leq x\right\} \geq 0.95\right\}. \tag{4}$$

Defining the threshold with respect to the counterfactual climate ensures that heatwaves are identified relative to a stationary baseline unaffected by anthropogenic warming, facilitating a consistent comparison between the factual and counterfactual worlds.

Finally, let $\mathcal{R} \subset \mathbb{R}^2$ denote a spatial region of interest, composed of a collection of grid cell centers $\{\mathbf{s}_1, \ldots, \mathbf{s}_{n_\mathcal{R}}\}$. Let $a(\mathbf{s})$ denote the surface area (in km^2) of the cell of which $\mathbf{s}$ is the center. We define the regional heatwave indicator as

$$H_{\mathcal{R}}(d, t) = \mathbb{1}\left\{\sum_{\mathbf{s}\in\mathcal{R}} a(\mathbf{s})\, H(\mathbf{s}, d, t) \ \geq \ \alpha \sum_{\mathbf{s}\in\mathcal{R}} a(\mathbf{s})\right\}, \tag{5}$$

where $H(\mathbf{s}, d, t)$ is the local heatwave indicator defined in (2), and $\alpha \in (0, 1)$ denotes the minimum fraction of the region's total surface area that must be affected to declare a regional heatwave event. In this work, we set $\alpha = 0.6$. This fractional area criterion is in accordance with Rousi et al. (2022), Stefanon et al. (2012), and Tian et al. (2024). Under this definition, a region $\mathcal{R}$ is said to experience a heatwave on day d of year t if all grid cells experiencing a heatwave simultaneously cover at least 60% of the region's total surface area.

3 Methods

Our objective is to model and simulate nonstationary daily summer temperature fields under the factual climate scenario, and stationary temperature fields under the counterfactual scenario, in order to study the occurrence, persistence, and spatial extent of heatwaves. Heatwaves, as defined in Section 2, Eqs. (2) and (5), depend jointly on high quantiles of the daily temperature distribution, persistence over consecutive days, and spatial coherence across a region. Capturing these features requires a framework that accommodates nonstationary marginal behavior driven by external covariates while allowing flexible spatio-temporal dependence at the daily scale. Recall that, as discussed in Section 2.1, seasonality has been approximately removed by restricting the analysis to summer months, under the assumption that seasonal effects primarily impact the mean of the temperature distribution.

We adopt the working assumption that, for a given year t, daily temperature anomalies are marginally stationary within the summer season. Let $F_{\mathbf{s},d,t}$ be the distribution function of $X(\mathbf{s}, d, t)$. We assume that

$$\forall\, \mathbf{s} \in \mathcal{R},\ \forall\, t \in \{1, \dots, T\},\ \exists\, F_{\mathbf{s},t} \quad \text{such that} \quad F_{\mathbf{s},1,t} = \cdots = F_{\mathbf{s},D,t} = F_{\mathbf{s},t}, \tag{6}$$

where $F_{\mathbf{s},t}$ denotes the distribution function of $X(\mathbf{s}, d, t)$ in year t. This assumption implies that, conditional on the year, daily observations at a given location are identically distributed, while the marginal distribution is allowed to evolve across years. Temporal nonstationarity is introduced exclusively through year-specific covariates, while within-season variability is captured through the dependence structure.

To address these challenges, we decompose the modeling problem into three complementary components, each targeting a specific aspect of the heatwave process:

- A BSQR model is used to capture the nonstationary evolution of the bulk of the temperature distribution across space and time. This component allows us to define heatwave thresholds relative to a counterfactual baseline and to account for changes in the distribution under climate forcing. However, as we show in Section 4, the BSQR model alone is insufficient to accurately represent the upper tail of the distribution.

- A GEV model to address the limitation of the BSQR. We model the marginal tail behavior using a nonstationary spatial GEV model applied to seasonal maxima. This component provides a theoretically justified representation of extreme values and corrects the underestimation of tail risk observed when relying solely on the bulk model.

- Finally, a copula-based spatio-temporal dependence model is introduced to capture the joint behavior of daily temperatures across space and time. This component is essential to represent the persistence and spatial coherence of heatwaves, which cannot be recovered from marginal models alone.

Together, these three components define a coherent generative model for daily temperature fields that preserves both marginal nonstationarity and realistic dependence structures.

3.1 Bayesian Spatial Quantile Regression Model

This section presents the Bayesian hierarchical model corresponding to the *approximate method* of Reich et al. (2011, Sec. 3). We use the BSQR framework to model the conditional quantile function of temperature anomalies $X(\mathbf{s}, d, t)$ given covariates $\mathbf{c}(t)$. Let $q(\tau \mid \mathbf{c}(t), \mathbf{s})$ denote the conditional τ-th quantile of $X(\mathbf{s}, d, t)$, modeled as

$$q(\tau \mid \mathbf{c}(t), \mathbf{s}) = \mathbf{c}(t)^\top \boldsymbol{\beta}^{\mathrm{QR}}(\tau, \mathbf{s}), \tag{7}$$

where $\boldsymbol{\beta}^{\mathrm{QR}}(\tau, \mathbf{s})$ are spatially varying quantile regression coefficients. Following Reich et al. (2011), inference is performed using a two-stage approximation. First, for each grid cell center $\mathbf{s}$, classical quantile regression is applied at levels $\tau_1, \dots, \tau_K$, yielding estimates $\hat{\boldsymbol{\beta}}^{\mathrm{QR}}(\mathbf{s}) \in \mathbb{R}^{pK}$ and associated covariance matrices $\hat{\Sigma}(\mathbf{s})$. These estimators are consistent and asymptotically normal; details on their computation are provided in Supplement S1.1.

In the second stage, these estimates are treated as observations and linked to the latent coefficients through a Gaussian likelihood:

$$\hat{\boldsymbol{\beta}}^{\mathrm{QR}}(\mathbf{s}) \sim \mathcal{N}\big(\boldsymbol{\beta}^{\mathrm{QR}}(\mathbf{s}),\, \hat{\Sigma}(\mathbf{s})\big), \qquad \mathbf{s} \in \mathcal{R}. \tag{8}$$

To model variation across quantile levels, each coefficient function is represented using a Bernstein polynomial basis,

$$\beta_j^{\mathrm{QR}}(\tau, \mathbf{s}) = \sum_{m=1}^{M} B_m(\tau)\, \alpha_{jm}(\mathbf{s}), \tag{9}$$

with M fixed. To enforce non-crossing quantiles, we adopt the increment parametrization of Reich et al. (2011), defining $\alpha_{jm}(\mathbf{s}) = \sum_{\ell=1}^{m} \delta_{j\ell}(\mathbf{s})$, where constraints on $\delta_{jm}(\mathbf{s})$ ensure monotonicity in τ. Spatial dependence is introduced by assigning Gaussian process priors to latent variables $\delta^*_{jm}(\mathbf{s})$, which determine the increments and hence the quantile coefficient functions. Their prior mean is centered on a parametric base distribution indexed by Ω, while $\tilde{\sigma}_j^2$ and ρ_j control respectively the marginal variance and spatial range of the Gaussian process associated with covariate j. For notational convenience, let $\boldsymbol{\beta}^{\mathrm{QR}}(\mathbf{s}) \in \mathbb{R}^{pK}$ denote the vector obtained by evaluating $\beta_j^{\mathrm{QR}}(\tau, \mathbf{s})$ at $\tau_1, \dots, \tau_K$ and stacking over $j = 1, \dots, p$.

Let $\Theta^{\mathrm{QR}}_{\mathcal{R}}$ denote the collection of all unknown parameters. The posterior distribution is given by

$$\begin{aligned}\pi\left(\Theta^{\mathrm{QR}}_{\mathcal{R}} \mid \{\hat{\boldsymbol{\beta}}^{\mathrm{QR}}(\mathbf{s})\}_{\mathbf{s}\in\mathcal{R}}\right) \propto \prod_{\mathbf{s}\in\mathcal{R}} \mathcal{N}\Big(\hat{\boldsymbol{\beta}}^{\mathrm{QR}}(\mathbf{s}) \mid \boldsymbol{\beta}^{\mathrm{QR}}(\mathbf{s}), \hat{\Sigma}(\mathbf{s})\Big) \\ \times \pi\big(\{\delta^*_{jm}(\cdot)\} \mid \Omega, \{\tilde{\sigma}^2_j, \rho_j\}\big)\, \pi_\Omega(\Omega) \prod_{j=1}^{p} \pi(\tilde{\sigma}^2_j)\pi(\rho_j).\end{aligned} \tag{10}$$

This formulation yields spatially smooth, non-crossing quantile functions while substantially reducing computational complexity compared to the full likelihood approach. Full details of the construction, including covariance estimation, the non-crossing parametrization, and prior specification, are provided in Supplement S1.1.

3.2 Marginal modeling of the tail (Univariate extreme value theory)

The classical starting point of EVT is the modeling of block maxima. Let $Y_1, \ldots, Y_n$ be independent and identically distributed random variables with common distribution function F, and define the block maximum $M_n = \max(Y_1, \ldots, Y_n)$. If sequences $\{a_n\}$ with $a_n > 0$ and $\{b_n\}$ exist such that the normalized maximum $M^*_n = (M_n - b_n)/a_n$ converges in distribution to a non-degenerate limit, then the limiting distribution must belong to the GEV family (Coles, 2001; Fisher & Tippett, 1928; Jenkinson, 1955). The GEV distribution function is given by

$$G(z) = \begin{cases} \exp\Big[-\{\max(1+\xi(z-\mu)/\sigma, 0)\}^{-1/\xi}\Big], & \xi \neq 0, \\ \exp\big[-\exp\{-(z-\mu)/\sigma\}\big], & \xi = 0, \end{cases} \tag{11}$$

where $\mu \in \mathbb{R}$ is a location parameter, $\sigma > 0$ is a scale parameter, and $\xi \in \mathbb{R}$ is a shape parameter controlling tail behavior.

For each grid cell center $\mathbf{s} \in \mathcal{R}$, we observe daily maxima within each summer season and extract the annual maximum. Let $M_{\mathbf{s},t} = \max_{d=1,\ldots,D} X(\mathbf{s}, d, t)$ denote the annual maximum over all days in year t for the grid cell of which $\mathbf{s}$ is the center. We further observe p time-varying covariates $\mathbf{c}(t) \in \mathbb{R}^p$ common to all sites as in Section 3.1. These covariates enter the marginal GEV parameters, yielding a nonstationary spatial GEV model (Coles, 2001)

$$M_{\mathbf{s},t} \sim \mathrm{GEV}\big(\mu_{\mathbf{s}}(t), \sigma_{\mathbf{s}}(t), \xi_{\mathbf{s}}\big). \tag{12}$$

We model the location and scale parameters through linear predictors

$$\mu_{\mathbf{s}}(t) = \mu_{0,\mathbf{s}} + c_1(t)\mu_{1,\mathbf{s}} + c_2(t)\mu_{2,\mathbf{s}} + \cdots + c_p(t)\mu_{p,\mathbf{s}}, \tag{13}$$

$$\log(\sigma_{\mathbf{s}}(t)) = \log(\sigma_{0,\mathbf{s}}) + c_1(t)\sigma_{1,\mathbf{s}} + c_2(t)\sigma_{1,\mathbf{s}} + \cdots + c_p(t)\sigma_{p,\mathbf{s}}, \tag{14}$$

while the shape parameter $\xi_{\mathbf{s}}$ is assumed to remain constant over time. The location and scale parameters $\mu_{\mathbf{s}}(t)$ and $\sigma_{\mathbf{s}}(t)$ are expressed in °C, while the shape parameter $\xi_{\mathbf{s}}$ is dimensionless.

Because nearby locations are exposed to the same regional extreme events, the GEV parameters are expected to vary smoothly over space. To encode this spatial structure while retaining interpretable global effects, we adopt a hierarchical Bayesian formulation based on an Intrinsic Conditional Autoregressive (ICAR) prior, which induce intrinsic Gaussian Markov random Field (GMRF) structures(Ferreira, 2024; Rue & Held, 2005). This approach is similar to what has been done by Cooley and Sain (2010), Dyrrdal et al. (2015), and Sang and Gelfand (2009). For each coefficient type ℓ (e.g., intercept or covariate slope in μ or $\log\sigma$, or the shape parameter ξ), define the spatial coefficient vector

$$\boldsymbol{\gamma}_\ell = (\gamma_{\ell,1}, \ldots, \gamma_{\ell,n})^\top. \tag{15}$$

Rather than assigning an ICAR prior directly to $\boldsymbol{\gamma}_{\mathcal{R},\ell}$, we decompose each spatial coefficient vector into a global mean plus a spatially structured deviation. We assume this decomposition is specified independently for each coefficient type ℓ. Specifically,

$$\boldsymbol{\gamma}_{\mathcal{R},\ell} = \gamma_{\ell,g}\mathbb{1} + s_\ell \boldsymbol{u}_\ell, \qquad \boldsymbol{u}_\ell \sim \text{ICAR}(\boldsymbol{A}), \tag{16}$$

where $\mathbb{1}$ denotes the n-vector of ones, $\gamma_{\ell,g} \in \mathbb{R}$ is a global parameter, $s_\ell > 0$ is a spatial scale parameter, and $\boldsymbol{u}_\ell = (u_{\ell,1}, \ldots, u_{\ell,n})^\top$ is a latent ICAR field defined with respect to the spatial adjacency $n \times n$ matrix $\boldsymbol{A} = (a_{ik})$. Let $\mathcal{A}_{\mathbf{s}}$ denote the set of neighbors of the grid cell of which $\mathbf{s}$ is the center and define

$$\boldsymbol{S} = \text{diag}(|\mathcal{A}_{\mathbf{s}_1}|, \ldots, |\mathcal{A}_{\mathbf{s}_n}|) - \boldsymbol{A}, \qquad a_{ik} = \begin{cases} 1, & k \in \mathcal{A}_{\mathbf{s}_i}, \\ 0, & \text{otherwise.} \end{cases} \tag{17}$$

The intrinsic GMRF prior for $\boldsymbol{u}_\ell$ has improper density proportional to

$$\pi(\boldsymbol{u}_\ell) \propto \exp\left\{-\frac{1}{2}\boldsymbol{u}_\ell^\top \boldsymbol{S}_{\mathcal{R}} \boldsymbol{u}_\ell\right\},$$

restricted to the identifiable subspace defined by a sum-to-zero constraint within each connected component of the adjacency graph. Note that no separate precision parameter appears in this expression, as the overall marginal variance of $\boldsymbol{\gamma}_\ell$ is controlled by the scale parameter s_ℓ of Eq. (16). The quadratic form can be written as

$$\boldsymbol{u}_\ell^\top \boldsymbol{S}_{\mathcal{R}} \boldsymbol{u}_\ell = \frac{1}{2} \sum_{i=1}^{n} \sum_{k=1}^{n} a_{ik} \, (u_{\ell,i} - u_{\ell,k})^2,$$

which makes explicit that the ICAR prior penalizes squared differences between neighboring grid cells. Under the decomposition above, the implied spatial precision is $1/s_\ell^2$, so that larger values of s_ℓ correspond to weaker spatial smoothing. In the general formulation, the index j may range over an arbitrary number of covariates p. However, in our application we restrict attention to an intercept and a single covariate effect ($j = 0, 1$), corresponding to the intercept and the yearly GMST anomaly. Applying this decomposition to each GEV coefficient yields, for all $\mathbf{s} \in \mathcal{R}$,

$$\begin{aligned}
\mu_{j,\mathbf{s}} &= \mu_{j,g} + s_{\mu_j} \, u_{\mu_j,\mathbf{s}}, & \boldsymbol{u}_{\mu_j} &\sim \mathrm{ICAR}(\boldsymbol{A}), \qquad j = 0, 1, \\
\log(\sigma_{0,\mathbf{s}}) &= \log(\sigma_{0,g}) + s_{\sigma_0} \, u_{\sigma_0,\mathbf{s}}, & \boldsymbol{u}_{\sigma_0} &\sim \mathrm{ICAR}(\boldsymbol{A}), \\
\sigma_{1,\mathbf{s}} &= \sigma_{1,g} + s_{\sigma_1} \, u_{\sigma_1,\mathbf{s}}, & \boldsymbol{u}_{\sigma_1} &\sim \mathrm{ICAR}(\boldsymbol{A}), \\
\xi_{\mathbf{s}} &= \xi_g + s_\xi \, u_{\xi,\mathbf{s}}, & \boldsymbol{u}_{\xi} &\sim \mathrm{ICAR}(\boldsymbol{A}).
\end{aligned}$$

We assign weakly informative priors to the global and spatial scale parameters,

$$\begin{aligned}
\mu_{0,g} &\sim \mathcal{N}(m_0, s_0^2), & \mu_{1,g} &\sim \mathcal{N}(0, 2^2), \\
\log(\sigma_{0,g}) &\sim \mathcal{N}(\log s_0, 1^2), & \sigma_{1,g} &\sim \mathcal{N}(0, 1^2), \\
\xi_g &\sim \mathcal{N}(0, 0.1^2) \, \mathbb{1}_{[-0.4,\, 0.4]},
\end{aligned}$$

where m_0 and s_0 denote the empirical mean and standard deviation, respectively, of the observed annual maxima and $\mathcal{HN}$ is the half normal distribution. A schematic summary of the hierarchical construction is provided in Supplement S1.2.

Let $\Theta_{\mathcal{R}}^{\mathrm{GEV}}$ denote the collection of all unknown parameters in the spatial GEV model in a region $\mathcal{R}$, namely

$$\Theta_{\mathcal{R}}^{\mathrm{GEV}} = \left\{\mu_{0,\mathbf{s}}, \mu_{1,\mathbf{s}}, \sigma_{0,\mathbf{s}}, \sigma_{1,\mathbf{s}}, \xi_{\mathbf{s}} : \mathbf{s} \in \mathcal{R}\right\} \,\cup\, \left\{s_\psi, \boldsymbol{u}_\psi : \psi \in \{\mu_0, \mu_1, \sigma_0, \sigma_1, \xi\}\right\}. \qquad (18)$$

Assuming conditional independence of annual maxima across years given the model parameters, the posterior distribution is

$$\pi(\Theta_{\mathcal{R}}^{\mathrm{GEV}} \mid \{M_{\mathbf{s}_i,t}\}) \propto \prod_{i=1}^{n} \prod_{t=1}^{T} f_{\mathrm{GEV}}\big(M_{\mathbf{s}_i,t}; \mu_{\mathbf{s}_i}(t), \sigma_{\mathbf{s}_i}(t), \xi_{\mathbf{s}_i}\big)\ \pi(\Theta_{\mathcal{R}}^{\mathrm{GEV}}), \tag{19}$$

where $f_{\mathrm{GEV}}(\,\cdot\,; \mu, \sigma, \xi)$ denotes the GEV density and $\pi(\Theta_{\mathcal{R}}^{\mathrm{GEV}})$ is the joint prior implied by the hierarchical specification. Inference is performed using Markov chain Monte Carlo, yielding posterior samples for all marginal GEV parameters and spatial effects.

3.3 Spatio-temporal modeling

3.3.1 Model construction

We adopt a copula-based approach to construct processes with flexible extremal dependence structures. For a random vector $X = (X_1, \ldots, X_m)$ with marginals $X_j \sim F_j$ that are continuous, the m-dimensional copula function C_X is defined by

$$C_X(u_1, \ldots, u_m) = \mathsf{P}(F_1(X_1) \le u_1, \ldots, F_m(X_m) \le u_m), \qquad u_1, \ldots, u_m \in [0, 1].$$

According to Sklar's theorem (Sklar, 1959), any multivariate joint distribution can be expressed in terms of a copula. Moreover, when the marginal distributions F_j, for $j = 1, \ldots, m$, are continuous, the copula representation is unique. In this work, we assume that covariates affect only the marginal distributions F_j, while the copula function remains invariant. In particular, climate forcing is allowed to modify the marginal behavior of temperatures through the BSQR and GEV components, but the spatio-temporal dependence structure is assumed to be stationary and independent of the covariates.

Our objective is to construct a process $Z(\mathbf{s}, d)$ that exhibits the desired within-year extremal space-time dependence. For a fixed year t, we use the copula associated with $Z(\mathbf{s}, d)$ to model the dependence structure of $(X(\mathbf{s}, d, t))_{\mathbf{s},d}$. Consider the following equation:

$$Z(\mathbf{s}, d) = R(\mathbf{s}, d)^{\delta} \cdot W(\mathbf{s}, d)^{1-\delta}, \qquad \delta \in [0, 1], \tag{20}$$

where $R(\mathbf{s}, d)$ and $W(\mathbf{s}, d)$ denote two independent processes that are asymptotically dependent and asymptotically independent, respectively, and that have unit-Pareto marginals:

$$\mathsf{P}(R(\mathbf{s}, d) > r) = \mathsf{P}(W(\mathbf{s}, d) > r) = 1/r, \qquad r \ge 1.$$

The dependence characteristics of the process $Z(\mathbf{s}, d)$ are determined by the asymptotic dependence or independence of the processes $R(\mathbf{s}, d)$ and $W(\mathbf{s}, d)$. If $\delta \leq 1/2$ then the tail of $Z(\mathbf{s}, d)$ is driven by $W(\mathbf{s}, d)$ and the process is asymptotically independent, while if $\delta > 1/2$, the tail is driven by $R(\mathbf{s}, d)$ and the process is asymptotially dependent. The marginal distribution function $Q(z) = \mathsf{P}(Z(\mathbf{s}, d) \leq z)$ for $z \geq 1$ is

$$Q(z) = \begin{cases} 1 - [\delta/(2\delta - 1)z^{-1/\delta} - (1-\delta)/(2\delta - 1)z^{-1/(1-\delta)}], & \text{if } \delta \neq 1/2, \\ 1 - z^{-2}[2\log(z) + 1], & \text{if } \delta = 1/2. \end{cases} \tag{21}$$

Additional technical specifications are available in Dell'Oro and Gaetan (2024) and Huser and Wadsworth (2019).

In our application, we construct $R(\mathbf{s}, d)$ by transforming an underlying kind of finite Smith storm profile model $R^*(\mathbf{s}, d)$ (J. A. Smith & Karr, 1990; R. L. Smith, 1990); see Section 3.3.3 below. For each site $\mathbf{s}$ and day $d = 1, \ldots, D$, we define

$$R(\mathbf{s}, d) = \frac{1}{1 - F_{R^*}(R^*(\mathbf{s}, d))}, \tag{22}$$

where F_{R^*} is the marginal distribution function, namely $F_{R^*}(r) = \mathsf{P}(R^*(\mathbf{s}, d) \leq r)$. Similarly, we model $W(\mathbf{s}, d)$ by transforming an underlying Gaussian field; see Section 3.3.2 below. Specifically, let $W^*(\mathbf{s}, d)$ denote an anisotropic Gaussian process with standard normal margins, and let Φ be the standard normal cumulative distribution function. For each site $\mathbf{s}$ and day $d = 1, \ldots, D$, we define

$$W(\mathbf{s}, d) = \frac{1}{1 - \Phi\big(W^*(\mathbf{s}, d)\big)}. \tag{23}$$

These transformations are applied componentwise and map the Gaussian field $W^*(\mathbf{s}, d)$ and the process $R^*(\mathbf{s}, d)$ to a unit-Pareto process.

3.3.2 Gaussian process

Let $\{W^*(\mathbf{s}, d) : \mathbf{s} \in \mathcal{R}, d \in \{1, \ldots, D\}\}$ be a space-time Gaussian random field. The function $(\mathbf{s}_1, \mathbf{s}_2, d_1, d_2) \to C_{s,d}(\mathbf{s}_1, \mathbf{s}_2, d_1, d_2)$ defined on the product space $\mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R} \times \mathbb{R}$ is called the spatio-temporal covariance function. This function needs to be positive definite, meaning that for every finite set of real coefficients a_i and points $(\mathbf{s}_i, d_i)$ and as soon as not all coefficients a_i are zero, we have the nonnegativity constraint

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j C_{s,d}(\mathbf{s}_i, \mathbf{s}_j, d_i, d_j) > 0.$$

Under the assumption of second-order stationarity we have without loss of generality that $\mathsf{E}(W^*(\mathbf{s}, d)) = 0$ and

$$\mathsf{cov}(W^*(\mathbf{s}_i, d_i), W^*(\mathbf{s}_j, d_j)) = C_{s,d}(\boldsymbol{h}, l) \quad \text{where} \quad (\boldsymbol{h}, l) = (\mathbf{s}_i - \mathbf{s}_j, d_i - d_j), \tag{24}$$

that is, the covariance depends only on the spatial and temporal separation. Finally, a stationary covariance function is isotropic if it is rotation and translation invariant, meaning that $C_{s,d}(\boldsymbol{h}, l) = \tilde{C}_{s,d}(\|\boldsymbol{h}\|, |l|)$ with $\| \cdot \|$ the usual Euclidean norm and $\tilde{C}_{s,d}$ a positive definite function (Porcu et al., 2007).

A covariance function is called separable if it can be written as the product of two covariance functions, that is, $C_{s,d}(\boldsymbol{h}, l) = C_s(\boldsymbol{h})C_d(l)$ where $C_s(\cdot)$ denotes the spatial covariance and $C_d(\cdot)$ the temporal covariance. Separability is computationally convenient, especially for large spatio-temporal fields. Under the separability assumption we can define the variance–covariance matrix of the process $W^*(\mathbf{s}, d)$ as $\Sigma = \Sigma_s \otimes \Sigma_d$, with Σ_s and Σ_d the spatial and temporal covariance matrices and $\otimes$ the Kronecker product. Below, we model the spatial and temporal covariance functions separately.

Spatial covariance function

The assumption of isotropy is often made because it simplifies the spatial model and facilitates parameter estimation. However, this is a restrictive assumption and does not hold in many empirical applications (Deng, 2008). In cases where the covariance of a second-order stationary process depends on the direction, the spatial dependence is referred to as anisotropic. Isotropy implies that correlation decays only with distance regardless of the direction, resulting in circular or spherical isolines on a variogram map. Anisotropy introduces directional variation, which may produce elongated, elliptical patterns of semi-variance. A specific and frequently encountered form, called geometric anisotropy, can often be managed by applying a suitable linear transformation to the spatial coordinates (Schabenberger & Gotway, 2017). To correct for geometric anisotropy, one changes the coordinate system with an appropriate linear transformation. If $\mathbf{s}$ is the spatial coordinate of a geometric anisotropic process $W^*(\mathbf{s}, t)$, then $W^*(\boldsymbol{T}\mathbf{s}, t)$ has an isotropic covariance function, where

$$\boldsymbol{T} = \begin{pmatrix} \cos(\omega) & -\sin(\omega) \\ \sin(\omega) & \cos(\omega) \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \eta \end{pmatrix}. \tag{25}$$

Hence, geometric anisotropy is corrected by first rotating the coordinate system to align with the principal directions of the elliptical contours, which is governed by the parameter ω [rad] $\in [0, \pi]$, followed by scaling the axis to restore circular symmetry governed by the parameter $\eta \geq 1$.

We consider a stationary but anisotropic spatial correlation function in dimension 2 of the form

$$C_s(\mathbf{s}_i, \mathbf{s}_j; \theta) = C\left(\frac{\|\boldsymbol{T}(\mathbf{s}_i - \mathbf{s}_j)\|}{\theta}\right), \tag{26}$$

where C is a valid isotropic correlation function and θ [km] > 0 is a spatial range parameter. For our application, we specify C as a Cauchy-type correlation function as in Dell'Oro and Gaetan (2024), yielding

$$C_s(\mathbf{s}_i, \mathbf{s}_j; \theta) = \frac{1}{1 + \left(\frac{\|\boldsymbol{T}(\mathbf{s}_i - \mathbf{s}_j)\|}{\theta}\right)^2}. \tag{27}$$

With $\boldsymbol{T}$ defined as in Eq. (25), the parameters $\omega \in [0, \pi]$ and $\eta \in [1, \infty[$ control the orientation and strength of the anisotropy, respectively, while $\theta > 0$ determines the spatial range over which the correlation decays.

Temporal covariance function

The temporal covariance function between time steps is specified using the Matérn class (Matérn, 1960; Rue & Held, 2005): for days d_r and d_q, we put

$$C_d(d_r, d_q) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}\,|d_r - d_q|}{\kappa}\right)^{\nu} K_\nu\left(\frac{\sqrt{2\nu}\,|d_r - d_q|}{\kappa}\right) + \mathbb{1}(d_r = d_q), \tag{28}$$

where $\nu > 0$ is a smoothness parameter, κ [d] > 0 is a temporal range parameter, $\Gamma(\cdot)$ denotes the Gamma function, and $K_\nu(\cdot)$ is the modified Bessel function of the second kind of order ν. The indicator term ensures that $C_t(d, d) = 1$, so that the process has unit marginal variance. The Matérn smoothness parameter ν is difficult to estimate reliably from data, particularly in spatio-temporal models. In practice, it is common to set ν to a fixed value; this is what has been done by Healy et al. (2025) in the context of extreme spatial temperature events in Ireland. Recently, Cognot et al. (2025) report temporal Matérn smoothness estimates consistently below one for daily temperatures over France, with values clustering around 0.6 up to 0.9. Following this empirical evidence and common practice in applied spatio-temporal modeling, we fix ν to 0.8 in the data application rather than estimating it.

3.3.3 Storm process

We construct the latent process $R^*(\mathbf{s}, d)$ to represent episodic heatwave-like events with spatial extent and temporal persistence. Let $K \sim \text{Poisson}(\lambda)$ denote the number of heatwave episodes occurring during a given year. Conditionally on K, we let

$$\{A_k\}_{k=1}^K \overset{\text{iid}}{\sim} \text{Pareto}(1), \qquad \mathbb{P}(A_k > x) = x^{-1}, \quad x \geq 1,$$

where the variables A_k represent independent event intensities. In addition to the amplitudes A_k, we sample for each episode k a pair $(\boldsymbol{C}_k, T_k)$, where the location $\boldsymbol{C}_k$ is sampled uniformly from the set of grid cell centres $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ and T_k is sampled uniformly from $\{1, \dots, D\}$. The variable $\boldsymbol{C}_k$ represents the spatial centre of the heatwave episode, from which the event propagates spatially, while T_k denotes the day at which the episode reaches its peak intensity. The latent process is then defined as

$$R^*(\mathbf{s}, d) = \max_{k=1,\dots,K} A_k \exp\left(-\frac{\|\boldsymbol{T}(\mathbf{s} - \boldsymbol{C}_k)\|}{\rho_s} - \frac{|d - T_k|}{\rho_t}\right). \tag{29}$$

Here $\boldsymbol{T}$ denotes the anisotropy matrix defined in Eq. (25). The parameter ρ_s [km] > 0 controls the spatial decay of the episode away from its centre $\boldsymbol{C}_k$, while ρ_t [d] > 0 controls the temporal decay away from the peak day T_k. For a fixed spatial location $\mathbf{s}$ and day d, let $F_{R^*(\mathbf{s},d)}(x) = \mathsf{P}\{R^*(\mathbf{s}, d) \leq x\}$ denote the distribution function of the latent process. Using the Poisson construction of the episodic field and the Pareto(1) distribution of the episode intensities, the distribution function admits the explicit form

$$F_{R^*(\mathbf{s},d)}(x) = \begin{cases} 0, & x < 0, \\ \exp(-\lambda), & x = 0, \\ \exp\left\{-\lambda\left(1 - \dfrac{1}{nD}\displaystyle\sum_{i=1}^{n}\sum_{t=1}^{D}\left(1 - \dfrac{\exp\left(-\frac{\|\boldsymbol{T}(\mathbf{s}-\mathbf{s}_i)\|}{\rho_s} - \frac{|d-t|}{\rho_t}\right)}{x}\right)_+\right)\right\}, & x > 0, \end{cases} \tag{30}$$

where $(u)_+ = \max(u, 0)$, n denotes the number of spatial grid cells $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$, and D the number of days in the season. The derivation of this expression is provided in the Supplement S1.3. Given this distribution function we can construct our process $R(\mathbf{s}, d)$ as in Eq. (22).

3.3.4 Estimation

For a given region $\mathcal{R}$, we let the parameter that controls the space-time dependence be denoted by

$$\Theta_{\mathcal{R}}^{\mathrm{DEP}} := (\delta, \theta, \omega, \eta, \kappa, \lambda, \rho_s, \rho_t)^{\top}. \tag{31}$$

The estimation of $\Theta_{\mathcal{R}}^{\mathrm{DEP}}$ is complex: we were unable to establish the likelihood function, and even if it had been possible, the maximization problem would likely have been intractable. Instead, we used likelihood-free parameter estimation with the neural Bayes estimator (Sainsbury-Dale et al., 2024). This estimation framework has been used by Dell'Oro and Gaetan (2024), Richards et al. (2024), Sainsbury-Dale et al. (2024), and Vihrs (2022) to estimate parameters of spatial processes when conventional methods, such as maximum likelihood estimation or pseudo-likelihood techniques, are unfeasible. The concept is straightforward and involves treating the estimation issue as a prediction challenge. The only prerequisite for the approach to work is the ability to simulate from the data-generating process to produce simulated data for model training and validation.

The method consists of establishing a prior distribution for the parameters and then conducting extensive random sampling over the parameter space. Following that, a dataset is created from the process for each parameter setting. A Neural Network (NN) is then trained to learn the map between the data to the parameters that have generated the data. Training the NN is a costly task; but, once completed, obtaining estimates from a dataset may be executed rapidly. This is the reason why Zammit-Mangion et al. (2025) designate it as an amortized estimator. In contrast to the method of Sainsbury-Dale et al. (2024), we employed the approach proposed by Dell'Oro and Gaetan (2024) and Vihrs (2022), so we did not input the data directly into the NN but rather utilized certain statistics. Selecting which statistics are appropriate instead of letting the NN learning them brings some subjectivity to the process. However, this can massively reduce the memory needed to train the neural network.

To train the NN to estimate the parameters of the underlying spatiotemporal process, we first transformed the value of $Z(\mathbf{s}, d)$ to the uniform scale $U(\mathbf{s}, d)$ by using Eq. (21), and then we computed several summary statistics that characterize extremal dependence. These include spatial extremogram grids at multiple lags (Dell'Oro & Gaetan, 2024), an auto-tail dependence function (ATDF) (Reiss & Thomas, 2007, Chapter 13), distance-binned spatial extremograms (Davis & Mikosch, 2009) summarized by quantile profiles and interquartile ranges (IQR) and a madogram map (Cooley et al., 2006). In practice, the observed data consist of T independent replicates of the spatio-temporal process $Z(\mathbf{s}, d, t)$ $t \in \{1, \ldots, T\}$, corresponding to distinct summer seasons (years).

All summary statistics defined in this section are computed on the copula scale $U(\mathbf{s}, d, t)$. To train the neural Bayes estimator, we can work directly with $U(\mathbf{s}, d, t)$ because the marginal distribution of the latent process $Z(\mathbf{s}, d, t)$ is known. Because we simulate $Z(\mathbf{s}, d, t)$ explicitly, the transformation in Eq. (21) is available in closed form. In contrast, for the observed temperature anomalies $X(\mathbf{s}, d, t)$ the marginal distribution varies with year through the covariates, and $U(\mathbf{s}, d, t)$ must be constructed in a way that removes this temporal nonstationarity. We achieve this by mapping $X(\mathbf{s}, d, t)$ to the copula scale using the estimated conditional quantile function from the BSQR model (Section 3.1): for each $X(\mathbf{s}, d, t)$ we define

$$\hat{U}(\mathbf{s}, d, t) = \hat{q}^{-1}\big(X(\mathbf{s}, d, t) \mid \mathbf{c}(t), \mathbf{s}\big), \tag{32}$$

where $\hat{q}(\tau \mid \mathbf{c}(t), \mathbf{s})$ denotes the fitted conditional quantile function of Section 3.1. Equivalently, $\hat{U}(\mathbf{s}, d, t)$ is the estimated conditional CDF value of $X(\mathbf{s}, d, t)$ given $\mathbf{c}(t)$, so that the transformed field has approximately uniform margins and can be compared across years. All dependence summaries (the χ-grids, ATDF, extremogram quantile profiles, IQR curves, and the thresholded madogram map) are then computed by replacing $U(\mathbf{s}, d, t)$ with $\hat{U}(\mathbf{s}, d, t)$.

For the χ-grids, we evaluate the pairwise tail dependence coefficient $\chi_{\mathbf{s}_i,\mathbf{s}_k}(h; u)$ defined as:

$$\chi_{\mathbf{s}_i,\mathbf{s}_k}(h; u) = \mathsf{P}\Big\{U(\mathbf{s}_k, d + h, t) > u \;\Big|\; U(\mathbf{s}_i, d, t) > u\Big\}. \tag{33}$$

which measures the probability of an extreme event at location $\mathbf{s}_k$ on day $d + h$ given an extreme event at location $\mathbf{s}_i$ on day d. We evaluate this quantity empirically over probability thresholds $u \in \{0.90, 0.95, 0.99\}$ with the following estimator. Define the binary indicator of exceedance:

$$I_{\mathbf{s},u}(t, d) = \mathbb{1}\{U(\mathbf{s}, d, t) > u\}.$$

Then, for locations $\mathbf{s}_i$ and $\mathbf{s}_k$ and lag $h \geq 0$. we estimate the tail dependence coefficient via:

$$\hat{\chi}_{\mathbf{s}_i,\mathbf{s}_k}(h; u) = \frac{\displaystyle\sum_{t=1}^{T}\sum_{d=1}^{D-h} I_{\mathbf{s}_i,u}(t, d)\, I_{\mathbf{s}_k,u}(t, d + h)}{\displaystyle\sum_{t=1}^{T}\sum_{d=1}^{D-h} I_{\mathbf{s}_i,u}(t, d)}. \tag{34}$$

To reduce dimensionality, we aggregate pairwise estimates according to the Euclidean distance between locations. Let $\|\mathbf{s}_i - \mathbf{s}_k\|$ denote the distance between locations $\mathbf{s}_i$ and $\mathbf{s}_k$, and define distance bins

$$B_l = [r_l, r_l + 100\,\text{km}), \qquad l = 1, \ldots, L,$$

where $r_l = 100(l-1)$ km. For a given lag h and threshold u, the bin-averaged tail dependence coefficient is defined as

$$\bar{\chi}(r_l, h; u) = \frac{1}{|B_l|} \sum_{\substack{i<k: \\ \|\mathbf{s}_i - \mathbf{s}_k\| \in B_m}} \hat{\chi}_{\mathbf{s}_i,\mathbf{s}_k}(h; u), \tag{35}$$

where $|B_l|$ denotes the number of location pairs whose distance falls in bin B_l. Repeating this process for lags $h = 0, \ldots, 10$ yields a two-dimensional matrix (χ-grid) representing extremal dependence across space and time. For numerical stability, we apply a logarithmic transformation to the averaged estimates, computing $\log\left(\bar{\chi}(r_l, h; u) + \varepsilon\right)$, with $\varepsilon = 10^{-6}$. The use of such grids has also been proposed by Dell'Oro and Gaetan (2024) in a similar setting. An illustration of the estimated χ-grids are visualized in Supplement S1.4 Figure S3.

To provide more information about the temporal persistence of extremes, we compute the auto-tail dependence function (ATDF) for lags $h = 0, \ldots, 10$. For each spatial location $\mathbf{s}$, the ATDF at lag h is estimated as

$$\widehat{\text{ATDF}}_{\mathbf{s}}(h) = \frac{\displaystyle\sum_{t=1}^{T} \sum_{d=1}^{D-h} I_{\mathbf{s},u}(t, d)\, I_{\mathbf{s},u}(t, d+h)}{\displaystyle\sum_{t=1}^{T} \sum_{d=1}^{D-h} I_{\mathbf{s},u}(t, d)}, \tag{36}$$

which estimates the conditional probability $\mathsf{P}\{U(\mathbf{s}, d+h, t) > u \mid U(\mathbf{s}, d, t) > u\}$. The final ATDF curve is obtained by averaging across all spatial locations $\mathbf{s}$ in the region. We have

$$\widehat{\text{ATDF}}_{\mathcal{R}}(h) = \frac{1}{n_{\mathcal{R}}} \sum_{i=1}^{n_{\mathcal{R}}} \widehat{\text{ATDF}}_{\mathbf{s}_i}(h), \tag{37}$$

where $n_{\mathcal{R}}$ denotes the number of grid cells in region $\mathcal{R}$. For numerical stability we apply a logarithmic transformation, computing $\log\left(\widehat{\text{ATDF}}_{\mathcal{R}}(h) + \varepsilon\right)$.

We further characterize the spatial heterogeneity of extremal dependence by computing a quantile profile of the spatial extremogram at lag $h = 0$. For a fixed exceedance threshold $u = 0.95$, we first estimate the pairwise tail dependence coefficients $\hat{\chi}_{\mathbf{s}_i,\mathbf{s}_k}(0; u)$ using Eq. (34) for all pairs of locations

$(\mathbf{s}_i, \mathbf{s}_k)$ in the region $\mathcal{R}$. These estimates are then grouped into distance bins of fixed width (100 km) according to the spatial separation $\|\mathbf{s}_i - \mathbf{s}_k\|$. Let B denote the total number of distance bins, and define

$$C_b = \Big\{\hat{\chi}_{\mathbf{s}_i,\mathbf{s}_k}(0;u) : \|\mathbf{s}_i - \mathbf{s}_k\| \in \text{bin } b\Big\}, \qquad b = 1, \dots, B.$$

Within each bin b, we compute the empirical α-quantiles of C_b, namely

$$\hat{q}_\alpha(b) = \inf\Big\{x : \frac{1}{|C_b|}\sum_{c\in C_b} \mathbb{1}(c \le x) \ge \alpha\Big\}, \qquad \alpha \in \{0.1,\, 0.5,\, 0.9\}. \tag{38}$$

This yields a $3 \times B$ summary describing the median decay of extremal dependence with distance together with its lower and upper variability envelopes. For numerical stability, we apply a logarithmic transformation to the bin-wise quantiles,

$$\log\big(\hat{q}_\alpha(b) + \varepsilon\big).$$

As a compact summary of dispersion, we additionally report the interquartile range (IQR),

$$\widehat{\mathrm{IQR}}(b) = \hat{q}_{0.75}(b) - \hat{q}_{0.25}(b), \tag{39}$$

which quantifies the variability of extremal dependence across spatial pairs at comparable separations. For numerical stability, a small constant ε is added prior to applying a logarithmic transformation to all binwise summaries $\log\left(\widehat{\mathrm{IQR}}(b) + \varepsilon\right)$.

To characterize directional features and spatial anisotropy in the dependence structure of extremes, we compute a thresholded madogram map based on the field $U(\mathbf{s}, d)$. Let $u = 0.95$ denote a high probability threshold so that only high values contribute to the dependence measure. For any pair of spatial locations $\mathbf{s}_i = (s_{i,x}, s_{i,y})$ and $\mathbf{s}_k = (\mathbf{s}_{k,x}, \mathbf{s}_{k,y})$, consider the displacement vector $\boldsymbol{h} = \mathbf{s}_k - \mathbf{s}_i = (\mathbf{s}_{k,x} - \mathbf{s}_{i,x}, \mathbf{s}_{k,y} - \mathbf{s}_{i,y}) = (h_x, h_y)$. For a spatial lag vector $\boldsymbol{h} = (h_x, h_y)$, the empirical extremal thresholded madogram is defined as

$$\hat{\nu}(\boldsymbol{h}) = \frac{1}{TD}\sum_{t=1}^{T}\sum_{d=1}^{D}\Bigg\{\frac{1}{2\,N(\boldsymbol{h}, d, t)} \sum_{i,k:\mathbf{s}_k - \mathbf{s}_i \in \mathcal{B}(\boldsymbol{h})} \big|U(\mathbf{s}_k, d, t) - U(\mathbf{s}_i, d, t)\big|\; \mathbb{1}\big\{U(\mathbf{s}_k, d, t) \ge u, U(\mathbf{s}_i, d, t) \ge u\big\}\Bigg\}, \tag{40}$$

where $\mathcal{B}(\boldsymbol{h})$ denotes a bin of displacement vectors centered at $\boldsymbol{h}$, and $N(\boldsymbol{h}, d, t)$ is the number of spatial pairs such that $\mathbf{s}_k - \mathbf{s}_i \in \mathcal{B}(\boldsymbol{h})$ and $U(\mathbf{s}_k, d, t) \ge u$ and $U(\mathbf{s}_i, d, t) \ge u$. The lag-vector bins

$\mathcal{B}(\boldsymbol{h})$ are defined on a regular two-dimensional grid in (h_x, h_y) with bin width 100 km in both directions and maximum extent 1000 km. A pair $(\mathbf{s}_i, \mathbf{s}_k)$ contributes to the bin centered at $\boldsymbol{h}$ if both components of $\mathbf{s}_k - \mathbf{s}_i$ fall within ±50 km of (h_x, h_y). This construction is common in geostatistics (Chiles & Delfiner, 2012, Chapter 2) and know as the variogram map; we only adjusted it to a thresholded madogram to be coherent with EVT. The resulting madogram map summarizes the average absolute difference between high values of U at pairs of locations separated by a given spatial displacement. Isotropic dependence would produce approximately symmetric patterns in this map, while elongated or directional features indicate spatial anisotropy in the structure of extremes. For stability purposes, the madogram values are log-transformed $\log(\hat{\nu}(\boldsymbol{h}) + \varepsilon)$.

3.4 Return periods and uncertainty quantification

In extreme value analysis, the return period N and the return level z_N are key metrics for quantifying the rarity of extreme events. The return period N represents the average time interval between events whose magnitude is at least z_N. Mathematically, if M denotes the annual maximum of the variable of interest (e.g., temperature anomalies), the return period associated with a level z_N is $T(z_N) = 1/\mathsf{P}(M > z_N)$. The return level z_N is defined as the magnitude that is expected to be exceeded on average once every N years, i.e., it satisfies $\mathsf{P}(M > z_N) = 1/N$. In the simple univariate case, one can fit a GEV distribution to the annual maxima. The estimated return level associated with a return period of N years is given by $z_N = G^{-1}(1 - 1/N)$ with $G(\cdot)$ defined as Eq. (11). However, when the analytical form of G is not available, for instance, when using complex stochastic models, the relationship between N and z_N can be obtained through numerical approaches such as Monte Carlo simulation.

To return simulated values from the latent process $Z(\mathbf{s}, d, t)$ to the original physical scale for a given year t, we apply a marginal back-transformation based on the nonstationary fitted GEV parameters of annual maxima at each location $\mathbf{s}$. Full derivations of this transformation are provided in Supplement S1.5. Specifically, we first transform the realizations of $Z(\mathbf{s}, d, t)$ to a uniform scale using Eq. (21), and then back-transform the uniform variates $U(\mathbf{s}, d, t)$ to the original scale using

$$X(\mathbf{s}, d, t) = \begin{cases} q(U(\mathbf{s}, d, t) \mid \mathbf{c}(t), \mathbf{s}), & U(\mathbf{s}, d, t) < \tau_{\mathbf{s},t}, \\ \vartheta_{\mathbf{s}}^{\text{thres}} + \dfrac{\tilde{\sigma}_{\mathbf{s},t}}{\xi_{\mathbf{s}}}\left[\left(\dfrac{1 - U(\mathbf{s}, d, t)}{1 - \tau_{\mathbf{s},t}}\right)^{-\xi_{\mathbf{s}}} - 1\right], & U(\mathbf{s}, d, t) \geq \tau_{\mathbf{s},t},\ \xi_{\mathbf{s}} \neq 0, \\ \vartheta_{\mathbf{s}}^{\text{thres}} - \tilde{\sigma}_{\mathbf{s},t} \log\left(\dfrac{1 - U(\mathbf{s}, d, t)}{1 - \tau_{\mathbf{s},t}}\right), & U(\mathbf{s}, d, t) \geq \tau_{\mathbf{s},t},\ \xi_{\mathbf{s}} = 0. \end{cases} \tag{41}$$

Here, $q(\cdot \mid \mathbf{c}(t), \mathbf{s})$ is the BSQR model described in Section 3.1 and $\tau_{\mathbf{s},t} = q^{-1}(\vartheta_{\mathbf{s}}^{\text{thres}} \mid \mathbf{c}(t), \mathbf{s})$ is obtained by inverting it. The quantity $\tilde{\sigma}_{\mathbf{s},t} = \sigma_{\mathbf{s}}(t) + \xi_{\mathbf{s}}\{\vartheta_{\mathbf{s}}^{\text{thres}} - \mu_{\mathbf{s}}(t)\}$ involves the nonstationary GEV parameters $\mu_{\mathbf{s}}(t)$, $\sigma_{\mathbf{s}}(t)$, and $\xi_{\mathbf{s}}$ defined in Eq. (12), while the threshold $\vartheta_{\mathbf{s}}^{\text{thres}}$ is given in Eq. (3).

Now that we can simulate from our random field, we can generate independent replicates and use a Monte Carlo technique to approximate return periods. For each replicate j, we are interested in a certain property $\mathcal{P}$ that summarizes aspects of the simulated field over the year e.g., the duration of an event, the frequency of occurrence, or the intensity. Given a property $\mathcal{P}$, we fix a return level $z_N^{(\mathcal{P})}$ and approximate the corresponding return period empirically from the simulated data. Specifically, we generate a large number J of replicates. Let $X_j^{(\mathcal{P})}$ denote the value of property $\mathcal{P}$ in replicate j. The empirical approximation of the return period associated with the threshold $z_N^{(\mathcal{P})}$ is then given by

$$\hat{T}_J^{(\mathcal{P})}(z_N^{(\mathcal{P})}) \;=\; \frac{J}{\sum_{j=1}^{J} \mathbb{1}\{X_j^{(\mathcal{P})} > z_N^{(\mathcal{P})}\}}. \tag{42}$$

The relative error of our Monte Carlo approximation is given by $1/\sqrt{Jp}$, where p is the probability of exceedance that we are trying to approximate.

To quantify the uncertainty in the return period $\hat{T}_J^{(\mathcal{P})}(z_N^{(\mathcal{P})})$ that arises from the estimation procedure, we propose to use the following approach. For a given region $\mathcal{R}$ let the parameter set be denoted by $\Theta_{\mathcal{R}} = \left(\Theta_{\mathcal{R}}^{\text{QR}}, \Theta_{\mathcal{R}}^{\text{GEV}}, \Theta_{\mathcal{R}}^{\text{DEP}}\right)^{\top}$, where $\Theta_{\mathcal{R}}^{\text{QR}}$ are the parameters of the BSQR as defined in Eq. (S18) in Supplement S1.1, $\Theta_{\mathcal{R}}^{\text{GEV}}$ are the nonstationary GEV parameters as in Eq. (18), and $\Theta_{\mathcal{R}}^{\text{DEP}}$ are the space-time dependence parameters of the process $Z(\mathbf{s}, t)$ as defined in Eq. (31). As a full bootstrap strategy is computationally infeasible, we adopted a twofold resampling strategy: (i) a parametric bootstrap for the space-time dependence parameters as suggested by Dell'Oro and Gaetan (2024), Sainsbury-Dale et al. (2025), and Zammit-Mangion et al. (2025), (ii) a posterior sampling for the marginal GEV parameters from the posterior distribution as described in Section 3.2, Eq. (19), and a posterior sampling for the BSQR coefficient as described in Section 3.1, Eq. (10). From these procedures, we can generate B triples of parameters

$$\Theta_{\mathcal{R}}^{(b)} = (\Theta_{\mathcal{R}}^{\mathrm{QR},(\mathrm{b})}, \Theta_{\mathcal{R}}^{\mathrm{GEV},(\mathrm{b})}, \Theta_{\mathcal{R}}^{\mathrm{DEP},(\mathrm{b})})^\top, \qquad b = 1, \ldots, B. \tag{43}$$

For each parameter replicate b, we simulate J' data replicates (J' possibly smaller than J) from the model; we then approximate the return period $\hat{T}_{N,b}^{(P)}(z_N^{(\mathcal{P})})$ via the previous Monte Carlo method. The empirical distribution of these estimates is then used to assess uncertainty, with pointwise $1 - \alpha$ confidence intervals derived from the empirical quantiles.

4 Data application

4.1 Data acquisition and preprocessing

The data application is based on climate model simulations from the Coupled Model Intercomparison Project Phase 6 (CMIP6). As described in Section 2, we focus on simulations produced by the Meteorological Research Institute Earth System Model version 2 (MRI–ESM2).

For both the factual and counterfactual climates, we use daily maximum near-surface air temperature (TASMAX). In all experiments, we rely on the `r1i1p1f1` ensemble member to ensure consistency across scenarios. The factual climate is represented by the historical experiment extended with the SSP2–4.5 scenario beyond 2014, while the counterfactual climate is represented by the historical natural-forcing-only (`hist-nat`) experiment, as detailed in Section 2.1. All TASMAX fields are processed according to the methodology described in Section 2.

To capture large-scale climatic forcing, we use GMST anomalies as a covariate in the marginal models. GMST is computed from monthly near-surface air temperature (TAS) fields over all 12 months of each year, provided by MRI–ESM2, again using the `r1i1p1f1` ensemble member for both factual and counterfactual simulations. Monthly TAS fields are aggregated to a global mean. The resulting global mean series are then averaged to yearly means. Yearly GMST anomalies are computed relative to the same early-industrial baseline period (1850–1949) used for the TASMAX anomalies, separately for the factual and counterfactual climates. For the factual climate, the GMST series is truncated in 2020 to match the temporal availability of the counterfactual simulation. The resulting GMST anomaly series serves as the sole covariate in the BSQR model and the Bayesian spatial GEV model introduced in Sections 3.1 and 3.2 , where it drives nonstationary changes in the

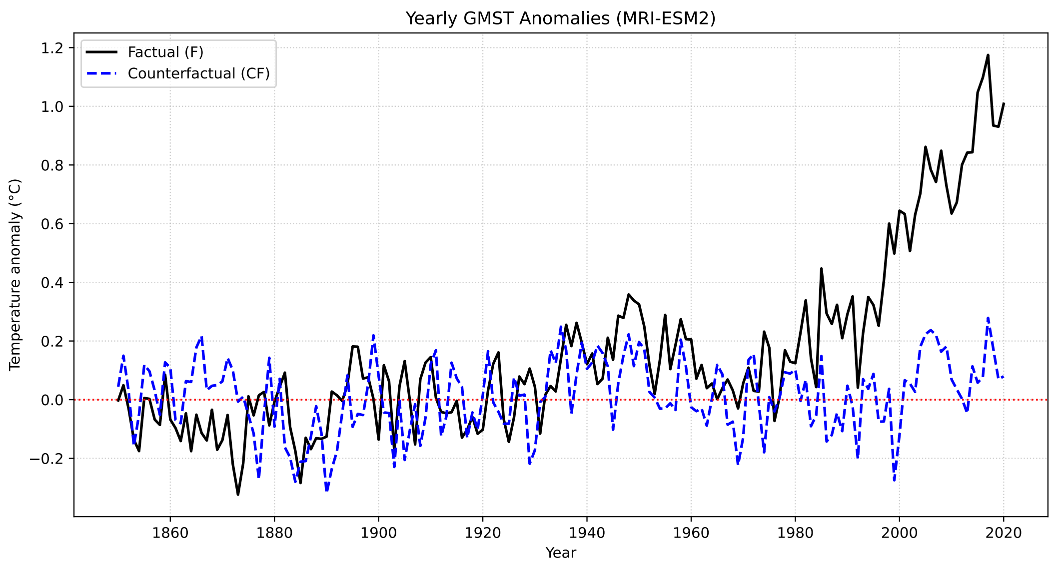


Figure 1: Yearly GMST anomalies from MRI–ESM2, computed relative to the 1850–1949 baseline. The factual series includes anthropogenic and natural forcings, while the counterfactual series includes natural forcings only. Both series are based on the `r1i1p1f1` ensemble member.

marginal distribution of daily temperature anomalies. The temporal evolution of this covariate in the factual and counterfactual climates is shown in Figure 1.

Europe spans a wide range of climate regimes, and fitting a single spatio-temporal model over the entire domain is unlikely to be realistic. In particular, both the marginal behavior and the dependence structure of extreme temperatures can vary substantially across regions (e.g., Mediterranean versus Scandinavian climates). This motivates a regionalization step prior to model fitting, presented in Supplement S2.1. We focus on Cluster 3, as shown in Figure 2.

4.2 BSQR application

We apply the BSQR methodology described in Section 3.1 to estimate spatially smooth conditional quantile functions of daily summer temperature anomalies. In the BSQR model, the conditional quantile function at level $\tau \in (0, 1)$ is expressed as a Bernstein-polynomial expansion; in the application we set the number of basis functions to $M = 15$ following Reich et al. (2011). The covariate entering the model is yearly GMST anomalies. To match the BSQR formulation, we rescale this covariate to the unit interval by a min–max transformation. Specifically, letting g_t denote the

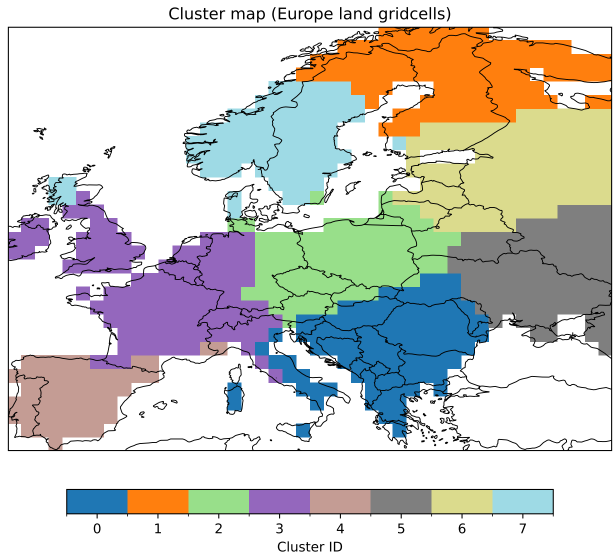


Figure 2: Spatial clustering of Europe based on extremal dependence of summer (JJA) temperature anomalies in the counterfactual climate. The partition was obtained using a k-medoids algorithm with the chi-based dissimilarity in Eq. (S21) computed at threshold $u = 0.8$, yielding $k = 8$ clusters.

yearly GMST anomaly at year t, we define

$$\tilde{g}_t = \frac{g_t - g_{\min}}{g_{\max} - g_{\min}}, \qquad g_{\min} = \min\{g_t^{\mathrm{F}}, g_t^{\mathrm{CF}}\}, \quad t \in \{1, \ldots, T\}, \quad g_{\max} = 4°\mathrm{C},$$

where g_t, $g_{\min}$ are expressed in °C. We choose $g_{\max} = 4°$C as this is the maximum value we allow for extrapolation, so that $\tilde{g}_t \in [0, 1]$ for the range of values encountered in the present application.

A key role of the fitted BSQR model is to remove temporal nonstationarity in the margins prior to dependence modeling. Using the estimated conditional quantile function $\hat{q}_{\mathbf{s},t}(\tau)$, we transform daily anomalies in both the factual and counterfactual worlds to the copula scale via the estimated conditional distribution function,

$$\widehat{U}(\mathbf{s}, d, t) = \hat{F}_{\mathbf{s},t}\big(X(\mathbf{s}, d, t)\big) = \hat{q}_{\mathbf{s},t}^{-1}\big(X(\mathbf{s}, d, t)\big).$$

A diagnostic check of this uniformization step for Cluster 3 is provided Supplement S2.2 Figure S4. The empirical distributions are close to the uniform distribution on $[0, 1]$, supporting the adequacy of the marginal standardization. Deviations are nevertheless visible in the lower and upper tails, indicating that the BSQR model captures the central part of the conditional distribution more accurately than the extremes.

Another key contribution of the BSQR model is its ability to quantify how a fixed counterfactual heatwave threshold shifts in the factual distribution of the temperature distribution under global warming. Specifically, we consider the counterfactual threshold $\hat{\vartheta}_{\mathbf{s}}^{\text{thres}}$ defined in Eq. (4) and examine how its associated probability level

$$\tau_{\mathbf{s},t} = q^{-1}\left(\vartheta_{\mathbf{s}}^{\text{thres}} \mid \mathbf{c}(t), \mathbf{s}\right)$$

evolves as a function of GMST. This provides a characterization of how temperatures that were extreme in the counterfactual climate are repositioned within the temperature distribution at each grid cell under factual warming. The results are presented in Figure 3. Panels (a)–(d) present the exceedance probability change

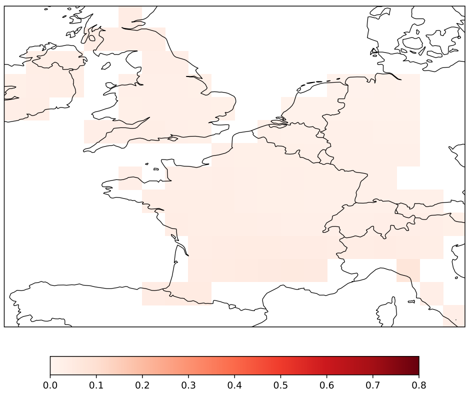
$$\left[1 - q_{\mathrm{F}}^{-1}(\hat{\vartheta}_{\mathbf{s}}^{\text{thres}} \mid \text{GMST} = g, \mathbf{s})\right] - \left[1 - q_{\mathrm{CF}}^{-1}(\hat{\vartheta}_{\mathbf{s}}^{\text{thres}} \mid \text{GMST} = 0, \mathbf{s})\right]$$

conditional on GMST anomalies g of 1 °C up to 4 °C.

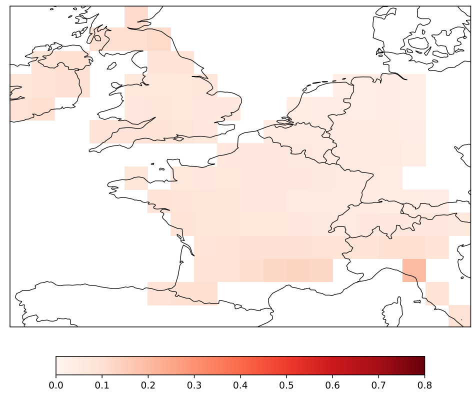
4.3 GEV application

We apply the spatially varying nonstationary GEV model in Section 3.2 to annual summer maxima of daily temperature anomalies over Cluster 3. Annual maxima are extracted independently for each grid cell and year, and the GEV parameters are modeled as functions of the rescaled GMST covariate, with spatial structure introduced through intrinsic Gaussian Markov random field priors. The model is fitted separately for the counterfactual and factual climates using Bayesian inference.

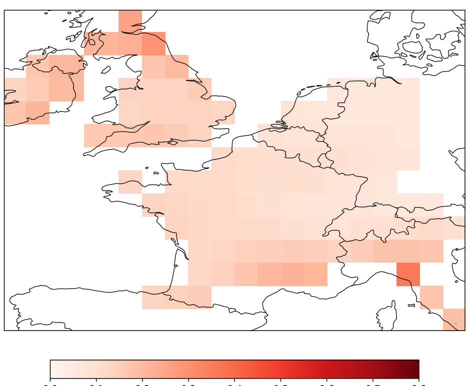
Figure 4 displays posterior mean estimates of the GEV parameters that are most directly relevant for attribution over Cluster 3, for factual climate. Specifically, we report the spatial patterns of the covariate effects in the location and scale parameters, μ_1 and σ_1, which quantify the sensitivity of extreme temperatures to GMST anomalies. We find that as GMST anomalies increase, extremes tend to become more intense, as indicated by the positive values of μ_1. On the other hand, the variability appears to decrease with increasing GMST, as suggested by σ_1. This behavior is also reported by Auld et al. (2023), who found that the effect of global warming on the scale parameter is heterogeneous across clusters. However, for clusters similar to ours, our results are consistent with their findings. Caterpillar plots by grid cell are shown in Figure S7, Figure S8 and Figure S9 in the Supplement S2.5, where the uncertainty around the parameters is also illustrated. See Supplement S2.3 (Figure S5) and Supplement S2.4 (Figure S6) for MCMC diagnostics and posterior predictive checks.



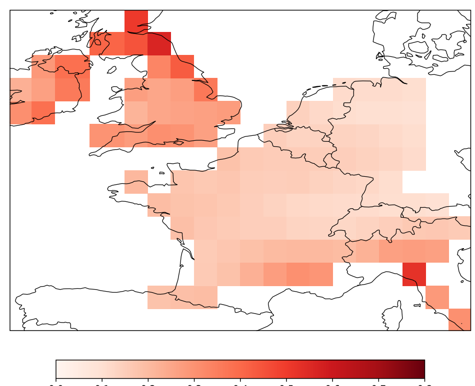
(a) Exceedance probability change for GMST = 1°C



(b) Same as (a), but for GMST = 2°C.

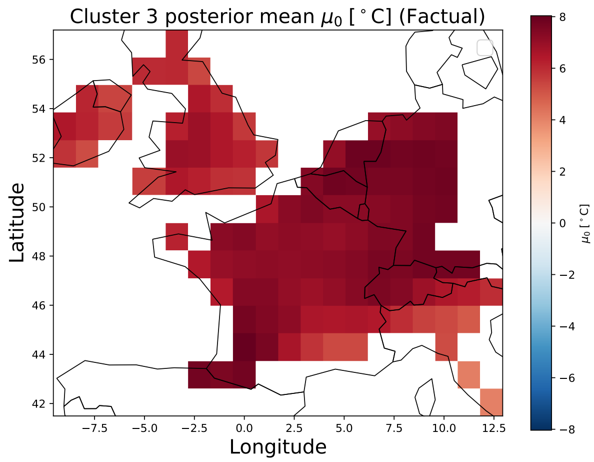


(c) Same as (a), but for GMST = 3°C.

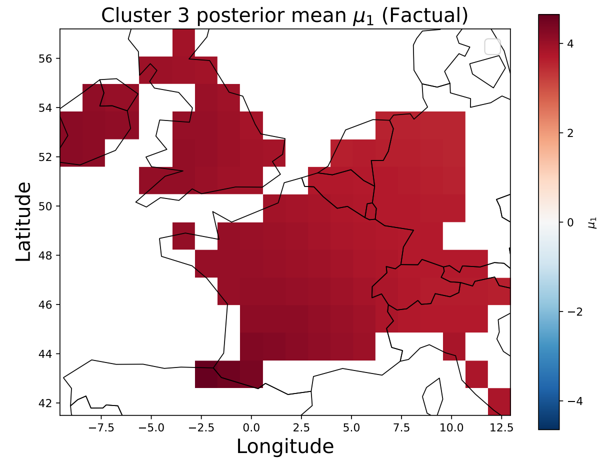


(d) Same as (a), but for GMST = 4°C.

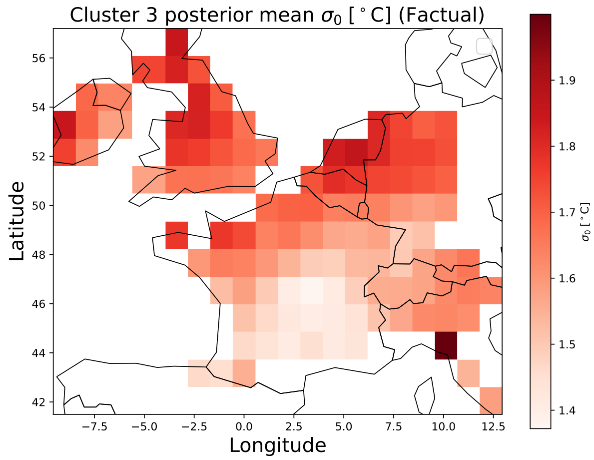
Figure 3: (a)–(d) Change in exceedance probability conditional on GMST anomalies.



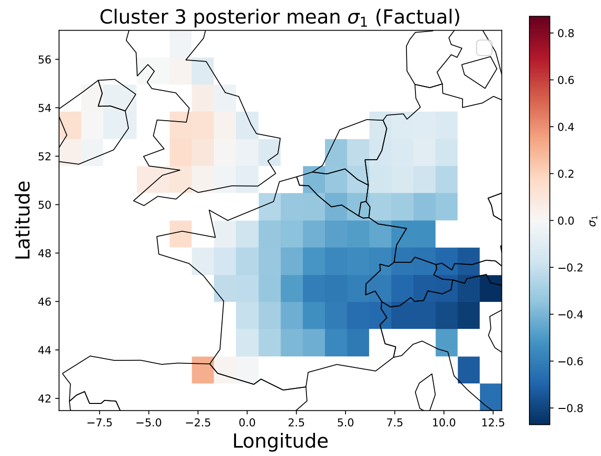
(a) μ_0 [°C]



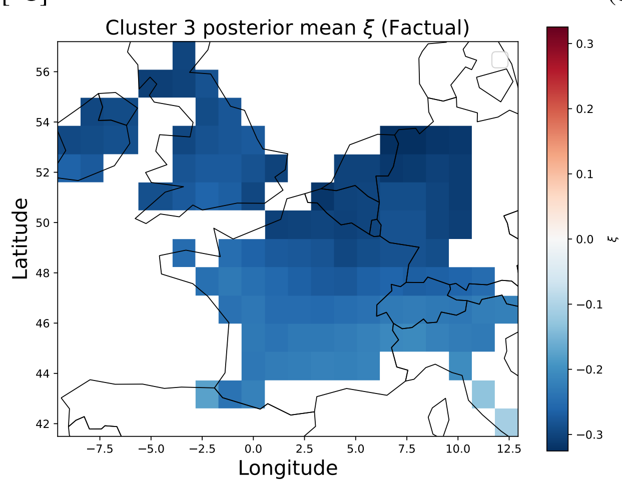
(b) μ_1



(c) σ_0 [°C]



(d) σ_1



(e) ξ

Figure 4: Posterior mean maps of key GEV parameters over Cluster 3 under the factual climate.

4.4 Neural Bayes estimator for Cluster 3

For Cluster 3, we trained a neural Bayes estimator on a large number of simulated datasets generated from the proposed spatio-temporal model. Each training sample consisted of the collection of summary statistics described above, namely the χ-grids, the thresholded madogram map, the spatial extremogram spread summaries, the ATDF curve in Eq. (37), and the IQR-by-distance curve in Eq. (39). The corresponding target output was the dependence parameter vector

$$\Theta_{\mathcal{R}}^{\mathrm{DEP}} = (\delta, \theta, \omega, \eta, \kappa, \lambda, \rho_s, \rho_t)^\top.$$

For the simulation of the training datasets, we specified independent prior distributions for each dependence parameter component. We adopted uniform priors over ranges chosen to cover plausible values of the model parameters while ensuring sufficient variability in the simulated data. Specifically, we used

$$\begin{aligned} \delta\,[-] &\sim \mathcal{U}(0,1), & \theta[\mathrm{km}] &\sim \mathcal{U}(50,1000), \\ \kappa[\mathrm{d}] &\sim \mathcal{U}(1,30), & \eta\,[-] &\sim \mathcal{U}(1,5), \\ \lambda\,[-] &\sim \mathcal{U}(0,5), & \omega\,[\mathrm{rad}] &\sim \mathcal{U}(0,\pi), \\ \rho_s[\mathrm{km}] &\sim \mathcal{U}(10,400), & \rho_t[\mathrm{d}] &\sim \mathcal{U}(0,15). \end{aligned}$$

with all parameters sampled independently. These ranges were selected to span a broad yet realistic domain given the cluster size for the spatio-temporal dependence structure, allowing the neural Bayes estimator to learn across diverse regimes of extremal dependence.

The NN was implemented as a multi-input architecture combining convolutional branches for image-like summaries and one-dimensional convolutional branches for curve-like summaries. More precisely, the χ-grids, madogram map, and spatial spread summaries were processed through two-dimensional convolutional layers, whereas the ATDF and IQR summaries were processed through one-dimensional convolutional layers. The features extracted from all branches were then concatenated and passed through fully connected layers to predict the eight model parameters jointly. The network was trained on 299 000 simulated samples of 171 years, while 1000 samples were used for validation during training. Furthermore, 1000 simulated samples were kept as an independent holdout set to assess predictive accuracy.

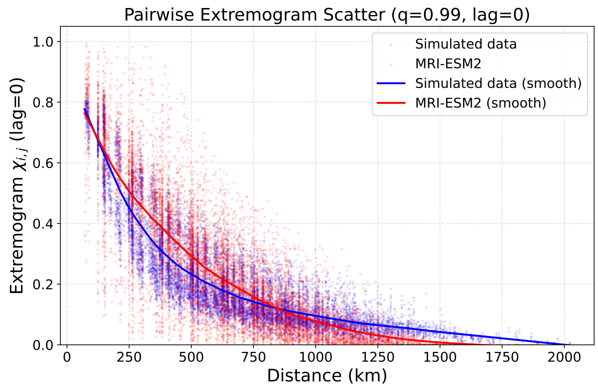
To evaluate the estimator, we compared the predicted and true parameter values on the holdout set. Figure S10 in Supplement S2.6 displays scatterplots of predicted versus true values for all eight parameters. Points concentrated around the diagonal indicate accurate recovery of the parameter. Figure S11 in Supplement S2.6 complements this analysis by showing boxplots of the prediction errors, defined as predicted minus true value, for each parameter. We observe that the estimation accuracy depends on the value of δ. Recall $Z(\mathbf{s}, d) = R(\mathbf{s}, d)^{\delta} \cdot W(\mathbf{s}, d)^{1-\delta}$ in Eq. (20). When δ is small, the data are primarily driven by the process $W(\mathbf{s}, d)$, and consequently the parameters associated with the component $R(\mathbf{s}, d)$, namely λ, ρ_s, and ρ_t, are not well identified. Conversely, when δ is close to 1, the data are mainly governed by the process $R(\mathbf{s}, d)$, and the parameters related to $W(\mathbf{s}, d)$ become difficult to estimate. In addition, the value of η in Eq. (25) also impacts identifiability. When η is close to 1, the process exhibits little to no anisotropy, which makes the anisotropy angle ω poorly identifiable.

We report posterior estimates for all parameters in the factual world. Table 1 presents the true parameter values alongside the corresponding 95% credible intervals derived from the parametric bootstrap based on 1000 draws.

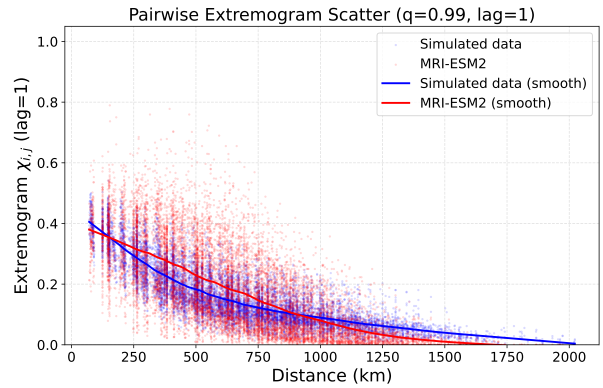
Table 1: Parameter estimates of dependence over Cluster 3 with 95% credible intervals for Factual world

Parameter	Estimated	95% CI
δ [−]	0.4494	(0.4155, 0.4691)
θ [km]	661.9112	(638.4598, 718.5402)
κ [d]	20.9955	(20.0814, 23.8779)
ω [rad]	2.3335	(2.2359, 2.3671)
η [−]	2.0366	(1.9770, 2.2302)
λ [−]	2.0668	(1.4143, 3.1116)
ρ_s [km]	300.9300	(261.5675, 352.3192)
ρ_t [d]	1.1806	(0.6109, 1.5983)

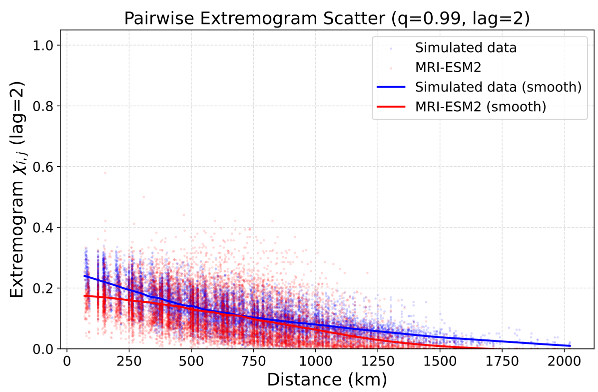
To assess whether our statistical model captures the correct extremal dependence structure, we simulate data from the statistical model and visually compare it with data generated by the MRI-ESM2 climate model. Figure 5 illustrates this comparison for a threshold of $u = 0.99$ and four lag times. We observe that our statistical model slightly underestimates extremal dependence at short distances. Moreover, although anisotropy is included, the climate model data exhibit greater variability than what is captured by our statistical model.



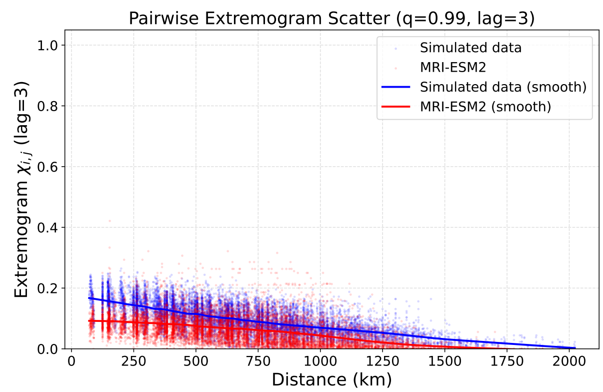
(a) Lag 0



(b) Lag 1



(c) Lag 2



(d) Lag 3

Figure 5: Pairwise extremogram plots for Cluster 3 at temporal lags 0 to 3 for 171 years of $Z(\mathbf{s}, d)$ vs the observed factual MRI-ESM2 data. Each panel shows the empirical extremogram as a function of spatial distance. The estimated dependence parameters are $\hat{\delta} = 0.45$, $\hat{\theta} = 661.98$, $\hat{\kappa} = 20.99$, $\hat{\omega} = 2.33$, $\hat{\eta} = 2.04$, $\hat{\lambda} = 2.07$, $\hat{\rho}_s = 301.01$, and $\hat{\rho}_t = 1.18$.

4.5 Data simulation, return period estimation & attribution

Given the parameters of the copula $Z(\mathbf{s}, d)$ and the parameters of the EVT and BSQR models, we can now use our simulation framework as described in Section 3.4. We first illustrate empirically why it is necessary to combine the BSQR model for the bulk of the data with the EVT model for the tail of the distribution at each location. Figure 6 shows a pooled upper-tail quantile-quantile (QQ) comparison between observed (MRI-ESM2) and simulated temperature anomalies. The empirical quantiles of the factual data from the climate model MRI-ESM2, $x^{\text{MRI-ESM2,F}}(\mathbf{s}, d, t)$ with $t \in \{1, \ldots, 171\}$, are compared to the corresponding quantiles of simulated fields obtained from the copula model. More precisely, we generate realizations $z(\mathbf{s}, d, t)$ of the latent process $Z(\mathbf{s}, d)$, which are transformed into uniform variables $u(\mathbf{s}, d, t)$ according to Eq. (21), using the parameters estimated by the Neural Bayes estimator. These uniform realizations are then backtransformed to the anomaly scale using two approaches. First, using only the BSQR model, we obtain $x^{\text{BSQR}}(\mathbf{s}, d, t) = \hat{q}\big(u(\mathbf{s}, d, t) \mid \mathbf{c}(t), \mathbf{s}\big)$, where $\hat{q}$ denotes the estimated conditional quantile function. Second, using the hybrid approach described in Eq. (41), we obtain $x^{\text{BSQR+EVT}}(\mathbf{s}, d, t)$, which combines the BSQR model for the bulk with a EVT model for the upper tail. The QQ plot focuses on high quantile levels and pools all locations and times in order to emphasize tail behavior. The BSQR only transformation clearly underestimates the upper quantiles, as shown by its systematic deviation below the 1:1 line. In contrast, the hybrid BSQR+EVT transformation corrects the underestimation of the upper tail and reaches the appropriate magnitude of extreme values. In the most extreme quantiles, it exceeds the 1:1 line, reflecting its ability to extrapolate beyond the range of the observed data.

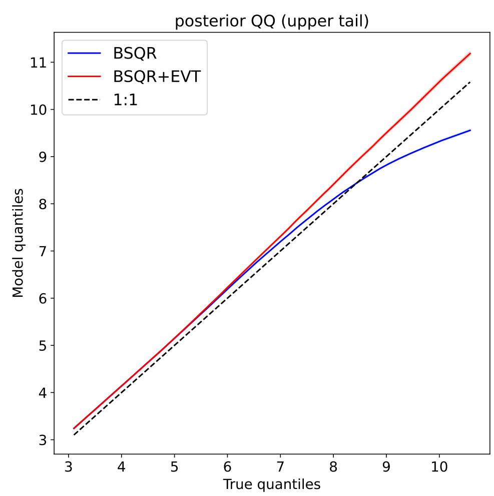
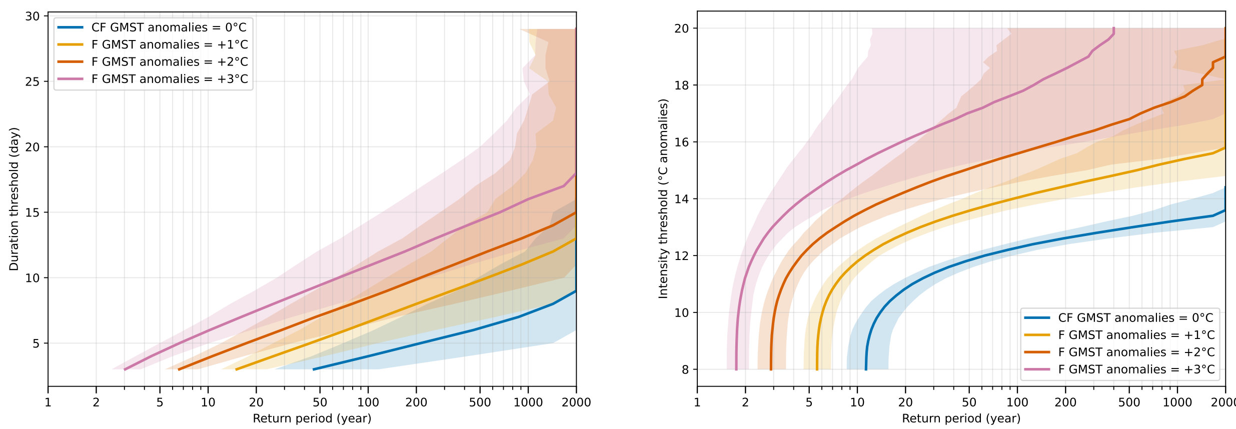


Figure 6: Pooled upper-tail QQ plot of observed MRI-ESM2 anomalies versus simulations from the fitted model. The BSQR-only transformation underestimates upper quantiles, whereas the hybrid BSQR+EVT approach is slightly above the 1:1 line at the highest quantiles. Which reflects its ability to extrapolate beyond the observed range.

Given the estimated model parameters, empirical return levels can now be computed using Monte Carlo simulation. For each parameter triple defined in Eq. (43), we simulated 10 000 years of data and applied the back-transformation described in Eq. (41). Figure 7 presents the return periods associated with the duration of heatwave episodes and with the maximum intensity observed during a heatwave episode under GMST anomaly scenarios of 0 °C up to 3 °C.



(a) Duration

(b) Intensity

Figure 7: Return period plots under different GMST anomaly scenarios. Panel (a) shows the return period plot for the duration of heatwave episodes, while panel (b) shows the return period plot for the maximum intensity observed during a heatwave episode. F and CF stands for factual and counterfactual climate.

5 Discussion & Conclusion

Our main objective was to develop a complete statistical pipeline to quantify the extent at which anthropogenic forcing has influenced the properties of heatwaves, including their duration and intensity. To address this question, we modeled temperature fields in both space and time within a nonstationary framework by combining marginal models with a flexible dependence model. The proposed approach integrates several modern statistical tools into a unified framework that allows the simulation of realistic spatio-temporal temperature fields. In particular, recent work has emphasized the importance of using sub-asymptotic models to accurately capture extremal dependence, as misspecification of the dependence structure can substantially affect attribution results (Huser et al., 2025; Li & Castro-Camilo, 2025). Although the estimation procedure is computationally demanding, once the model is fitted it enables practitioners to generate coherent spatio-temporal temperature fields and to investigate a wide range of heatwave characteristics.

Our results based on the MRI-ESM2 climate model are consistent with findings reported in the attribution literature (Engdaw et al., 2023; Intergovernmental Panel on Climate Change (IPCC), 2021; Quilcaille et al., 2025), namely that anthropogenic forcing has increased both the severity and the intensity of heatwaves. However, our approach differs from most previous studies in that we explicitly model heatwaves as spatio-temporal events rather than relying on scalar temperature indices or heatwave indicators. By simulating full temperature fields, the proposed framework makes it possible to study a broad set of heatwave properties, including their spatial extent, persistence, and intensity, within a coherent probabilistic framework.

Despite these promising results, several areas of improvement remain. First, the diagnostic analysis presented in Section 4.4 suggests that the proposed copula slightly underestimates the tail dependence. Over large spatial domains such as the one considered in this study, we observe substantial variability in the tail dependence that cannot be fully captured by a simple anisotropic structure. One possible improvement would be to introduce non-stationary covariance structures; however, such an extension would substantially increase the number of parameters and the computational complexity of the model. A second limitation concerns the assumption that the tail dependence regime—whether asymptotically dependent or independent—is constant across the entire study region. This assumption may be unrealistic in practice. More flexible models, such as those proposed by Hazra et al. (2025) or Shi et al. (2026), allow different dependence regimes to coexist within the same spatial domain, although these approaches are currently restricted to purely spatial settings. Finally, the dependence structure was assumed to be stationary over time. An interesting extension would be to allow the mixture parameter δ to vary with covariates, following ideas similar to those proposed by Maume-Deschamps et al. (2024).

Future work could therefore focus on improving the dependence model and extending the proposed framework to incorporate more flexible copula structures. Another natural direction would be to apply the methodology to ensembles of climate models in order to assess the robustness of the results and to evaluate whether consistent conclusions are obtained across different climate simulations. In addition, applying the framework to observational datasets would provide an important benchmark. The methodology could also be extended to other spatio-temporally coherent processes, such as droughts, and to different regions of the world.

References

Auld, G., Hegerl, G. C., & Papastathopoulos, I. (2023). Changes in the distribution of annual maximum temperatures in Europe. *Advances in Statistical Climatology, Meteorology and Oceanography*, *9*(1), 45–66.

Ballester, J., Quijal-Zamorano, M., Méndez Turrubiates, R. F., Pegenaute, F., Herrmann, F. R., Robine, J. M., Basagaña, X., Tonne, C., Antó, J. M., & Achebak, H. (2023). Heat-related mortality in Europe during the summer of 2022. *Nature Medicine*, *29*(7), 1857–1866.

Bernard, E., Naveau, P., Vrac, M., & Mestre, O. (2013). Clustering of maxima: Spatial dependencies among heavy rainfall in France. *Journal of Climate*, *26*(20), 7929–7937.

Bône, C., Gastineau, G., Thiria, S., Gallinari, P., & Mejia, C. (2023). Detection and attribution of climate change using a neural network. *Journal of Advances in Modeling Earth Systems*, *15*(10), e2022MS003475.

Chiles, J.-P., & Delfiner, P. (2012). *Geostatistics: Modeling spatial uncertainty*. John Wiley & Sons.

Cognot, C., Bel, L., Métivier, D., & Parey, S. (2025). A spatio-temporal weather generator for the temperature over France. *Advances in Statistical Climatology, Meteorology and Oceanography*, *11*(2), 203–228.

Coles, S. (2001). *An introduction to statistical modeling of extreme values* (Vol. 208). Springer.

Cooley, D., Naveau, P., & Poncet, P. (2006). Variograms for spatial max-stable random fields. In *Dependence in probability and statistics* (pp. 373–390). Springer.

Cooley, D., & Sain, S. R. (2010). Spatial hierarchical modeling of precipitation extremes from a regional climate model. *Journal of agricultural, biological, and environmental statistics*, *15*(3), 381–402.

Davis, R. A., & Mikosch, T. (2009). The extremogram: A correlogram for extreme events. *Bernoulli*, *15*(4), 977–1009.

Davison, A. C., Huser, R., & Thibaud, E. (2019). Spatial extremes. In *Handbook of environmental and ecological statistics* (pp. 711–744). Chapman; Hall/CRC.

Dell'Oro, L., & Gaetan, C. (2024). Flexible space-time models for extreme data. *arXiv preprint arXiv:2411.19184*.

Deng, M. (2008). An anisotropic model for spatial processes. *Geographical Analysis*, *40*(1), 26–51.

Dyrrdal, A. V., Lenkoski, A., Thorarinsdottir, T. L., & Stordal, F. (2015). Bayesian hierarchical modeling of extreme hourly precipitation in norway. *Environmetrics*, *26*(2), 89–106.

Engdaw, M. M., Steiner, A. K., Hegerl, G. C., & Ballinger, A. P. (2023). Attribution of observed changes in extreme temperatures to anthropogenic forcing using CMIP6 models. *Weather and Climate Extremes*, *39*, 100548.

Ferreira, M. A. (2024). *Modeling spatio-temporal data: Markov random fields, objective bayes, and multiscale models*. CRC Press.

Fischer, E. M., & Schär, C. (2010). Consistent geographical patterns of changes in high-impact European heatwaves. *Nature Geoscience*, *3*(6), 398–403.

Fisher, R. A., & Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, *24*(2), 180–190.

Genton, M. G., Padoan, S. A., & Sang, H. (2015). Multivariate max-stable spatial processes. *Biometrika*, *102*(1), 215–230.

Hazra, A., Huser, R., & Bolin, D. (2025). Efficient modeling of spatial extremes over large geographical domains. *Journal of Computational and Graphical Statistics*, *34*(3), 795–811.

Healy, D., Tawn, J., Thorne, P., & Parnell, A. (2025). Inference for extreme spatial temperature events in a changing climate with application to Ireland. *Journal of the Royal Statistical Society Series C: Applied Statistics*, *74*(2), 275–299.

Huser, R., & Davison, A. C. (2014). Space–time modelling of extreme events. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *76*(2), 439–461.

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

Huser, R., Opitz, T., & 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., & Wadsworth, J. L. (2019). Modeling spatial processes with unknown extremal dependence class. *Journal of the American Statistical Association*, *114*(525), 434–444.

Huser, R., & Wadsworth, J. L. (2022). Advances in statistical modeling of spatial extremes. *Wiley Interdisciplinary Reviews: Computational Statistics*, *14*(1), e1537.

Intergovernmental Panel on Climate Change (IPCC). (2021). *Climate change 2021: The physical science basis* (V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, & B. Zhou, Eds.). Cambridge University Press. https://doi.org/10.1017/9781009157896

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*(348), 158–171.

Kiriliouk, A., & Naveau, P. (2020). Climate extreme event attribution using multivariate peaks-over-thresholds modeling and counterfactual theory. *The Annals of Applied Statistics*, *14*(3), 1342–1358.

Kiriliouk, A., Segers, J., & Warcho l, M. (2015). Nonparametric estimation of extremal dependence. *Extreme Value Modelling and Risk Analysis: Methods and Applications, Eds D Dey, J Yan. Boca Raton, FL: Chapman and Hall/CRC*.

Koenker, R. (2005). *Quantile regression* (Vol. 38). Cambridge University Press.

Li, M., & Castro-Camilo, D. (2025). On the importance of tail assumptions in climate extreme event attribution. *arXiv preprint arXiv:2507.14019*.

Matérn, B. (1960). *Spatial variation: Stachastic models and their application to some problems in forest surveys and other sampling investigations* [Doctoral dissertation, Stockholm University].

Maume-Deschamps, V., Ribereau, P., & Zeidan, M. (2024). A spatio-temporal model for temporal evolution of spatial extremal dependence. *Spatial Statistics*, *64*, 100860.

Perkins, S. E. (2015). A review on the scientific understanding of heatwaves—their measurement, driving mechanisms, and changes at the global scale. *Atmospheric Research*, *164*, 242–267.

Philip, S., Kew, S., Van Oldenborgh, G., Anslow, F., Seneviratne, S., Vautard, R., Coumou, D., Ebi, K., Arrighi, J., Singh, R., et al. (2022). EL: Rapid attribution analysis of the extraordinary heat wave on the Pacific coast of the US and Canada in June 2021. *Earth Syst. Dynam.*, *13*, 1689–1713.

Porcu, E., Mateu, J., & Bevilacqua, M. (2007). Covariance functions that are stationary or nonstationary in space and stationary in time. *Statistica Neerlandica*, *61*(3), 358–382.

Quilcaille, Y., Gudmundsson, L., Schumacher, D. L., Gasser, T., Heede, R., Heri, C., Lejeune, Q., Nath, S., Naveau, P., Thiery, W., et al. (2025). Systematic attribution of heatwaves to the emissions of carbon majors. *Nature*, *645*(8080), 392–398.

Reich, B. J., Fuentes, M., & Dunson, D. B. (2011). Bayesian spatial quantile regression. *Journal of the American Statistical Association*, *106*(493), 6–20.

Reiss, R.-D., & Thomas, M. (2007). *Statistical analysis of extreme values: With applications to insurance, finance, hydrology and other fields*. Springer.

Richards, J., Sainsbury-Dale, M., Zammit-Mangion, A., & Huser, R. (2024). Neural Bayes estimators for censored inference with peaks-over-threshold models. *Journal of Machine Learning Research*, *25*(390), 1–49.

Risser, M. D., Zhang, L., & Wehner, M. F. (2025). Data-driven upper bounds and event attribution for unprecedented heatwaves. *Weather and Climate Extremes*, *47*, 100743.

Rousi, E., Kornhuber, K., Beobide-Arsuaga, G., Luo, F., & Coumou, D. (2022). Accelerated western European heatwave trends linked to more-persistent double jets over Eurasia. *Nat. Commun.*, *13*, 3851.

Rue, H., & Held, L. (2005). *Gaussian Markov random fields: Theory and applications*. Chapman; Hall/CRC.

Sainsbury-Dale, M., Zammit-Mangion, A., & Huser, R. (2024). Likelihood-free parameter estimation with neural Bayes estimators. *The American Statistician*, *78*(1), 1–14.

Sainsbury-Dale, M., Zammit-Mangion, A., Richards, J., & Huser, R. (2025). Neural Bayes estimators for irregular spatial data using graph neural networks. *Journal of Computational and Graphical Statistics*, *34*(3), 1153–1168.

Sang, H., & Gelfand, A. E. (2009). Hierarchical modeling for extreme values observed over space and time. *Environmental and ecological statistics*, *16*(3), 407–426.

Schabenberger, O., & Gotway, C. A. (2017). *Statistical methods for spatial data analysis*. Chapman; Hall/CRC.

Seneviratne, S. I., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Di Luca, A., et al. (2021). Weather and climate extreme events in a changing climate. In *Climate change 2021: The* *physical science basis. contribution of working group i to the sixth assessment report of the intergovernmental panel on climate change*. Cambridge University Press.

Shi, M., Zhang, L., Risser, M. D., & Shaby, B. A. (2026). Spatial scale-aware tail dependence modeling for high-dimensional spatial extremes. *Journal of the American Statistical Association*, (just-accepted), 1–23.

Simpson, E. S., & Wadsworth, J. L. (2021). Conditional modelling of spatio-temporal extremes for Red Sea surface temperatures. *Spatial Statistics*, *41*, 100482.

Sklar, M. (1959). Fonctions de repartition à *n* dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, *8*, 229–231.

Smith, J. A., & Karr, A. F. (1990). A statistical model of extreme storm rainfall. *Journal of Geophysical Research: Atmospheres*, *95*(D3), 2083–2092.

Smith, R. L. (1990). Max-stable processes and spatial extremes. *Unpublished manuscript*, *205*, 1–32.

Stefanon, M., D'Andrea, F., & Drobinski, P. (2012). Heatwave classification over Europe and the Mediterranean region. *Environmental Research Letters*, *7*(1), 014023.

Tebaldi, C., Debeire, K., Eyring, V., Fischer, E., Fyfe, J., Friedlingstein, P., Knutti, R., Lowe, J., O'Neill, B., Sanderson, B., et al. (2021). Climate model projections from the scenario model intercomparison project (ScenarioMIP) of CMIP6. *Earth System Dynamics*, *12*(1), 253–293.

Tian, Y., Kleidon, A., Lesk, C., Zhou, S., Luo, X., Ghausi, S. A., Wang, G., Zhong, D., & Zscheischler, J. (2024). Characterizing heatwaves based on land surface energy budget. *Communications Earth & Environment*, *5*(1), 617.

van Oldenborgh, G. J., Wehner, M. F., Vautard, R., Otto, F. E. L., Seneviratne, S. I., & Stott, P. A. (2022). Attributing and projecting heatwaves is hard: We can do better. *Earth's Future*, *10*(6), e2021EF002271.

Vautard, R., Gobiet, A., Jacob, D., Belda, M., Colette, A., Déqué, M., Fernández, J., García-Díez, M., Goergen, K., Güttler, I., et al. (2013). The simulation of european heat waves from an ensemble of regional climate models within the EURO-CORDEX project. *Climate dynamics*, *41*, 2555–2575.

Vautard, R., van Aalst, M., Boucher, O., Drouin, A., Haustein, K., Kreienkamp, F., van Oldenborgh, G. J., Otto, F. E., Ribes, A., Robin, Y., et al. (2020). Human contribution to the recordbreaking June and July 2019 heatwaves in Western Europe. *Environmental Research Letters*, *15*(9), 094077.

Vihrs, N. (2022). Using neural networks to estimate parameters in spatial point process models. *Spatial Statistics*, *51*, 100668.

Wadsworth, J., & Tawn, J. (2018). Spatial conditional extremes. *Manuscript submitted for publication*.

Wake, B. (2025). Heatwave attribution in seconds: Extreme events. *Nature Climate Change*, 1–1.

Yukimoto, S., Kawai, H., Koshiro, T., Oshima, N., Yoshida, K., Urakawa, S., Tsujino, H., Deushi, M., Tanaka, T., Hosaka, M., et al. (2019). The Meteorological Research Institute Earth System Model version 2.0, MRI-ESM2. 0: Description and basic evaluation of the physical component. *Journal of the Meteorological Society of Japan. Ser. II*, *97*(5), 931–965.

Zaitchik, B. F., Macalady, A. K., Bonneau, L. R., & Smith, R. B. (2006). Europe's 2003 heat wave: A satellite view of impacts and land–atmosphere feedbacks. *International Journal of Climatology*, *26*(6), 743–769.

Zammit-Mangion, A., Sainsbury-Dale, M., & Huser, R. (2025). Neural methods for amortized inference. *Annual Review of Statistics and Its Application*, *12*(1), 311–335.

Supplementary Material for " A spatio-temporal statistical framework for heatwave attribution under climate change "

S1 Supplementary Methods

S1.1 Bayesian Spatial Quantile Regression

This section presents the Bayesian hierarchical model corresponding to the *approximate method* of Reich et al. (2011, Sec. 3). We adopt the BSQR model to characterize the conditional distribution of temperature anomalies $X(\mathbf{s}, d, t)$ given covariates $\mathbf{c}(t) = (c_1(t), \ldots, c_p(t))^\top$. The covariates are transformed such that $c_1(t) = 1$ is the intercept and $c_j(t) \in [0, 1]$ for $j \geq 2$. In practice, this transformation can be performed using a min–max rescaling or by applying the inverse-CDF–based normalization. In principle, the covariates could also depend on the day d, allowing for higher temporal resolution (e.g., monthly or daily climate indicators). However, in our application the only used large-scale covariate is the GMST, which is observed at monthly resolution and varies slowly over time. Because GMST primarily reflects long-term climate forcing rather than day-to-day variability, we treat it as a yearly covariate in the model. This simplification reduces model complexity while remaining consistent with the temporal scale at which the forcing evolves. The model allows covariate effects to vary smoothly across quantile level τ and spatial location $\mathbf{s}$ as we expect nearby locations to be exposed to the same regional events, while enforcing non-crossing quantile curves. The approximation replaces the full data likelihood with a Gaussian likelihood for site-specific quantile regression coefficient estimates, while retaining the spatial quantile-process prior of Reich et al. (2011, Sec. 2). Let $q(\tau \mid \mathbf{c}(t), \mathbf{s})$ denote the conditional τ-th quantile of $X(\mathbf{s}, d, t)$. Following Reich et al. (2011), we assume

$$q(\tau \mid \mathbf{c}(t), \mathbf{s}) = \mathbf{c}(t)^\top \boldsymbol{\beta}^{\mathrm{QR}}(\tau, \mathbf{s}), \tag{S1}$$

where $\boldsymbol{\beta}^{\mathrm{QR}}(\tau, \mathbf{s}) = (\beta_1^{\mathrm{QR}}(\tau, \mathbf{s}), \ldots, \beta_p^{\mathrm{QR}}(\tau, \mathbf{s}))^\top$ are spatially varying quantile coefficient functions. The approximation starts by fitting for each spatial location $\mathbf{s} \in \mathcal{R}$, a classical quantile regression at quantile levels $\tau_1, \ldots, \tau_K$. For each τ_k and site $\mathbf{s}$, the Koenker–Bassett estimator (Koenker, 2005) is

$$\hat{\boldsymbol{\beta}}^{\mathrm{QR}}(\tau_k, \mathbf{s}) = \big(\hat{\beta}_1^{\mathrm{QR}}(\tau_k, \mathbf{s}), \ldots, \hat{\beta}_p^{\mathrm{QR}}(\tau_k, \mathbf{s})\big)^\top = \underset{\boldsymbol{\beta}^{\mathrm{QR}}}{\arg\min} \sum_t \sum_d \rho_{\tau_k}\big(x(\mathbf{s}, d, t) - \mathbf{c}(t)^\top \boldsymbol{\beta}^{\mathrm{QR}}\big), \quad \text{(S2)}$$

where $\rho_\tau(u) = u \cdot (\tau - \mathbb{1}\{u < 0\})$ is the pinball loss. These estimators are consistent for the true quantile coefficients and possess a known asymptotic covariance structure. Specifically, for any two quantile levels τ_k and τ_l, we have

$$\mathsf{cov}\Big(\sqrt{DT}\,\hat{\boldsymbol{\beta}}^{\mathrm{QR}}(\tau_k, \mathbf{s}),\ \sqrt{DT}\,\hat{\boldsymbol{\beta}}^{\mathrm{QR}}(\tau_l, \mathbf{s})\Big) \approx H(\tau_k, \mathbf{s})^{-1} J(\tau_k, \tau_l, \mathbf{s}) H(\tau_l, \mathbf{s})^{-1}. \quad \text{(S3)}$$

where $H(\tau, \mathbf{s})$ and $J(\tau_k, \tau_l, \mathbf{s})$ are the standard quantile regression sandwich $p \times p$ matrices defined in Koenker (2005). We define the stacked vector of estimated coefficients

$$\hat{\boldsymbol{\beta}}^{\mathrm{QR}}(\mathbf{s}) = \Big(\hat{\boldsymbol{\beta}}^{\mathrm{QR}}(\tau_1, \mathbf{s})^\top, \ldots, \hat{\boldsymbol{\beta}}^{\mathrm{QR}}(\tau_K, \mathbf{s})^\top\Big)^\top \in \mathbb{R}^{pK}, \quad \text{(S4)}$$

and $\mathsf{cov}(\hat{\boldsymbol{\beta}}^{\mathrm{QR}}(\mathbf{s})) = \Sigma(\mathbf{s})$ is the $pK \times pK$ covariance matrix, where the elements are given by Eq. (S3). Let $\hat{\Sigma}(\mathbf{s}) \in \mathbb{R}^{pK \times pK}$ denote the estimated asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}^{\mathrm{QR}}(\mathbf{s})$ obtained from standard quantile regression theory. Its $((j,k),(j',\ell))$ entry is

$$\hat{\Sigma}_{(j,k),(j',\ell)}(\mathbf{s}) = \hat{H}(\tau_k, \mathbf{s})^{-1}\, \hat{J}(\tau_k, \tau_\ell, \mathbf{s})\, \hat{H}(\tau_\ell, \mathbf{s})^{-1}, \quad \text{(S5)}$$

for $j, j' = 1, \ldots, p$ and $k, \ell = 1, \ldots, K$, where

$$\hat{H}(\tau, \mathbf{s}) = \frac{1}{DT} \sum_{t,d} \mathbf{c}(t)\mathbf{c}(t)^\top\, \hat{f}_{\mathbf{s}}\Big(\mathbf{c}(t)^\top \hat{\boldsymbol{\beta}}^{\mathrm{QR}}(\tau, \mathbf{s})\Big), \quad \text{(S6)}$$

$$\hat{J}(\tau_k, \tau_\ell, \mathbf{s}) = (\tau_k \wedge \tau_\ell - \tau_k \tau_\ell)\ \frac{1}{DT} \sum_{t,d} \mathbf{c}(t)\mathbf{c}(t)^\top, \quad \text{(S7)}$$

with $\hat{f}_{\mathbf{s}}(\cdot)$ a nonparametric density estimate at location $\mathbf{s}$ evaluated at the fitted quantile $\mathbf{c}(t)_i^\top \hat{\boldsymbol{\beta}}^{\mathrm{QR}}(\tau, \mathbf{s})$.

In the second step, the quantile-regression coefficients $\hat{\boldsymbol{\beta}}^{\mathrm{QR}}(\mathbf{s})$ are assumed to follow a normal distribution,

$$\hat{\boldsymbol{\beta}}^{\mathrm{QR}}(\mathbf{s}) \sim \mathcal{N}\big(\boldsymbol{\beta}^{\mathrm{QR}}(\mathbf{s}),\ \Sigma(\mathbf{s})\big), \quad \text{(S8)}$$

independently over $\mathbf{s} \in \mathcal{R}$. Where $\Sigma(\mathbf{s})$ will be replaced by its estimator $\hat{\Sigma}(\mathbf{s})$ defined in Eq. (S5). For each covariate $j = 1, \ldots, p$ and location $\mathbf{s}$, the quantile-regression coefficient function is modeled using a Bernstein polynomial expansion,

$$\beta_j^{\mathrm{QR}}(\tau, \mathbf{s}) = \sum_{m=1}^{M} B_m(\tau)\, \alpha_{jm}(\mathbf{s}), \qquad B_m(\tau) = \binom{M}{m} \tau^m (1-\tau)^{M-m}, \qquad \tau \in [0,1]. \quad \text{(S9)}$$

We choose $M = 15$ basis functions. Following Reich et al. (2011), moderate values of M are sufficient to accurately approximate smooth quantile curves, and their simulation results were shown to be robust to the choice of M. The vector $\boldsymbol{\beta}^{\mathrm{QR}}(\mathbf{s})$ in Eq. (S8) is obtained by evaluating $\beta_j^{\mathrm{QR}}(\tau, \mathbf{s})$ at $\tau_1, \ldots, \tau_K$ and stacking over j as in Eq. (S4). Define the increment parameters $\{\delta_{jm}(\mathbf{s})\}$ by

$$\delta_{j1}(\mathbf{s}) = \alpha_{j1}(\mathbf{s}), \tag{S10}$$

$$\delta_{jm}(\mathbf{s}) = \alpha_{jm}(\mathbf{s}) - \alpha_{j,m-1}(\mathbf{s}), \qquad m = 2, \ldots, M, \tag{S11}$$

so that

$$\alpha_{jm}(\mathbf{s}) = \sum_{\ell=1}^{m} \delta_{j\ell}(\mathbf{s}), \qquad m = 1, \ldots, M. \tag{S12}$$

Next, let us introduce unconstrained latent variables $\delta_{jm}^*(\mathbf{s})$. For $m = 2, \ldots, M$, define the constrained increments by

$$\delta_{jm}(\mathbf{s}) = \begin{cases} \delta_{jm}^*(s), & \text{if } \delta_{1m}^*(\mathbf{s}) + \sum_{i=2}^{p} \mathbb{1}\left\{\delta_{im}^*(\mathbf{s}) < 0\right\} \delta_{im}^*(\mathbf{s}) \geq 0, \\ 0, & \text{otherwise,} \end{cases} \tag{S13}$$

and for $m = 1$ set $\delta_{j1}(\mathbf{s}) = \delta_{j1}^*(\mathbf{s})$. Under the assumption that the intercept covariate $c_1(t)$ equals 1 and that $c_j(t) \in [0, 1]$ for $j = 2, \ldots, p$, this construction ensures that the conditional quantile function is nondecreasing in τ for all $\mathbf{c}(t)$; see Reich et al. (2011, Section 2.2 Eq. (11)).

To introduce spatial smoothness, for each pair (j, m), the latent variables are assumed to follow independent Gaussian spatial processes:

$$\{\delta_{jm}^*(\mathbf{s}) : \mathbf{s} \in \mathcal{R}\} \sim GP\left(\bar{\delta}_{jm}(\Omega),\ \tilde{\sigma}_j^2 \exp\left(-\frac{\|\mathbf{s} - \mathbf{s}'\|}{\rho_j}\right)\right). \tag{S14}$$

Where $\bar{\delta}_{jm}(\Omega)$ is chosen to center the quantile process on a parametric base distribution $f_0(y \mid \Omega)$. Following Reich et al. (2011), we take f_0 to be a skew-normal distribution with parameter vector $\Omega = (\psi_1, \psi_2, \psi_3)^\top$. For the intercept term ($j = 1$), the prior mean coefficients $\bar{\delta}_{1m}(\Omega)$ are selected so that the implied quantile function is centered on the corresponding skew-normal quantile function $q_0(\tau \mid \Omega)$. Specifically,

$$q_0(\tau \mid \Omega) \approx \sum_{m=1}^{M} B_m(\tau)\, \bar{\alpha}_{1m}(\Omega), \qquad \bar{\alpha}_{1m}(\Omega) = \sum_{\ell=1}^{m} \bar{\delta}_{1\ell}(\Omega).$$

The increment parameters $\{\bar{\delta}_{1m}(\Omega)\}$ are obtained by solving the constrained ridge regression problem with $\lambda = 1$ following Reich et al. (2011)

$$(\bar{\delta}_{11}(\Omega), \ldots, \bar{\delta}_{1M}(\Omega)) = \underset{\{d_m\}}{\arg\min} \sum_{k=1}^{K} \left[q_0(\tau_k \mid \Omega) - \sum_{m=1}^{M} B_m(\tau_k) \left(\sum_{\ell=1}^{m} d_\ell \right) \right]^2 + \sum_{m=1}^{M} d_m^2, \tag{S15}$$

subject to $d_m \geq 0$ for $m \geq 2$, and we set $\bar{\delta}_{1m}(\Omega) = d_m$. For all non-intercept terms ($j = 2, \ldots, p$), we set

$$\bar{\delta}_{jm}(\Omega) = 0, \qquad j = 2, \ldots, p; m = 1, \ldots, M,$$

shrinking covariate effects toward no effect in the absence of strong evidence from the data. For each covariate $j = 1, \ldots, p$, the spatial variance and range parameters in Eq. (S14) are assigned independent priors

$$\tilde{\sigma}_j^2 \sim \text{Inv-Gamma}(a_\sigma, b_\sigma), \tag{S16}$$

$$\rho_j \sim \text{Gamma}(a_\rho, b_\rho), \tag{S17}$$

and the base-distribution parameters Ω receive a prior $\Omega \sim \pi_\Omega(\Omega)$, independent of $\{\tilde{\sigma}_j^2, \rho_j\}_{j=1}^{p}$. A schematic summary of the hierarchical construction is provided in Figure S1.

Let $\Theta_{\mathcal{R}}^{\text{QR}}$ denote the collection of all unknown parameters,

$$\Theta_{\mathcal{R}}^{\text{QR}} = \left\{ \delta_{jm}^*(\mathbf{s}) : j = 1, \ldots, p,\ m = 1, \ldots, M,\ \mathbf{s} \in \mathcal{R} \right\} \cup \left\{ \tilde{\sigma}_j^2, \rho_j : j = 1, \ldots, p \right\} \cup \{\Omega\}. \tag{S18}$$

For each spatial location $\mathbf{s}$, the Bernstein polynomial representation in Eq. (S9) induces a finite-dimensional vector of quantile regression coefficients evaluated at the grid $\tau_1, \ldots, \tau_K$. Specifically, we define

$$\boldsymbol{\beta}^{\text{QR}}(\mathbf{s}) := \left(\beta_1^{\text{QR}}(\tau_1, \mathbf{s}), \ldots, \beta_1^{\text{QR}}(\tau_K, \mathbf{s}),\ \ldots,\ \beta_p^{\text{QR}}(\tau_1, \mathbf{s}), \ldots, \beta_p^{\text{QR}}(\tau_K, \mathbf{s}) \right)^\top \in \mathbb{R}^{pK}, \tag{S19}$$

where $\beta_j^{\text{QR}}(\tau, \mathbf{s})$ is given by Eq. (S9). The posterior distribution of Θ_{QR} given the estimated quantile regression coefficients $\hat{\boldsymbol{\beta}}^{\text{QR}}$ is

$$\begin{aligned} \pi\left(\Theta_{\mathcal{R}}^{QR} \mid \{\hat{\boldsymbol{\beta}}^{\text{QR}}(\mathbf{s})\}_{\mathbf{s}\in\mathcal{R}} \right) \propto{} & \prod_{\mathbf{s}\in\mathcal{R}} \mathcal{N}\left(\hat{\boldsymbol{\beta}}^{\text{QR}}(\mathbf{s}) \mid \boldsymbol{\beta}^{\text{QR}}(\mathbf{s}), \hat{\Sigma}(\mathbf{s}) \right) \\ & \times \prod_{j=1}^{p} \prod_{m=1}^{M} GP\left(\delta_{jm}^*(\cdot) \mid \bar{\delta}_{jm}(\Omega), \tilde{\sigma}_j^2, \rho_j \right) \pi_\Omega(\Omega) \prod_{j=1}^{p} \pi(\tilde{\sigma}_j^2) \pi(\rho_j). \end{aligned} \tag{S20}$$

This posterior induces spatially smooth, noncrossing quantile functions while borrowing strength across spatial locations and quantile levels. From Eq. (S20), we see that the posterior distribution of the model parameters is conditioned on frequentist quantile regression estimates, rather than on the original data directly. This constitutes the key step of the approximate approach proposed by Reich et al. (2011), and leads to a substantial reduction in computational complexity. For a fully Bayesian formulation, we refer the reader to Reich et al. (2011, Section 2).

S1.2 GEV method

This section provides a schematic overview of the hierarchical Bayesian GEV model described in the main text. Figure S2 summarizes the full model structure, linking observed annual maxima $M_{\mathbf{s},t}$ and global covariates $\mathbf{c}(t)$ to the GEV parameters. Spatial variability in the GEV coefficients is introduced through the decomposition $\boldsymbol{\gamma}_\ell = \gamma_{\ell,g}\mathbb{1} + s_\ell \boldsymbol{u}_\ell$, where $\boldsymbol{u}_\ell$ follows an ICAR prior. These spatially varying coefficients are then combined with covariates to define the location, scale, and shape parameters of the GEV distribution at each site.

S1.3 Marginal distribution of $R^*(\mathbf{s}, d)$

For a fixed location $\mathbf{s}$ and day d, the process $R^*(\mathbf{s}, d)$ is the maximum over the contributions of the K episodes. Let

$$H_k(\mathbf{s}, d) = A_k \exp\left(-\frac{\|\boldsymbol{T}(\mathbf{s} - \boldsymbol{C}_k)\|}{\rho_s} - \frac{|d - T_k|}{\rho_t}\right)$$

denote the contribution of episode k at $(\mathbf{s}, d)$, so that

$$R^*(\mathbf{s}, d) = \max_{k=1,\dots,K} H_k(\mathbf{s}, d).$$

Let $F_{\mathbf{s},d}$ denote the distribution function of $R^*(\mathbf{s}, d)$. Conditionally on K, the variables $\{H_k(\mathbf{s}, d)\}_{k=1}^K$ are independent and identically distributed, which yields

$$\mathbb{P}(R^*(\mathbf{s}, d) \le x \mid K) = \mathbb{P}(H_1(\mathbf{s}, d) \le x)^K .$$

Since $K \sim \text{Poisson}(\lambda)$, taking expectation with respect to K gives

$$F_{\mathbf{s},d}(x) = \mathbb{P}(R^*(\mathbf{s}, d) \le x) = \exp\big\{-\lambda\big(1 - p_{\mathbf{s},d}(x)\big)\big\},$$

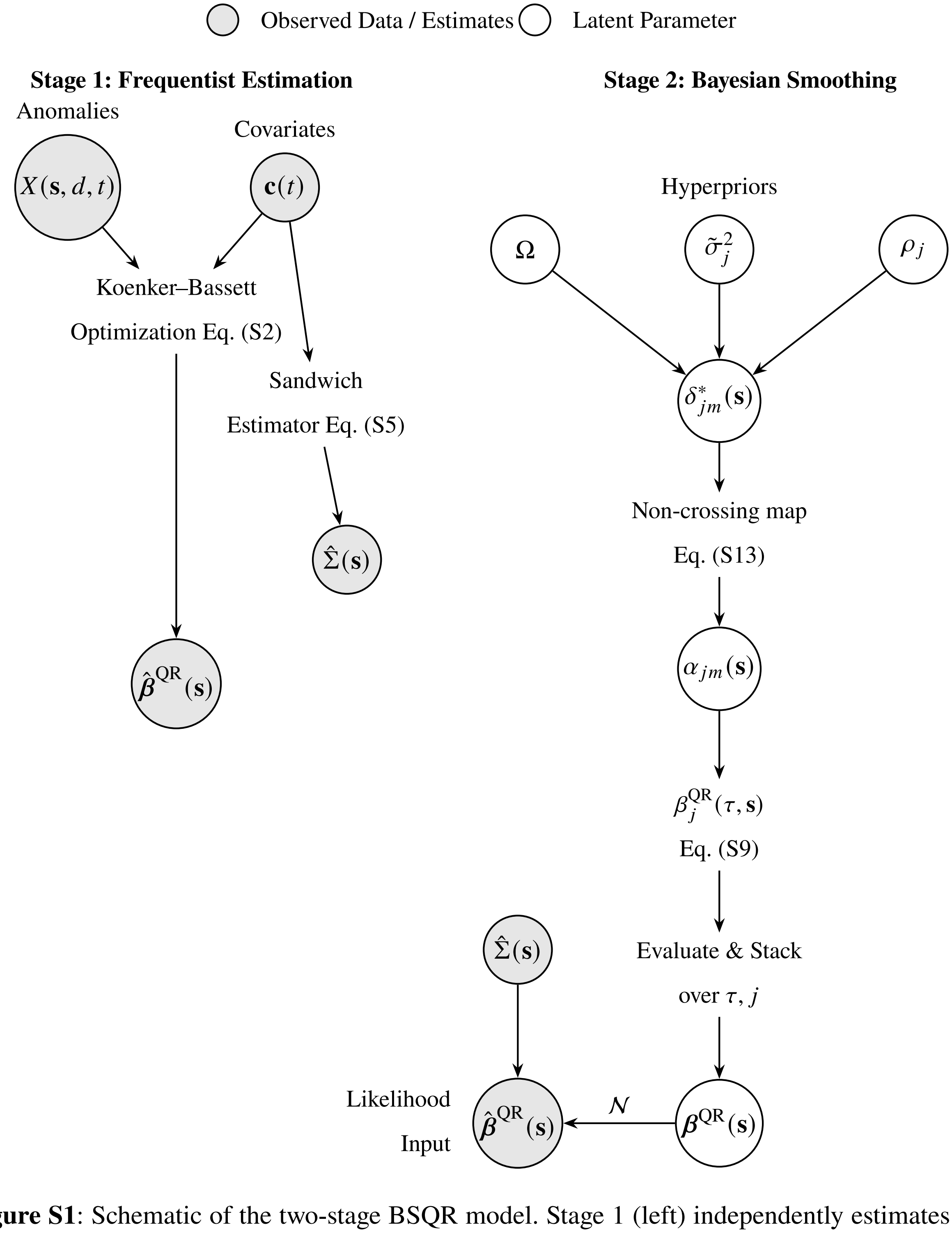


Figure S1: Schematic of the two-stage BSQR model. Stage 1 (left) independently estimates the quantile coefficients and their covariance at each location $\mathbf{s}$. Stage 2 (right) re-incorporates these estimates as observed data (shaded nodes) to perform spatial smoothing via a Bernstein polynomial-based Gaussian process prior.

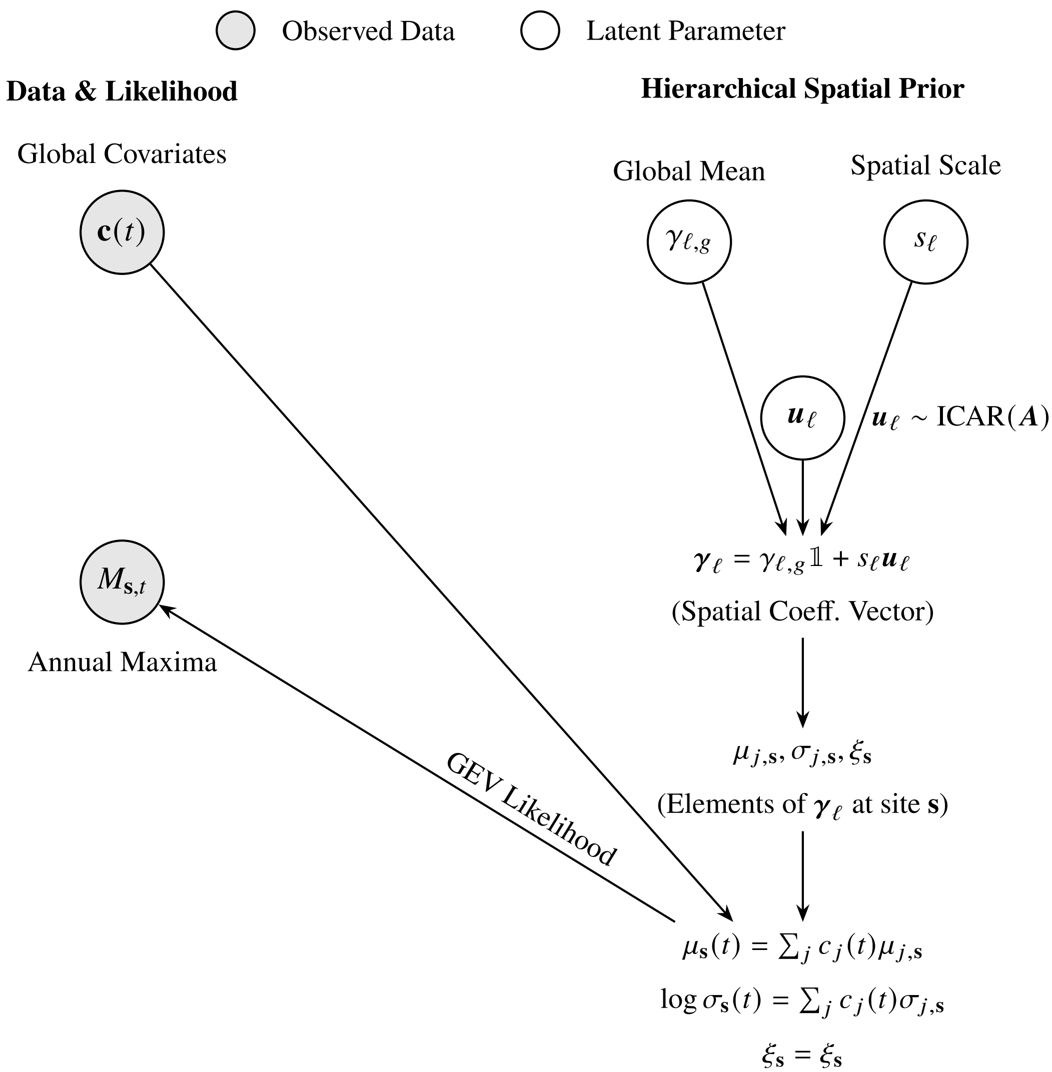


Figure S2: Schematic representation of the spatial GEV framework. Spatial dependence is modeled via ICAR priors on latent fields $\mathbf{u}_\ell$, forming spatially varying coefficients $\boldsymbol{\gamma}_\ell$. These coefficients are combined with global covariates $\mathbf{c}(t)$ through linear predictors to define the time-dependent GEV distribution at each location $\boldsymbol{s}$.

where

$$p_{\mathbf{s},d}(x) = \mathbb{P}(H_1(\mathbf{s}, d) \leq x)\,.$$

Using the uniform distribution of $(\boldsymbol{C}_1, T_1)$ over $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\} \times \{1, \ldots, D\}$, we obtain

$$p_{\mathbf{s},d}(x) = \frac{1}{nD} \sum_{i=1}^{n} \sum_{t=1}^{D} \mathbb{P}\left(A_1 \exp\left(-\frac{\|\boldsymbol{T}(\mathbf{s} - \mathbf{s}_i)\|}{\rho_s} - \frac{|d-t|}{\rho_t}\right) \leq x\right).$$

Because $A_1 \sim \text{Pareto}(1)$, the inner probability can be written explicitly as

$$\mathbb{P}(A_1 w \leq x) = \left(1 - \frac{w}{x}\right)_+, \qquad w = \exp\left(-\frac{\|\boldsymbol{T}(\mathbf{s} - \mathbf{s}_i)\|}{\rho_s} - \frac{|d-t|}{\rho_t}\right),$$

where $(u)_+ = \max(u, 0)$. Consequently,

$$p_{\mathbf{s},d}(x) = \frac{1}{nD} \sum_{i=1}^{n} \sum_{t=1}^{D} \left(1 - \frac{\exp\left(-\frac{\|\boldsymbol{T}(\mathbf{s}-\mathbf{s}_i)\|}{\rho_s} - \frac{|d-t|}{\rho_t}\right)}{x}\right)_+\,.$$

S1.4 Chi grid

This section illustrates one of the summary statistics used as input to the neural network. Figure S3 shows the tail dependence grids based on the transformed statistic $\log(\bar{\chi} + \varepsilon)$, computed across spatial distance bins and temporal lags for different thresholds u. Each grid encodes the dependence structure of extremes over space and time, and serves as a compact representation of extremal dependence patterns used for model training.

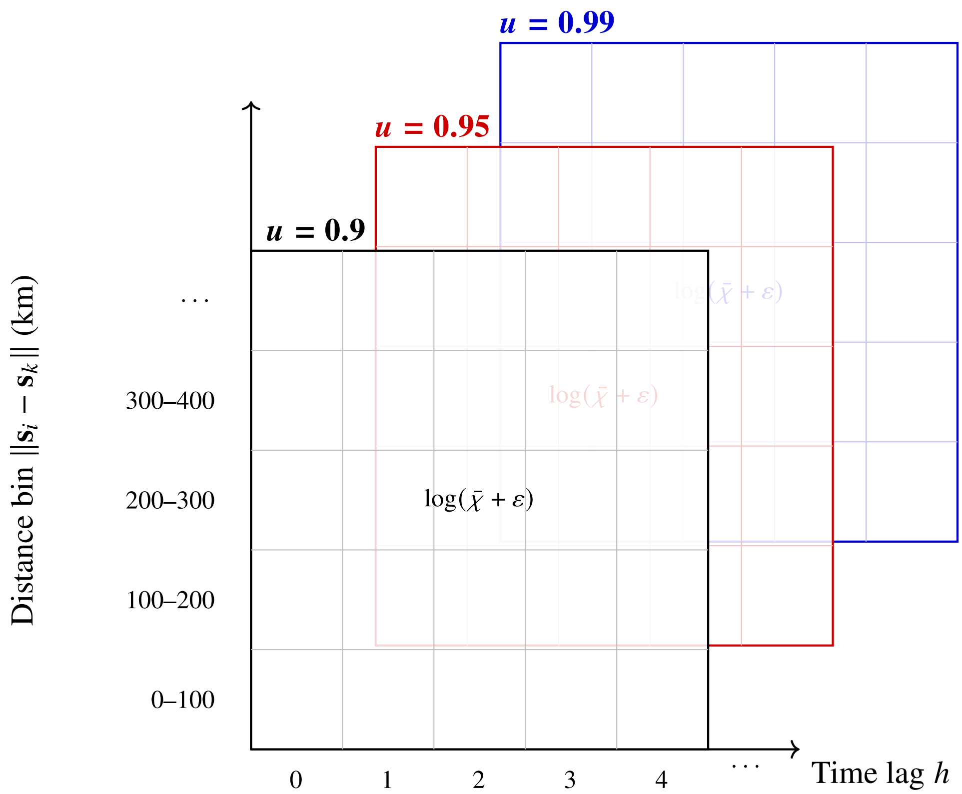


Figure S3: Visual representation of the estimated tail dependence grids $\log(\bar{\chi} + \varepsilon)$ for thresholds $u \in \{0.9, 0.95, 0.99\}$. Each grid layer represents a two-dimensional matrix where rows correspond to spatial distance bins and columns to discrete temporal lags h. Each cell represents the bin-averaged tail dependence estimate $\log(\bar{\chi} + \varepsilon)$ for a specific distance-lag combination.

S1.5 Back-transformation

This section provides the detailed derivation of the marginal back-transformation presented in Section 3.4 of the main document. We assume throughout that the shape parameter $\xi_{\mathbf{s}}$ is constant in time (i.e., does not depend on t). Furthermore, we assume that the temporal process $\{X(\mathbf{s}, d, t)\}$ admits an extremal index $\theta_{\mathbf{s}} \in (0, 1]$, characterizing the clustering behavior of extremes. In addition, we impose Leadbetter's $D(u_n)$ and $D'(u_n)$ mixing conditions to control the dependence structure of exceedances and ensure the validity of the asymptotic arguments for block maxima. Starting from the standard tail decomposition, for a large value $x(\mathbf{s}, d, t)$ exceeding a high marginal threshold

$\vartheta_{\mathbf{s}}^{\text{thres}}$ of $X(\mathbf{s}, d, t)$, we decompose

$$\mathsf{P}(X(\mathbf{s}, d, t) > x(\mathbf{s}, d, t)) = \mathsf{P}\left(X(\mathbf{s}, d, t) > \vartheta_{\mathbf{s}}^{\text{thres}}\right) \mathsf{P}\left(X(\mathbf{s}, d, t) > x(\mathbf{s}, d, t) \mid X(\mathbf{s}, d, t) > \vartheta_{\mathbf{s}}^{\text{thres}}\right).$$

Let $M_{\mathbf{s},t} = \max_{d=1,\dots,D} X(\mathbf{s}, d, t)$ denote the annual maximum. Under the above mixing conditions, the series $(M_{\mathbf{s},t})_t$ has an extremal index $\theta_{\mathbf{s}}$ and converges in distribution to a GEV distribution with parameters $(\mu_{\mathbf{s},t}, \sigma_{\mathbf{s},t}, \xi_{\mathbf{s}})$:

$$G_{\mathbf{s},t}(x) = \begin{cases} \exp\Big[-\big(1 + \xi_{\mathbf{s}}\frac{x-\mu_{\mathbf{s},t}}{\sigma_{\mathbf{s},t}}\big)^{-1/\xi_{\mathbf{s}}}\Big], & \xi_{\mathbf{s}} \neq 0, \quad 1 + \xi_{\mathbf{s}}\frac{x-\mu_{\mathbf{s},t}}{\sigma_{\mathbf{s},t}} > 0, \\ \exp\Big[-\exp\big(-\frac{x-\mu_{\mathbf{s},t}}{\sigma_{\mathbf{s},t}}\big)\Big], & \xi_{\mathbf{s}} = 0. \end{cases}$$

Since $M_{\mathbf{s},t}$ is the maximum of D dependent replicates,

$$\mathsf{P}(M_{\mathbf{s},t} \leq x(\mathbf{s}, d, t)) = F_{\mathbf{s},t}(x(\mathbf{s}, d, t))^{\theta_{\mathbf{s}} D} \approx \exp\Big[-\theta_{\mathbf{s}} D\{1 - F_{\mathbf{s},t}(x(\mathbf{s}, d, t))\}\Big],$$

where the last approximation uses the first-order tail expansion $\log(1 - y) \approx -y$ valid when $y = 1 - F_{\mathbf{s},t}(x) \to 0$. Using the GEV limit relation

$$F_{\mathbf{s},t}(x)^{\theta_{\mathbf{s}} D} \approx G_{\mathbf{s},t}(x),$$

we obtain the tail approximation

$$1 - F_{\mathbf{s},t}(x(\mathbf{s}, d, t)) \approx \frac{-\log G_{\mathbf{s},t}(x(\mathbf{s}, d, t))}{\theta_{\mathbf{s}} D} = \begin{cases} \dfrac{\big(1 + \xi_{\mathbf{s}}\frac{x(\mathbf{s},d,t)-\mu_{\mathbf{s},t}}{\sigma_{\mathbf{s},t}}\big)^{-1/\xi_{\mathbf{s}}}}{\theta_{\mathbf{s}} D}, & \xi_{\mathbf{s}} \neq 0, \\ \dfrac{\exp\big(-\frac{x(\mathbf{s},d,t)-\mu_{\mathbf{s},t}}{\sigma_{\mathbf{s},t}}\big)}{\theta_{\mathbf{s}} D}, & \xi_{\mathbf{s}} = 0. \end{cases}$$

Hence, the conditional tail of the parent distribution is

$$\begin{aligned} \mathsf{P}\left(X(\mathbf{s}, d, t) > x(\mathbf{s}, d, t) \mid X(\mathbf{s}, d, t) > \vartheta_{\mathbf{s}}^{\text{thres}}\right) &= \frac{1 - F_{\mathbf{s},t}(x(\mathbf{s}, d, t))}{1 - F_{\mathbf{s},t}(\vartheta_{\mathbf{s}}^{\text{thres}})} \\ &\approx \begin{cases} \left(\dfrac{1 + \xi_{\mathbf{s}}\frac{x(\mathbf{s},d,t)-\mu_{\mathbf{s},t}}{\sigma_{\mathbf{s},t}}}{1 + \xi_{\mathbf{s}}\frac{\vartheta_{\mathbf{s}}^{\text{thres}}-\mu_{\mathbf{s},t}}{\sigma_{\mathbf{s},t}}}\right)^{-1/\xi_{\mathbf{s}}}, & \xi_{\mathbf{s}} \neq 0, \\ \exp\left(-\dfrac{x(\mathbf{s}, d, t) - \vartheta_{\mathbf{s}}^{\text{thres}}}{\sigma_{\mathbf{s},t}}\right), & \xi_{\mathbf{s}} = 0. \end{cases} \end{aligned}$$

Introducing the generalized Pareto scale parameter

$$\tilde{\sigma}_{\mathbf{s},t} = \sigma_{\mathbf{s},t} + \xi_{\mathbf{s}}(\vartheta_{\mathbf{s}}^{\text{thres}} - \mu_{\mathbf{s},t}),$$

the conditional tail takes the standard GPD form:

$$\mathsf{P}\left(X(\mathbf{s},d,t) > x(\mathbf{s},d,t) \mid X(\mathbf{s},d,t) > \vartheta_{\mathbf{s}}^{\text{thres}}\right) \approx \begin{cases} \left(1+\xi_{\mathbf{s}}\dfrac{x(\mathbf{s},d,t)-\vartheta_{\mathbf{s}}^{\text{thres}}}{\tilde{\sigma}_{\mathbf{s},t}}\right)^{-1/\xi_{\mathbf{s}}}, & \xi_{\mathbf{s}} \neq 0, \\ \exp\left(-\dfrac{x(\mathbf{s},d,t)-\vartheta_{\mathbf{s}}^{\text{thres}}}{\tilde{\sigma}_{\mathbf{s},t}}\right), & \xi_{\mathbf{s}} = 0. \end{cases}$$

For a given year t and simulated uniform value $u(\mathbf{s},d) \in [\tau_{\mathbf{s},t}, 1)$, where

$$\tau_{\mathbf{s},t} = q_{\mathbf{s},t}^{-1}(\vartheta_{\mathbf{s}}^{\text{thres}}),$$

the inverse transformation is obtained by solving

$$1-u(\mathbf{s},d) = (1-\tau_{\mathbf{s},t}) \cdot \begin{cases} \left(1+\dfrac{\xi_{\mathbf{s}}(x(\mathbf{s},d,t)-\vartheta_{\mathbf{s}}^{\text{thres}})}{\tilde{\sigma}_{\mathbf{s},t}}\right)^{-1/\xi_{\mathbf{s}}}, & \xi_{\mathbf{s}} \neq 0, \\ \exp\left(-\dfrac{x(\mathbf{s},d,t)-\vartheta_{\mathbf{s}}^{\text{thres}}}{\tilde{\sigma}_{\mathbf{s},t}}\right), & \xi_{\mathbf{s}} = 0. \end{cases}$$

Solving for $x(\mathbf{s},d,t)$ gives the explicit back-transformation:

$$x(\mathbf{s},d,t) = \begin{cases} \vartheta_{\mathbf{s}}^{\text{thres}} + \dfrac{\tilde{\sigma}_{\mathbf{s},t}}{\xi_{\mathbf{s}}}\left[\left(\dfrac{1-u(\mathbf{s},d)}{1-\tau_{\mathbf{s},t}}\right)^{-\xi_{\mathbf{s}}} - 1\right], & u(\mathbf{s},d) \geq \tau_{\mathbf{s},t},\ \xi_{\mathbf{s}} \neq 0, \\ \vartheta_{\mathbf{s}}^{\text{thres}} - \tilde{\sigma}_{\mathbf{s},t}\log\left(\dfrac{1-u(\mathbf{s},d)}{1-\tau_{\mathbf{s},t}}\right), & u(\mathbf{s},d) \geq \tau_{\mathbf{s},t},\ \xi_{\mathbf{s}} = 0. \end{cases}$$

S2 Supplementary Data Application

S2.1 Clustering for European heatwaves

To this end, we propose a clustering strategy that identifies spatial subregions exhibiting similar patterns of concomitant extremes, in the spirit of Kiriliouk and Naveau (2020) and related approaches (Bernard et al., 2013). We compute clusters using only the counterfactual climate simulations, which are assumed to be stationary by construction.

We first construct pseudo-observations on the uniform scale via within-location ranks. Denoting by DT the number of available summer days and by $\mathrm{rank}\{X(\mathbf{s}, d, t)\}$ the rank of $X(\mathbf{s}, d, t)$ among $\{X(\mathbf{s}, 1, 1), \ldots, X(\mathbf{s}, D, T)\}$, we define

$$\widehat{U}(\mathbf{s}, d, t) = \frac{\mathrm{rank}\{X(\mathbf{s}, d, t)\}}{DT}, \qquad d = 1, \ldots, D,\ t = 1, \ldots, T,$$

so that $\widehat{U}(\mathbf{s}, d, t)$ has approximately uniform margins on $[0, 1]$.

Extremal dependence between two locations $\mathbf{s}_i$ and $\mathbf{s}_k$ is then quantified through a high-threshold tail dependence coefficient at level $u \in (0, 1)$:

$$\hat{\chi}_{\mathbf{s}_i,\mathbf{s}_k}(u) = \frac{1}{DT(1-u)} \sum_{t=1}^{T} \sum_{d=1}^{D} \mathbb{1}\big\{\widehat{U}(\mathbf{s}_i, d, t) > u,\ \widehat{U}(\mathbf{s}_k, d, t) > u\big\}.$$

This quantity is an empirical analogue of the conditional exceedance probability used in Kiriliouk et al. (2015) and yields a simple dependence measure based on simultaneous extremes at identical time instances. Based on $\hat{\chi}_{\mathbf{s}_i,\mathbf{s}_k}(u)$, we follow Kiriliouk and Naveau (2020) and define the dissimilarity

$$d(\mathbf{s}_i, \mathbf{s}_k \mid u) = \frac{1 - \chi_{\mathbf{s}_i,\mathbf{s}_k}(u)}{2\big(3 - \chi_{\mathbf{s}_i,\mathbf{s}_k}(u)\big)}, \tag{S21}$$

which is symmetric, equals zero when $\mathbf{s}_i = \mathbf{s}_k$, and increases as extremal dependence weakens.

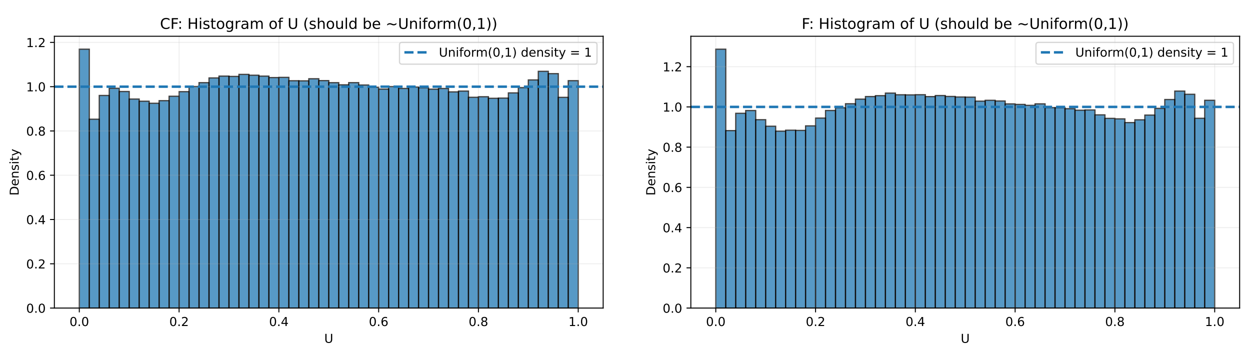
Given the resulting pairwise dissimilarity matrix, we obtain a partition of the spatial domain using a k-medoids algorithm. In particular, we consider partitioning around medoids and its scalable variants, and select the number of clusters k using a silhouette criterion computed from the precomputed dissimilarities. To reduce sensitivity to initialization, the silhouette criterion can be evaluated over multiple random restarts, and a stable choice of k is retained. The final output is a cluster label for each land grid cell and a set of representative medoid locations.

In the application, we fix the exceedance threshold at $u = 0.8$. This choice is necessarily somewhat arbitrary and reflects a classical bias-variance trade-off. Based on this choice, the clustering procedure selects $k = 8$ spatial clusters. The resulting partition of Europe into regions with similar extremal dependence properties is displayed in Figure 2. The clusters we obtained are highly similar to those obtained by Stefanon et al. (2012) and those used by Auld et al. (2023) in his analysis of annual maximum temperatures in Europe.

In the remainder of this work, subsequent analyses focus on Cluster 3. This cluster predominantly covers Western Europe and parts of Central Europe, including regions such as France, the Benelux countries, and neighboring areas.

S2.2 Diagnostic of the BSQR Uniform Transformation

This section provides a diagnostic check of the marginal transformation used in the BSQR framework. Figure S4 displays histograms of the transformed variables $\widehat{U}(\mathbf{s}, d, t) = \hat{F}_{\mathbf{s},t}(X(\mathbf{s}, d, t))$ for both counterfactual and factual settings. Under a correct specification of the marginal model, these transformed values should follow a uniform distribution on $[0, 1]$. The figure confirms that the transformation performs adequately overall, with minor deviations primarily observed in the tails.

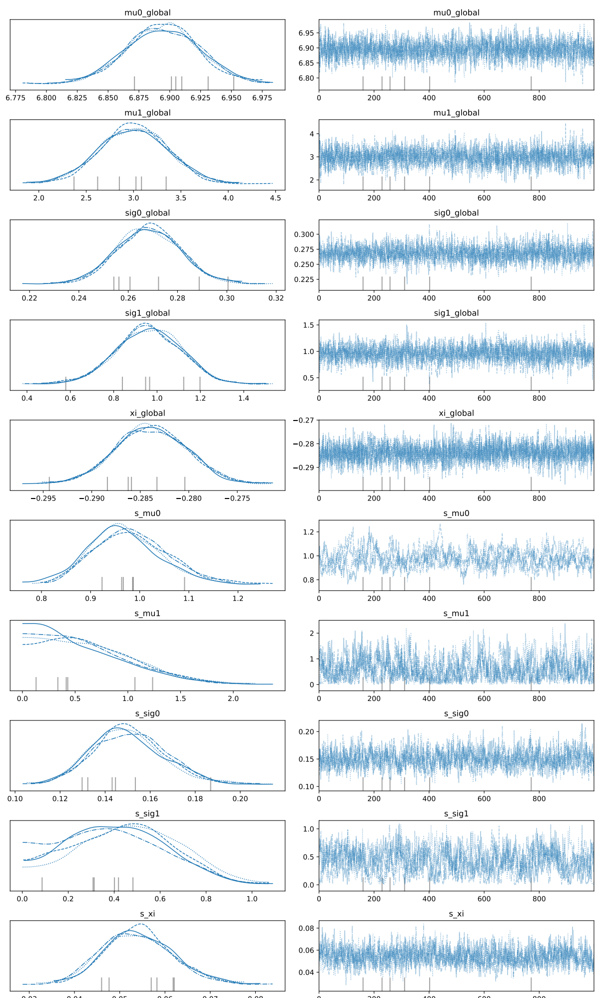


(a) Counterfactual (CF). (b) Factual (F).

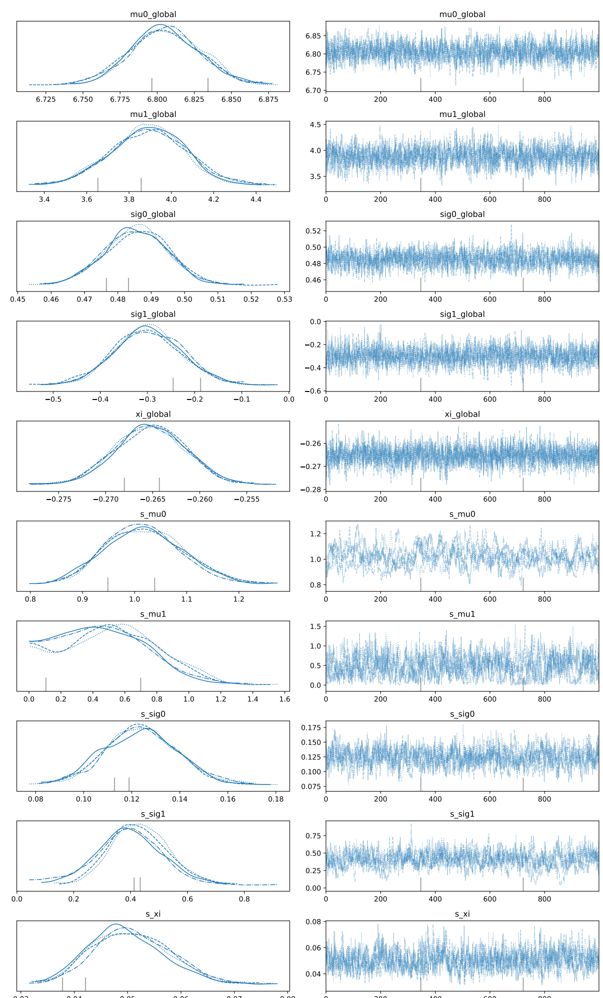
Figure S4: Diagnostic check of the BSQR uniformization step for Cluster 3. Histograms of $\widehat{U}(\mathbf{s}, d, t) = \hat{F}_{\mathbf{s},t}(X(\mathbf{s}, d, t))$ (months JJA) should be approximately uniform on $[0, 1]$ if the marginal nonstationarity is adequately removed. The fit is generally satisfactory, with remaining discrepancies most visible in the tails.

S2.3 MCMC convergence diagnostics

This section presents convergence diagnostics for the Bayesian GEV model. Figure S5 shows trace plots of selected global and spatial hyperparameters under both counterfactual and factual climates. The chains exhibit good mixing and no visible trends, indicating satisfactory convergence of the Markov chain Monte Carlo algorithm.



(a) Counterfactual (CF) climate.



(b) Factual (F) climate.

Figure S5: Trace plots for selected global and spatial hyperparameters of the GEV model under the counterfactual (CF) and factual (F) climates. The chains show satisfactory mixing and no apparent trends, supporting convergence of the Markov chain Monte Carlo algorithm.

S2.4 Posterior predictive checks

We assess the adequacy of the fitted nonstationary spatial GEV model using a posterior predictive Q-Q diagnostic based on standardized residuals. For each grid cell $\mathbf{s}_i$ and year t, recall that

$$M_{\mathbf{s}_i,t} \sim \mathrm{GEV}\big(\mu_{\mathbf{s}_i}(t), \sigma_{\mathbf{s}_i}(t), \xi_{\mathbf{s}_i}\big), \qquad \mu_{\mathbf{s}_i}(t) = \mathbf{c}(t)^\top \boldsymbol{\mu}_{\mathbf{s}_i}, \qquad \log \sigma_{\mathbf{s}_i}(t) = \mathbf{c}(t)^\top \boldsymbol{\sigma}_{\mathbf{s}_i},$$

with $\xi_{\mathbf{s}_i}$ constant over time. Let

$$V_{\mathbf{s}_i,t} = \begin{cases} \dfrac{1}{\xi_{\mathbf{s}_i}} \log\left(1 + \xi_{\mathbf{s}_i} \dfrac{M_{\mathbf{s}_i,t} - \mu_{\mathbf{s}_i}(t)}{\sigma_{\mathbf{s}_i}(t)}\right), & \xi_{\mathbf{s}_i} \neq 0, \\[2ex] \dfrac{M_{\mathbf{s}_i,t} - \mu_{\mathbf{s}_i}(t)}{\sigma_{\mathbf{s}_i}(t)}, & \xi_{\mathbf{s}_i} = 0, \end{cases}$$

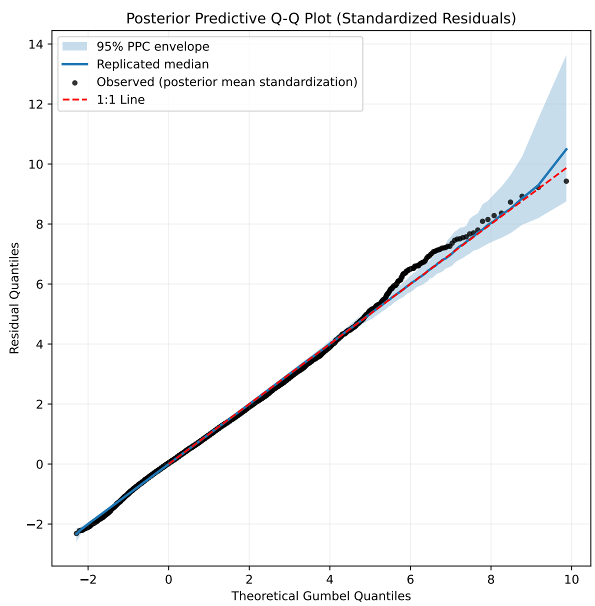
denote the GEV residual transformation, which maps $M_{\mathbf{s}_i,t}$ to a scale where the theoretical quantiles are those of a Gumbel distribution. We pool these transformed residuals over all available $(\mathbf{s}_i, t)$ within Cluster 3 and compare the empirical residual quantiles to theoretical Gumbel quantiles

$$t_q = -\log\{-\log(p_q)\}, \qquad p_q = \frac{q}{N+1}, \quad q = 1, \ldots, N,$$

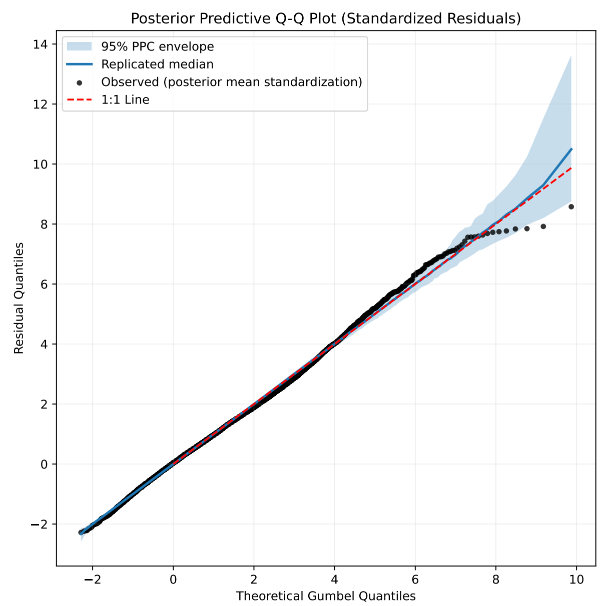
where $N = n_c * T$ is the number of pooled residuals with n_c the number of grid cells in Cluster 3.

To construct a posterior predictive envelope, we draw parameter samples from the posterior distribution and, for each draw, simulate replicated maxima $M^{\mathrm{rep}}_{\mathbf{s}_i,t}$ from the corresponding GEV model. The replicated data are transformed into residuals using the same mapping, yielding a replicated Q-Q curve. Repeating this procedure produces an ensemble of replicated Q-Q curves, from which we compute a pointwise 95% posterior predictive envelope and the posterior median replicated curve.

In Figure S6, the black dots is the Q-Q plot of the observed residuals computed using posterior mean parameter estimates, the blue solid line is the posterior median replicated Q-Q curve, the shaded band is the pointwise posterior predictive envelope, and the red dashed line is the 1:1 reference.



(a) Counterfactual (CF) climate.



(b) Factual (F) climate.

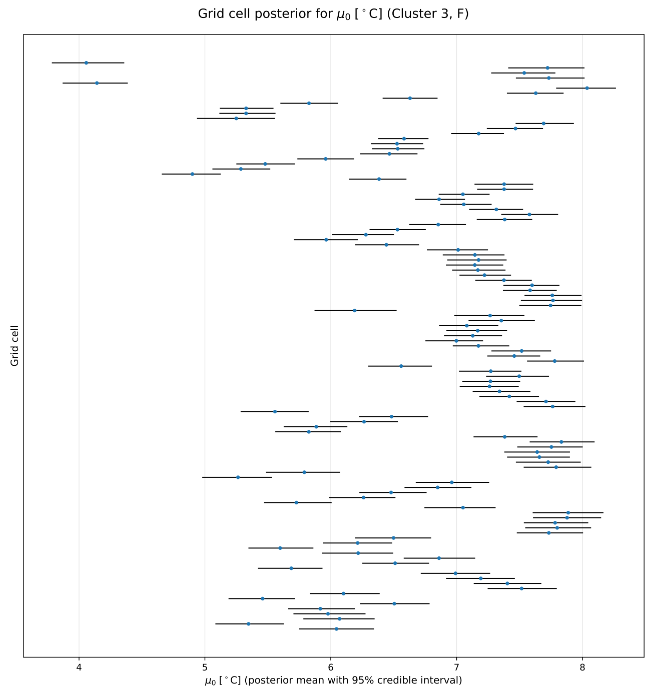
Figure S6: Posterior predictive Q-Q diagnostics for the fitted GEV model based on standardized residuals pooled over Cluster 3. The black curve corresponds to observed residual quantiles (computed using posterior mean parameters), the blue curve is the posterior median replicated Q–Q curve, the shaded region is the pointwise posterior predictive envelope, and the red dashed line is the 1:1 reference.

S2.5 Spatial posterior summaries for baseline GEV parameters

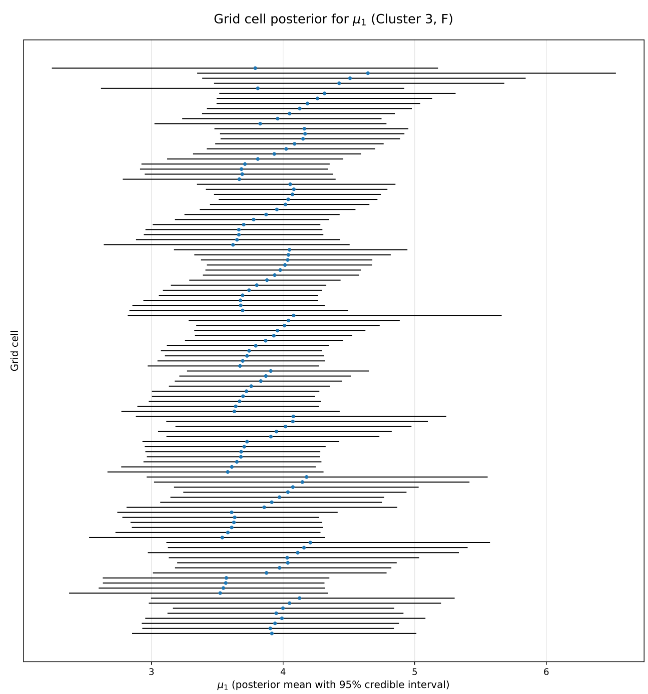
This section presents posterior summaries of the spatially varying GEV parameters under the factual climate for Cluster 3. Figures S7–S9 display caterpillar plots for the location, scale, and shape parameters, respectively. These plots summarize the posterior distributions across grid cells, highlighting spatial variability and associated uncertainty in the estimated GEV coefficients.

S2.6 Diagnostic estimation Neural Bayes estimator

This section presents validation diagnostics for the neural Bayes estimator on a holdout dataset. Figures S10 and S11 compare predicted and true parameter values for Cluster 3. The scatterplots assess agreement between predictions and ground truth, while the error boxplots summarize bias and variability of the estimates. Together, these diagnostics provide an overall evaluation of the predictive performance of the neural network.

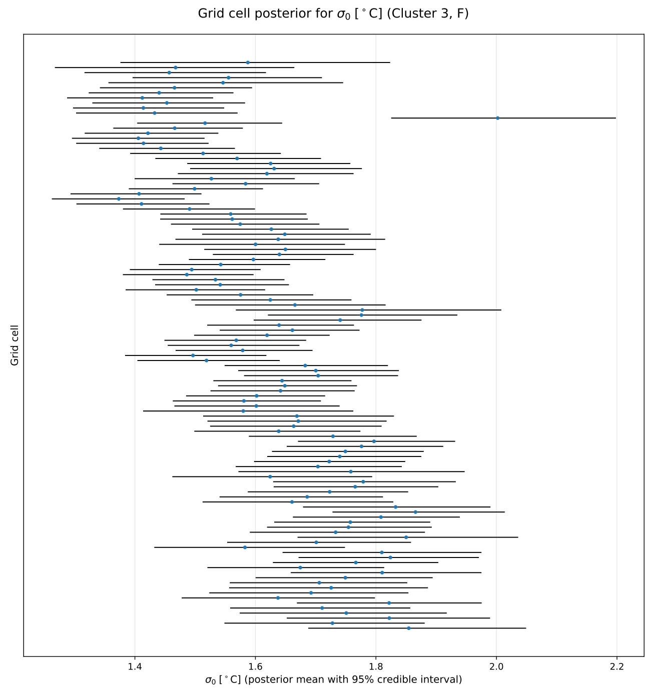


(a) μ_0 [$^\circ$C]

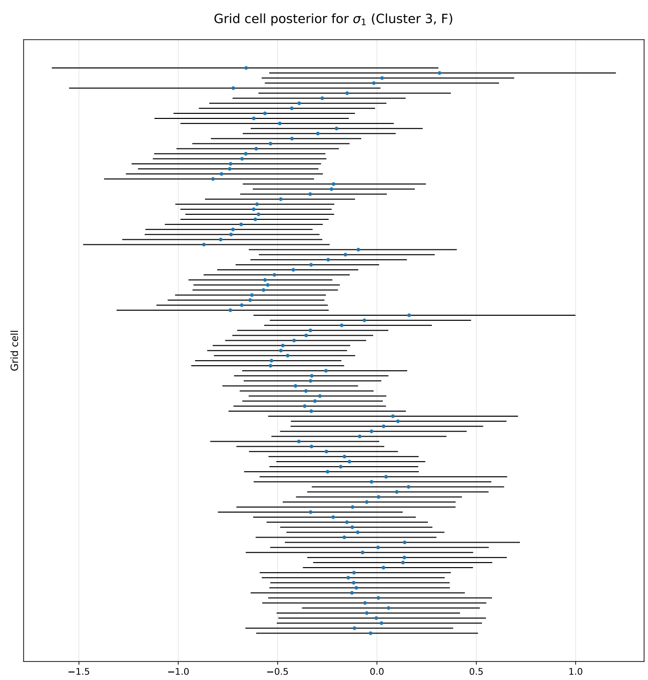


(b) μ_1

Figure S7: Caterpillar plots of the location parameters for Cluster 3 under the factual climate.



(a) σ_0 [°C]



(b) σ_1

Figure S8: Caterpillar plots of the scale parameters for Cluster 3 under the factual climate.

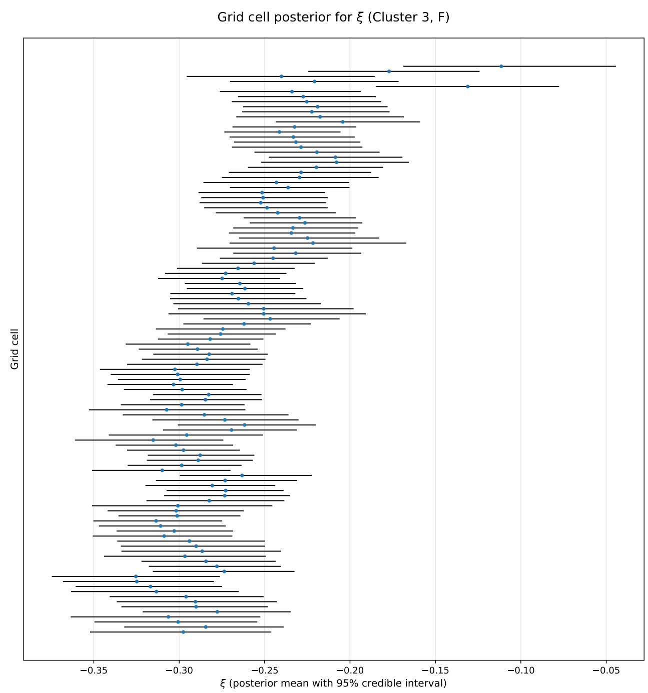
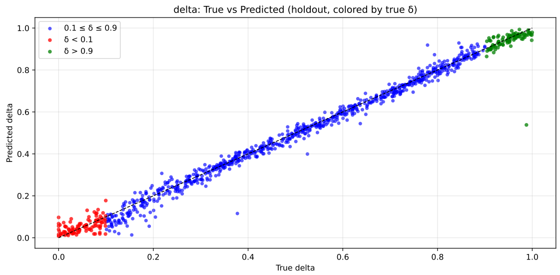
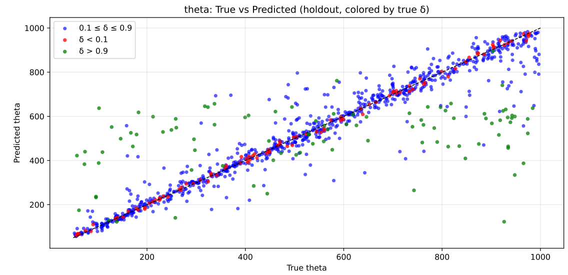


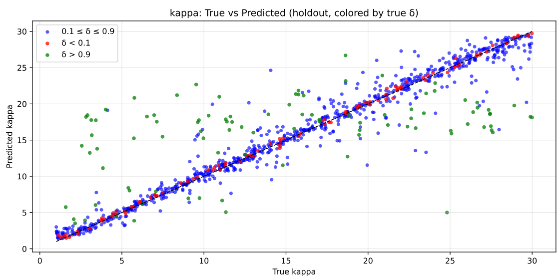
Figure S9: Caterpillar plot of the shape parameter ξ for Cluster 3 under the factual climate.



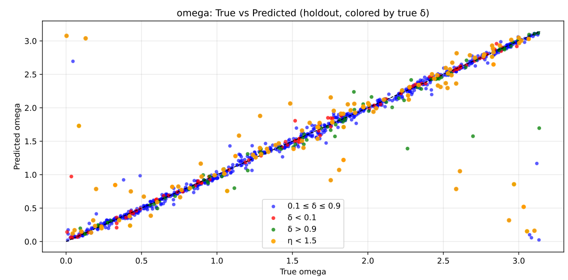
(a) δ



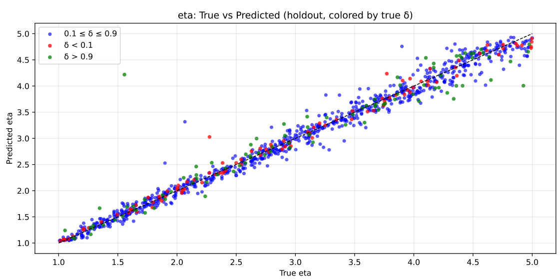
(b) θ



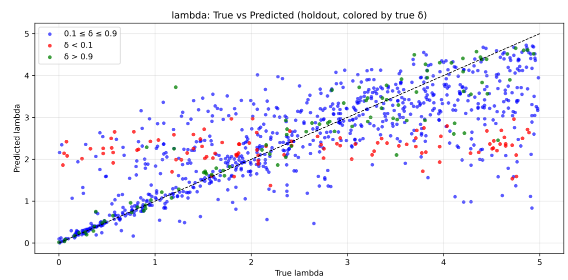
(c) κ



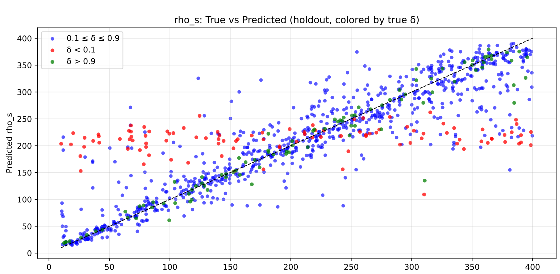
(d) ω



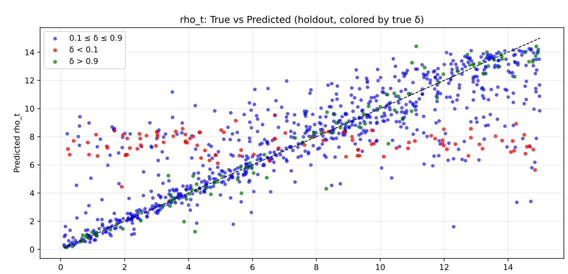
(e) η



(f) λ

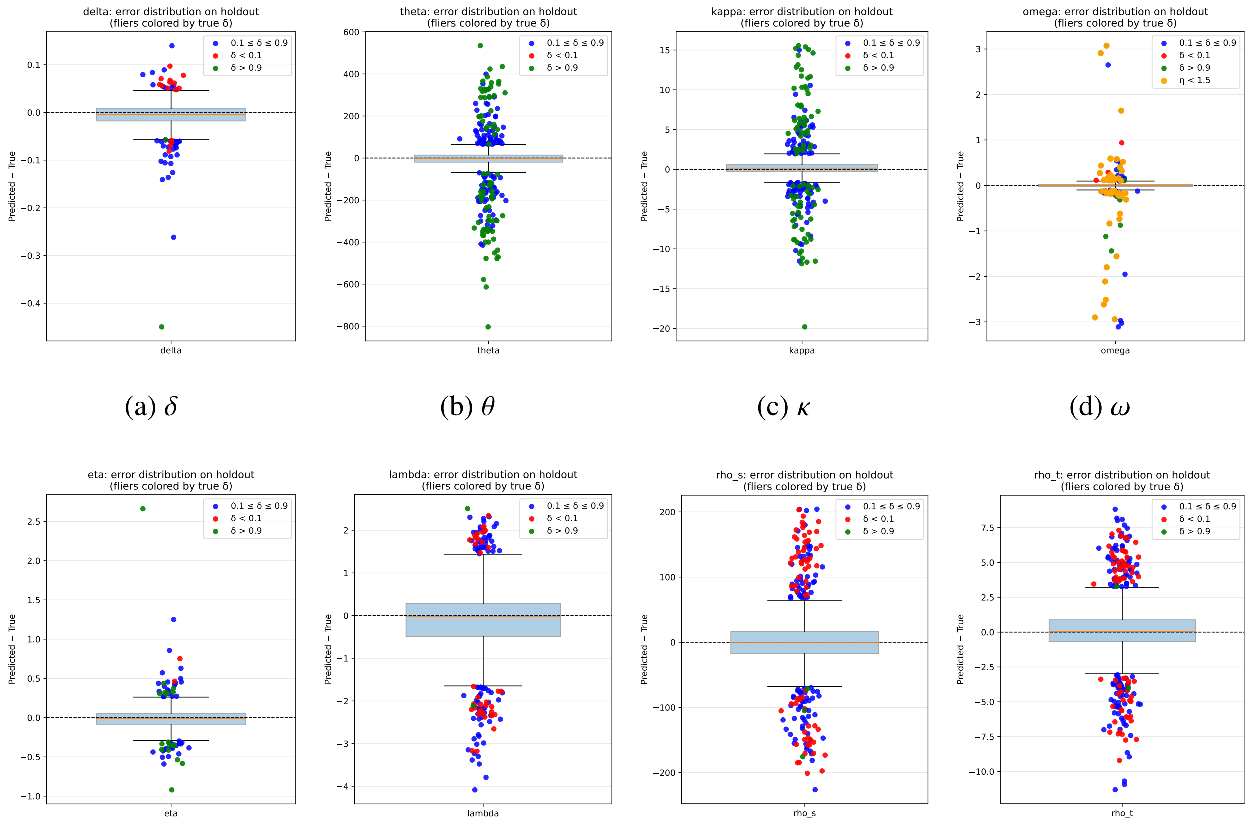


(g) ρ_s



(h) ρ_t

Figure S10: Scatterplots of predicted versus true parameter values on the holdout set for Cluster 3. The dashed line in each panel indicates perfect agreement between prediction and truth.



(a) δ (b) θ (c) κ (d) ω

(e) η (f) λ (g) ρ_s (h) ρ_t

Figure S11: Boxplots of the prediction errors (predicted minus true value) on the holdout set for Cluster 3. Values centered around zero indicate low bias, while the spread reflects estimation uncertainty.