

MODELING EXTREMAL STREAMFLOW USING DEEP LEARNING APPROXIMATIONS AND A FLEXIBLE SPATIAL PROCESS

BY REETAM MAJUMDER^{1,a} , BRIAN J. REICH^{2,b}  AND BENJAMIN A. SHABY^{3,c} 

¹*Southeast Climate Adaptation Science Center, North Carolina State University, rmajumd3@ncsu.edu*

²*Department of Statistics, North Carolina State University, bjreich@ncsu.edu*

³*Department of Statistics, Colorado State University, bshaby@colostate.edu*

Quantifying changes in the probability and magnitude of extreme flooding events is key to mitigating their impacts. While hydrodynamic data are inherently spatially dependent, traditional spatial models, such as Gaussian processes, are poorly suited for modeling extreme events. Spatial extreme value models with more realistic tail dependence characteristics are under active development. They are theoretically justified but give intractable likelihoods, making computation challenging for small datasets and prohibitive for continental-scale studies. We propose a process mixture model (PMM) which specifies spatial dependence in extreme values as a convex combination of a Gaussian process and a max-stable process, yielding desirable tail dependence properties but intractable likelihoods. To address this, we employ a unique computational strategy where a feed-forward neural network is embedded in a density regression model to approximate the conditional distribution at one spatial location, given a set of neighbors. We then use this univariate density function to approximate the joint likelihood for all locations by way of a Vecchia approximation. The PMM is used to analyze changes in annual maximum streamflow within the U.S. over the last 50 years and is able to detect areas which show increases in extreme streamflow over time.

1. Introduction. The Intergovernmental Panel on Climate Change released its Sixth Assessment in 2021 and projected an increased frequency of hydroclimatic extremes. In addition to changes in the mean of climate variables, the impact of climate change is more severe with changes in the frequency and magnitude of hydroclimatic extremes. Floods are responsible for huge economic and human costs (Hirabayashi et al. (2013), Winsemius et al. (2018)), and this cost is projected to increase due to sea level rise and extreme precipitation events brought about by our changing climate (Winsemius et al. (2018)). Effective prediction of future flooding events is required for water infrastructure design but is challenging due to the complexity of flooding events and uncertain climate predictions (Condon, Gangopadhyay and Pruitt (2015), François et al. (2019), Kundzewicz et al. (2017), Merz et al. (2014)). Extensive research has been conducted looking at changing climate signals in historical extreme rainfall (Knox (1993), Kunkel et al. (2020)) and in flooding (Archfield et al. (2016), Blöschl et al. (2019), Franks (2002), Hirsch (2011), Meehl et al. (2000), Milly, Dunne and Vecchia (2005), Sharma, Wasko and Lettenmaier (2018), Vogel, Yaindl and Walter (2011), Walter (2010)). For example, Hirsch and Ryberg (2012) found a significant change in annual maximum streamflow (a key measure of flood risk) at 48 of 200 U.S. Geological Survey (USGS) gauges and spatial clustering in the direction and magnitude of the changes. As a result, there is a need to account for spatial and temporal variability (i.e., nonstationarity) in flood frequency patterns when assessing current and future risk (Kundzewicz et al. (2014), Merz et al. (2014), Milly et al. (2008), Milly et al. (2015), Salas and Obeysekera (2014), Vogel, Yaindl and Walter (2011)).

Received August 2022; revised October 2023.

Key words and phrases. Gaussian process, max-stable process, neural networks, spatial extremes, Vecchia approximation.

One approach to projecting flood risk on the basis of extreme streamflow involves the statistical extrapolation of the spatiotemporal trends observed in the historical record. Of particular interest is estimating the joint probability of extremal streamflow at multiple locations, which is useful for understanding regional flood impacts and assessment to support federal and state emergency management agencies. For example, [Dawdy, Griffis and Gupta \(2012\)](#), [Lima et al. \(2016\)](#), [Sraj et al. \(2016\)](#) use extreme value analysis (EVA) methods to model nonstationarity with regressions or hierarchical models for the relationship between flooding and watershed characteristics and weather. Classic nonspatial EVA ([Coles et al. \(2001\)](#)) begins by isolating the extreme events of interest. This is done systematically by either selecting all exceedances over a threshold or computing the block maximum, for example, the annual maximum of daily streamflow. A spatial EVA analyzes exceedances or pointwise maxima (i.e., computed separately at each spatial location) as a stochastic process over space. Modeling spatial dependence allows for predictions at ungauged locations and the estimation of the joint probability of extremes at multiple locations. It also facilitates the borrowing of information across locations to estimate the marginal distribution at each location, which is particularly useful for EVA where data are sparse and low-probability events are of interest, and gives valid statistical inference for model parameters by properly accounting for spatial dependence.

In this study we consider extreme streamflow data from the United States Geological Survey's Hydro-Climatic Data Network (HCDN) ([Lins \(2012\)](#)). Our primary objective is to identify regions within the U.S. where the distribution of extreme streamflow has changed over time. The HCDN has a long historical record and consists of locations that are minimally impacted by anthropogenic activity while excluding sites where human activities affect the flow of the watercourse. We focus on the modeling of block maxima of streamflow with the help of the max-stable process (MSP) ([de Haan and Ferreira \(2006\)](#)). MSPs are a limiting class of models for spatial extremes, featuring strong forms of tail dependence ([Kabluchko, Schlather and de Haan \(2009\)](#), [Reich and Shaby \(2012b\)](#), [Schlather \(2002\)](#), [Smith \(1990\)](#), [Tawn \(1990\)](#), [Wadsworth and Tawn \(2012b\)](#)). They are a natural asymptotic model for block maxima but can also be applied to peaks over a threshold using a censored likelihood (e.g., [Huser and Davison \(2014\)](#), [Reich, Shaby and Cooley \(2013\)](#)).

In practice, MSPs pose two challenges. First, the analytic forms of (censored) MSP densities are computationally intractable for all but a small number of spatial locations ([Kabluchko, Schlather and de Haan \(2009\)](#), [Schlather \(2002\)](#), [Wadsworth \(2015\)](#), [Wadsworth and Tawn \(2012b\)](#), [Wadsworth and Tawn \(2014\)](#)). For general MSPs, [Castruccio, Huser and Genton \(2016\)](#) stated that full likelihood inference seemed limited to $n = 13$ locations. These low-dimensional results have led to the use of composite likelihood (CL) approximations ([Padoan, Ribatet and Sisson \(2010b\)](#)). However, CL suffers from statistical inefficiency for large n ([Huser, Davison and Genton \(2016\)](#)), finite-sample bias when using all pairs of observations ([Castruccio, Huser and Genton \(2016\)](#), [Sang and Genton \(2014b\)](#), [Wadsworth \(2015\)](#)), and computational challenges posed by computing likelihoods at all $O(n^2)$ pairs. More recently, [Huser et al. \(2019\)](#) proposed an expectation-maximization algorithm for full likelihood inference, with computation time of 19.8 hours with $n = 20$. [Huser, Stein and Zhong \(2022\)](#) have also applied the Vecchia approximation that requires only moderate-dimensional (say 10 or 15) joint distribution functions, which are available for some MSPs. Deep learning has also been used to estimate parameters in spatial models within a framework of simulation-based inference ([Gerber and Nychka \(2021\)](#), [Lenzi et al. \(2021\)](#), [Sainsbury-Dale, Zammit-Mangion and Huser \(2023\)](#)). They leverage a likelihood-free approach by simulating datasets with different parameter values and using deep learning to identify features of the simulated data that are indicative of particular parameter values. [Gerber and Nychka \(2021\)](#) used it to estimate covariance parameters for spatial Gaussian process (GP) models by training convolutional

neural networks (CNNs) to take moderate size spatial fields or variograms as input and return the range and noise-to-signal covariance parameters as output. [Lenzi et al. \(2021\)](#) used simulated data as input and trained CNNs to learn the parameters of an MSP. Finally, [Sainsbury-Dale, Zammit-Mangion and Huser \(2023\)](#) have used permutation invariant neural networks for large spatial extremes datasets in a Bayesian setting for estimating parameters from independent replicates. However, it is difficult to extend them to problems with large numbers of parameters; for example, a crucial assumption in our application is that the marginal distributions have spatiotemporally varying coefficients (STVC), which substantially expands the parameter space. Bayesian approaches have also been proposed since they provide stability by incorporating prior information as available and are often preferred for uncertainty quantification. But they are restricted to either small n ([Ribatet, Cooley and Davison \(2012\)](#)) or very specific models ([Bopp, Shaby and Huser \(2021\)](#), [Morris, Reich and Thibaud \(2019\)](#), [Reich and Shaby \(2012b\)](#)). For lower dimensional problems, approximate Bayesian computation (ABC) can replace likelihood evaluation with repeated simulation from the MSP model ([Erhardt and Smith \(2012\)](#)). For general intractable likelihood estimation problems, neural networks can be leveraged for conditional density estimation. For example, [Greenberg, Nonnenmacher and Macke \(2019\)](#) use a sequential neural posterior estimation method for simulation-based inference. A related method is normalizing flows ([Kobyzev, Prince and Brubaker \(2021\)](#), [Papamakarios et al. \(2021\)](#)), where a simple density is pushed through a series of transformations, often involving neural networks, to obtain more complex densities.

A second challenge posed by MSPs is that they are restrictive in the class of dependence types they can incorporate. Environmental data often have weakening spatial dependence with increasing levels of extreme quantiles, as we go farther out into the tails of the distributions; however, MSPs are unable to accommodate this behavior. [Wadsworth and Tawn \(2012b\)](#) addressed this with a max-mixture model that took an MSP and incorporated asymptotic independence at the boundary point of the parameter space using a mixing parameter. A more general approach was taken in [Huser and Wadsworth \(2019\)](#), which combined a Pareto random variable with a GP resulting in a hybrid model which interpolates between perfect dependence and asymptotic independence, indexed similarly by a mixing parameter. This flexible model can establish asymptotic dependence or asymptotic independence from the data without needing a prior assumption. A limitation of this model is that the Pareto random variable is shared by the spatial locations, inducing dependence between distant sites. This might be unrealistic for an analysis over a large spatial domain. Finally, [Hazra, Huser and Bolin \(2021\)](#) consider a mixture of a GP with a stochastic scale process; it can capture a range of extremal dependence structures but does not employ a mixing parameter and, therefore, assumes equal contribution from both its constituent processes.

In this paper we propose a spatial EVA model and an associated computational algorithm to address the aforementioned limitations of the MSP and related approaches. The EVA model is specified as a convex combination of an MSP and a GP for residual dependency and has generalized extreme value (GEV) distributed margins with STVC. We refer to it as the process mixture model (PMM). From a modeling perspective, the mixture of the two spatial processes allows asymptotic dependence or independence for locations separated by distance h with independence as $h \rightarrow \infty$ (long-range independence). Furthermore, the STVC can account for temporal nonstationarity, which is key for large-scale climate studies. This flexibility comes at a computational cost: the model has hundreds of parameters and even bivariate PDFs do not have a closed form to the best of our knowledge. Therefore, we develop a new computational algorithm that uses a feed-forward neural network (FFNN) embedded in a density regression model ([Xu and Reich \(2021\)](#)) to approximate the conditional distribution at one spatial location given a set of neighbors. Following this, the univariate density functions are used to approximate the joint likelihood for all locations by means of a

Vecchia approximation (Vecchia (1988b)). This specification partitions the parameter space into a low-dimensional vector of spatial dependence parameters and a higher dimensional vector of marginal parameters and decouples the likelihood approximation from parameter estimation. The FFNN is trained on synthetic data generated from a design distribution using different parameter values; this allows us to avoid data scarcity issues and accommodate a range of marginal densities. Parameter estimation is carried out using Markov chain Monte Carlo (MCMC) sampling. This computational framework is quite general. Unlike many of the approaches mentioned above, it can be applied to virtually any spatial process (e.g., GP, MSP, and mixtures), can accommodate high-dimensional STVC margins, as well as missing and censored data. We use the PMM to analyze changes in annual maximum streamflow within the U.S. over the past 50 years.

The rest of this paper is organized as follows. Section 2 provides background on the construction and dependence measures for MSPs. Section 3 introduces the PMM for block maxima. Section 4 describes inference for the PMM which employs a deep learning Vecchia approximated density regression approach. Section 5 consists of a detailed simulation study demonstrating the method. Section 6 analyzes annual streamflow maxima data for HCDN stations across the U.S. and identifies changes in their behavior over the past 50 years. Section 7 concludes with a discussion. Additional theoretical details, simulation studies, and results from our application are provided in the Supplementary Material (Majumder, Reich and Shaby (2024)).

2. Background.

2.1. The max-stable process. A random process $\{R(\mathbf{s}) : \mathbf{s} \in \mathcal{S} \subset \mathbb{R}^d\}$, indexed by spatial locations $\mathbf{s}$, is called max-stable if there exists a sequence $\{X_i(\mathbf{s}) : i \in \mathbb{N}\}$ of independent copies of the process $\{X(\mathbf{s}) : \mathbf{s} \in \mathcal{S}\}$ and normalizing functions $a_n(\mathbf{s}) > 0$, $b_n(\mathbf{s}) \in \mathbb{R}$ such that

$$R(\mathbf{s}) \stackrel{d}{=} \frac{\max_{i=1:n} X_i(\mathbf{s}) - b_n(\mathbf{s})}{a_n(\mathbf{s})}.$$

Further, it can be shown that if there exist continuous functions $c_n(\mathbf{s}) > 0$, $d_n(\mathbf{s}) \in \mathbb{R}$ such that, as $n \rightarrow \infty$,

$$\frac{\max_{i=1:n} X_i(\mathbf{s}) - d_n(\mathbf{s})}{c_n(\mathbf{s})} \rightarrow R(\mathbf{s}),$$

then $R(\mathbf{s})$ is either degenerate or an MSP (de Haan (1984)). If $R(\mathbf{s})$ is nondegenerate, the pointwise distributions of $R(\mathbf{s})$ are in the GEV family (de Haan and Ferreira (2006)).

Max-stable processes arise as the pointwise maxima taken over an infinite number of appropriately rescaled stochastic processes and are, therefore, widely applied for studying spatial extremes, in particular block maxima (with a block size of n). MSPs can be constructed through a spectral representation (de Haan (1984), Penrose (1992)). Let $\{Z_i : i \in \mathbb{N}\}$ be the points of a Poisson process on $(0, \infty)$ with intensity $1/z^2 dz$. Then there exists a nonnegative stochastic process $W(\mathbf{s})$ with continuous sample paths and with $\mathbb{E}W(\mathbf{s}) = 1$ for all $\mathbf{s} \in \mathcal{S}$ such that

$$R(\mathbf{s}) \stackrel{d}{=} \max_{i \geq 1} Z_i W_i(\mathbf{s}),$$

where $W_i(\mathbf{s})$ are independent copies of $W(\mathbf{s})$, and $R(\mathbf{s})$ is max-stable with unit Fréchet marginal distributions. Common parametric subclasses of MSPs include mixed moving maxima processes (Wang and Stoev (2010)), the Schlather processes (Schlather (2002)), and Brown–Resnick processes (Brown and Resnick (1977), Kabluchko, Schlather and de Haan (2009)).

The finite dimensional distribution of an MSP $R(\mathbf{s})$ at a set of locations $(\mathbf{s}_1, \dots, \mathbf{s}_k) \in \mathcal{S}$ has the form $\Pr\{R(\mathbf{s}_j) < r_j, j = 1 : k\} = \exp\{-\Lambda(r_1, \dots, r_k)\}$, where Λ is known as the exponent function and is given by

$$\Lambda(r_1, \dots, r_k) = \mathbb{E} \left[\max_{j=1:k} \frac{W(\mathbf{s}_j)}{r_j} \right].$$

2.2. Dependence properties. Let F_1 and F_2 be the cumulative distribution functions (CDFs) of $R(\mathbf{s}_1)$ and $R(\mathbf{s}_2)$, and let $U(\mathbf{s}_i) = F_i(R(\mathbf{s}_i))$ for $i = 1, 2$. The joint tail behavior of the two random variables $U(\mathbf{s}_1)$ and $U(\mathbf{s}_2)$ with uniform marginals can be studied in terms of the conditional exceedance probability, given by

$$\chi_u(\mathbf{s}_1, \mathbf{s}_2) := \Pr[U(\mathbf{s}_1) > u | U(\mathbf{s}_2) > u] \in (0, 1),$$

where $u \in (0, 1)$ is a threshold. A commonly used measure of extremal dependence is the upper-tail coefficient (Joe (1997)), defined as

$$\chi(\mathbf{s}_1, \mathbf{s}_2) = \lim_{u \rightarrow 1} \chi_u(\mathbf{s}_1, \mathbf{s}_2).$$

The random variables $U(\mathbf{s}_1)$ and $U(\mathbf{s}_2)$ are considered *asymptotically dependent* if the upper-tail coefficient is strictly positive and *asymptotically independent* if it is zero.

We note that asymptotic (in)dependence is different from complete (in)dependence, since asymptotic (in)dependence is determined specifically by the joint behavior of the tails of the distribution as $u \rightarrow 1$. Asymptotic independence is also different from long-range independence which is determined by the asymptotic behavior of $\chi_u(\mathbf{s}_1, \mathbf{s}_2)$ as $\|\mathbf{s}_1 - \mathbf{s}_2\| \rightarrow \infty$.

In the asymptotic independence scenario, the coefficient of tail dependence proposed by Ledford and Tawn (1996), Ledford and Tawn (1997) is useful to study the joint tail behavior of the process. Consider $R(\mathbf{s})$ with unit Fréchet margins. The joint survivor function of $R(\mathbf{s}_1)$ and $R(\mathbf{s}_2)$ is assumed to have the expression

$$\bar{F}(r, r) := \Pr[R(\mathbf{s}_1) > r, R(\mathbf{s}_2) > r] \sim \mathcal{L}(r)r^{-1/\eta}, \quad \text{as } r \rightarrow \infty,$$

where $\mathcal{L}$ is a slowly varying function that satisfies $\mathcal{L}(tr)/\mathcal{L}(r) \rightarrow 1$ as $r \rightarrow \infty$ for all fixed $t > 0$, and $\eta \in (0, 1]$ is a constant that effectively determines the decay rate of $\bar{F}(r, r)$ for large r . The parameter η is known as the coefficient of tail dependence. A value of $\eta = 1/2$ indicates independent marginal variables; values lower and higher than $1/2$ correspond to a negative and a positive association respectively between the pair of variables.

2.3. The Brown–Resnick process. Consider $W(s) = \exp\{\epsilon(s) - \gamma(s)\}$ in the spectral representation of an MSP, where $\epsilon(s)$ is an intrinsically stationary Gaussian process with semi-variogram $\gamma(\cdot)$, and $\epsilon(0) = 0$ almost surely. $W(s)$ is continuous and nonnegative. Then $R(s)$ is a strictly stationary MSP known as the Brown–Resnick process, whose distribution depends only on $\gamma(\cdot)$. The Brown–Resnick process is attractive since a wide range of variograms can be used with them, and they are relatively easily simulated. The exponent function that defines the joint distribution for the pair $R(\mathbf{s}_1)$ and $R(\mathbf{s}_2)$ is

$$\Lambda(r_1, r_2) = \frac{1}{r_1} \Phi \left\{ \frac{a}{2} - \frac{1}{a} \log \left(\frac{r_1}{r_2} \right) \right\} + \frac{1}{r_2} \Phi \left\{ \frac{a}{2} - \frac{1}{a} \log \left(\frac{r_2}{r_1} \right) \right\},$$

where $a = \{2\gamma(\mathbf{s}_1 - \mathbf{s}_2)\}^{1/2}$, and $\Phi(\cdot)$ denotes the standard normal distribution function.

3. A process mixture model for spatial extremes. Let $\{Y(\mathbf{s}); \mathbf{s} \in \mathcal{S}\}$ be a spatial extremes process indexed by the set $\mathcal{S} \subset \mathbb{R}^2$. In this section we consider $Y(\mathbf{s})$ to be defined as a block maximum, but the methods can be extended to peaks over a threshold. We assume a potentially different marginal distribution for each spatial location $\mathbf{s}$ and denote $F_{\mathbf{s}}$ as the marginal CDF for site $\mathbf{s}$. For example, we assume that $F_{\mathbf{s}}$ is the GEV distribution with location $\mu(\mathbf{s}) \in \mathbb{R}$, scale $\sigma(\mathbf{s}) > 0$, and shape $\xi(\mathbf{s}) \in \mathbb{R}$ so that marginally

$$Y(\mathbf{s}) \sim \text{GEV}\{\mu(\mathbf{s}), \sigma(\mathbf{s}), \xi(\mathbf{s})\}.$$

Its CDF $F_{\mathbf{s}}(y|\mu(\mathbf{s}), \sigma(\mathbf{s}), \xi(\mathbf{s})) := \Pr[Y(\mathbf{s}) < y]$ is

$$(1) \quad F_{\mathbf{s}}(y|\mu(\mathbf{s}), \sigma(\mathbf{s}), \xi(\mathbf{s})) = \begin{cases} \exp\left[-\left\{1 + \xi(\mathbf{s})\left(\frac{y - \mu(\mathbf{s})}{\sigma(\mathbf{s})}\right)\right\}_+^{-1/\xi(\mathbf{s})}\right], & \xi(\mathbf{s}) \neq 0, \\ \exp\left\{-\exp\left(\frac{y - \mu(\mathbf{s})}{\sigma(\mathbf{s})}\right)\right\}, & \xi(\mathbf{s}) = 0, \end{cases}$$

with $\{y\}_+ := \max(0, y)$, and support over the set $\{y : 1 + \xi(\mathbf{s})(y - \mu(\mathbf{s}))/\sigma(\mathbf{s}) > 0\}$ for the CDF. The shape parameter $\xi(\mathbf{s})$ controls the lower and upper bounds of the distribution; the GEV distribution is bounded above for $\xi(\mathbf{s}) < 0$ and bounded below for $\xi(\mathbf{s}) > 0$. Therefore, the transformed variables

$$(2) \quad U(\mathbf{s}) = F_{\mathbf{s}}\{Y(\mathbf{s})\}$$

share common uniform marginal distributions across the spatial domain. This transformation separates residual spatial dependence in $U(\mathbf{s})$ from the spatial dependence induced by spatial variation in the GEV parameters, which we model using GP priors over $\mathbf{s}$. We note that although we describe the marginal and residual models separately, we fit a joint hierarchical model to simultaneously estimate all model parameters.

We define our spatial dependence model on the residual model $U(\mathbf{s})$ by taking $U(\mathbf{s}) = G\{V(\mathbf{s})\}$ such that

$$(3) \quad V(\mathbf{s}) = \delta R(\mathbf{s}) + (1 - \delta)W(\mathbf{s}),$$

where $R(\mathbf{s})$ and $W(\mathbf{s})$ are an MSP and a GP, respectively, both transformed to have standard exponential margins, and $\delta \in [0, 1]$ is the weight parameter to control relative contribution of the two spatial processes. Mixing the asymptotically dependent MSP with the asymptotically independent GP provides a rich model for spatial dependence. This generalizes the extremal process of [Huser and Wadsworth \(2019\)](#), who assumed a standard exponential random variable R common to all locations, by replacing it with an MSP. Since (3) mixes two processes, we refer to it as the process mixture model (PMM). Further details regarding the transformations required to obtain standard exponential margins and the Huser–Wadsworth model are provided in the Supplementary Material ([Majumder, Reich and Shaby \(2024\)](#), Appendix A.2–A.3).

Let $G(\cdot)$ denote the CDF of the marginal distribution of $V(\mathbf{s})$. By construction, $V(\mathbf{s})$ marginally follows the two-parameter hypoexponential distribution, and its CDF has the following functional form:

$$(4) \quad G(v) = 1 - \frac{1 - \delta}{1 - 2\delta} \exp\left\{-\frac{v}{1 - \delta}\right\} + \frac{\delta}{1 - 2\delta} \exp\left\{-\frac{v}{\delta}\right\},$$

where $\delta \in (0, 1)$, $\delta \neq 1/2$, and $v > 0$. When $\delta = 1/2$, the marginal distribution of $V(\mathbf{s})$ is Gamma(2, 2) with a rate parameter of 2. Although other options are possible, we model the correlation of $W(\mathbf{s})$ using the isotropic powered-exponential correlation function $\text{Cor}\{W(\mathbf{s}_1), W(\mathbf{s}_2)\} = \exp\{-(h/\rho_W)^{\alpha_W}\}$ with distance $h = \|\mathbf{s}_1 - \mathbf{s}_2\|$ measured as the L_2 norm, smoothness $\alpha_W \in (0, 2]$, and range $\rho_W > 0$. $R(\mathbf{s})$ is assumed to be an isotropic Brown–Resnick process with the variogram $\gamma(h) = (h/\rho_R)^{\alpha_R}$, for smoothness $\alpha_R \in (0, 2]$ and range

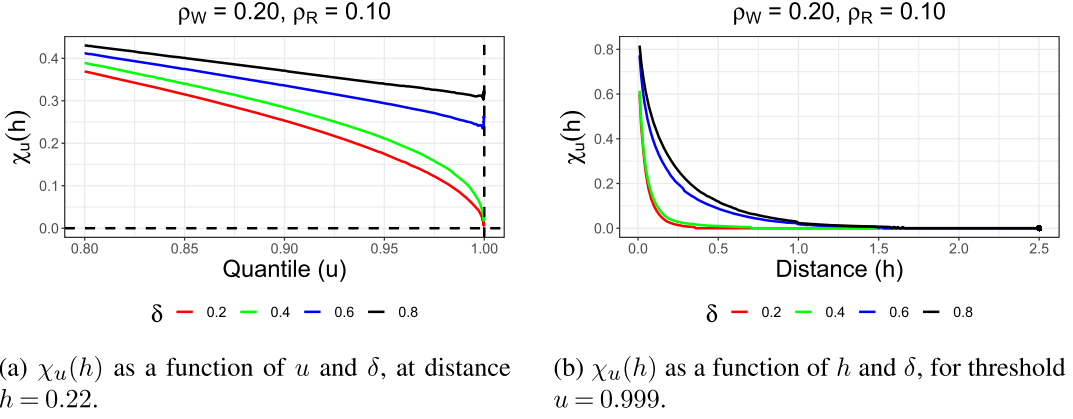


FIG. 1. Behavior of the empirical conditional exceedance: Approximate $\chi_u(h)$ for the PMM plotted as a function of threshold u , distance h , and asymptotic dependence parameter δ . Smoothness parameters $\alpha_W = \alpha_R = 1$, GP range $\rho_W = 0.2$, and MSP range $\rho_R = 0.1$ are fixed for both plots.

$\rho_R > 0$. We also incorporate a nugget into the process mixture. Denoting the proportion of the variance explained by the spatial process by r , we construct $W(\mathbf{s})$ and $R(\mathbf{s})$ to satisfy

$$\begin{aligned} \text{Cor}(W(\mathbf{s}_1), W(\mathbf{s}_2)) &= r \cdot \exp\{-(h/\rho_W)^{\alpha_W}\}, \\ R(\mathbf{s}) &= \max\{r \cdot R_1(\mathbf{s}), (1 - r) \cdot R_2(\mathbf{s})\}, \end{aligned}$$

where $R_1(\mathbf{s})$ is an MSP, and $R_2(\mathbf{s}) \stackrel{\text{iid}}{\sim} \text{GEV}(1, 1, 1)$ is distributed independently of $R_1(\mathbf{s})$. Since $R(\mathbf{s})$ and $W(\mathbf{s})$ are assumed to be isotropic processes, going forward, we can rewrite $\chi_u(\mathbf{s}_1, \mathbf{s}_2)$ and $\chi(\mathbf{s}_1, \mathbf{s}_2)$ as functions of the distance between locations, $\chi_u(h)$ and $\chi(h)$, respectively.

Figure 1 plots Monte Carlo approximations of $\chi_u(h)$ as a function of u and h for the PMM. For these plots we fix $\rho_R = 0.1$, $\rho_W = 0.2$, $\alpha_R = \alpha_W = 1$, and $\delta \in \{0.2, 0.4, 0.6, 0.8\}$. Figure 1(a) sets the correlation to 0.4 by fixing $h = 0.22$ and plots $\chi_u(h)$ as a function of the threshold u . The limit is zero for $\delta < 0.5$ and positive for $\delta > 0.5$. For small values of h , $R(\mathbf{s})$ is approximately the same for both sites (i.e., $R(\mathbf{s}_1) \approx R(\mathbf{s}_2) = R$) and thus the univariate R result of Huser and Wadsworth (2019) that the process is asymptotically dependent if and only if $\delta > 0.5$ emerges. An analytical expression for $\chi_u(h)$ in this special case is provided in Majumder, Reich and Shaby ((2024), Appendix A.4). From Figure 1(b) we see that as the distance h increases, $\chi_u(h)$ converges to zero for all δ because both $R(\mathbf{s})$ and $W(\mathbf{s})$ have diminishing spatial dependence for long distances. The rate of convergence of $\chi_u(h)$ to zero also depends on the value of δ with much slower convergence when $\delta > 0.5$. We note that $\chi_u(h)$ does not converge to zero for large h under the common R model of Huser and Wadsworth (2019), which is unrealistic for studies on a large spatial domain. Additional plots for different values of ρ_R and ρ_W are provided in Majumder, Reich and Shaby ((2024), Appendix A.5).

4. Deep learning Vecchia approximation for the process mixture model. Fitting the PMM introduced in Section 3 poses computational challenges, especially for large datasets. The joint distribution for $W(\mathbf{s})$ is available in closed form but is cumbersome for large datasets; the joint distribution of $R(\mathbf{s})$ is available only for a moderate number of spatial locations, and the joint distribution of the mixture model is more complicated than either of its components. An alternative is to build a surrogate likelihood for Bayesian computation (e.g., Järvenpää et al. (2021), Li et al. (2019), Price et al. (2018), Rasmussen (2003), Wang

and Li (2018b), Wilkinson (2014)). Below we develop a surrogate likelihood based on a Vecchia decomposition (Vecchia (1988b)) and deep learning density regression.

Assume the process is observed at n locations $\mathbf{s}_1, \dots, \mathbf{s}_n$. Partition the parameters into those that affect the marginal distributions, denoted $\boldsymbol{\theta}^{\text{MARG}}$, and those that affect the spatial dependence, denoted $\boldsymbol{\theta}^{\text{SPAT}}$. For the model in Section 3, $\boldsymbol{\theta}^{\text{MARG}}$ includes the GEV parameters $\boldsymbol{\theta}^{\text{MARG}} = \{\mu(\mathbf{s}_i), \sigma(\mathbf{s}_i), \xi(\mathbf{s}_i); i = 1, \dots, n\}$ and $\boldsymbol{\theta}^{\text{SPAT}} = \{\delta, \rho_R, \alpha_R, \rho_W, \alpha_W\}$. Let $Y(\mathbf{s}_i) \equiv Y_i$ and $U_i = F(Y_i; \boldsymbol{\theta}^{\text{MARG}})$ be the transformation of the response so that the distribution of $U_i \in [0, 1]$ does not depend on $\boldsymbol{\theta}^{\text{MARG}}$. We approximate the spatial model on this scale and use the standard change of variables formula to define the joint likelihood on the original scale

$$(5) \quad f_y(y_1, \dots, y_n; \boldsymbol{\theta}^{\text{MARG}}, \boldsymbol{\theta}^{\text{SPAT}}) = f_u(u_1, \dots, u_n; \boldsymbol{\theta}^{\text{SPAT}}) \prod_{i=1}^n \left| \frac{dF(y_i; \boldsymbol{\theta}^{\text{MARG}})}{dy_i} \right|,$$

where $f_y(\cdot)$ and $f_u(\cdot)$ are the joint density functions of $Y_1, \dots, Y_n$ and $U_1, \dots, U_n$, respectively.

We approximate the joint likelihood in (5) using a Vecchia approximation (Datta et al. (2016), Katzfuss and Guinness (2021), Stein, Chi and Welty (2004b), Vecchia (1988b)),

$$(6) \quad f_u(u_1, \dots, u_n; \boldsymbol{\theta}^{\text{SPAT}}) = \prod_{i=1}^n f(u_i | \boldsymbol{\theta}^{\text{SPAT}}, u_1, \dots, u_{i-1}) \approx \prod_{i=1}^n f_i(u_i | \boldsymbol{\theta}^{\text{SPAT}}, u_{(i)})$$

for $u_{(i)} = \{u_j; j \in \mathcal{N}_i\}$ and $\mathcal{N}_i \subseteq \{1, \dots, i-1\}$, for example, the m locations in $\mathcal{N}_i$ that are closest to $\mathbf{s}_i$. The set of locations $\mathbf{s}_{(i)}$ are analogously defined as $\mathbf{s}_{(i)} = \{\mathbf{s}_j; j \in \mathcal{N}_i\}$; the set is referred to as the Vecchia neighboring set and its members as the Vecchia neighbors of location $\mathbf{s}_i$. Of course, not all locations that are dependent on location $\mathbf{s}_i$ need be included in $\mathcal{N}_i$ because distant observations may be approximately independent after conditioning on more local observations. The approximation, therefore, entails truncating the dependence that u_i has on all its previous $i-1$ ordered sites to instead consider dependence on only up to m sites, that is, $|\mathcal{N}_i| \leq m$. The first term on the right-hand side of (6) is the marginal density $f_1(u_1)$. Different choices are possible for ordering the locations prior to the Vecchia approximation (Guinness (2018)). In our work the spatial locations are scaled to be on the unit square and ordered by their distance from the origin.

The conditional distributions for the PMM do not have closed-form expressions. We consider two related approximations—the local approximation, where individual conditional density functions $f_i(\cdot)$ are fit for each location, and a global approximation where a single conditional density function $f(\cdot)$ is estimated for all locations $\mathbf{s}_i, i = 2, \dots, n$. We primarily focus on the local SPQR and present it in this section. Details of the global SPQR approach can be found in the Supplementary Material (Majumder, Reich and Shaby (2024), Appendix A.6).

For the local SPQR approximation at location $\mathbf{s}_i$, we fit a density regression viewing $u_{(i)}$ and $\boldsymbol{\theta}^{\text{SPAT}}$ as features (covariates), denoted by $\mathbf{x}_i$. We approximate the univariate conditional densities for density regression using the model introduced in Xu and Reich (2021),

$$(7) \quad f_i(u_i | \mathbf{x}_i, \mathcal{W}) = \sum_{k=1}^K \pi_{ik}(\mathbf{x}_i, \mathcal{W}_i) B_k(u_i),$$

for $i = 2, \dots, n$, where $\pi_{ik}(\mathbf{x}_i, \mathcal{W}_i) \geq 0$ are probability weights with $\sum_{k=1}^K \pi_{ik}(\mathbf{x}_i) = 1$ that depend on the parameters $\mathcal{W}_i$ and $B_k(u_i) \geq 0$ are M-spline basis functions that, by definition, satisfy $\int B_k(u) du = 1$ for all k . By increasing the number of basis functions K and appropriately selecting the weights $\pi_{ik}(\mathbf{x}_i)$, this mixture distribution can approximate any continuous density function (e.g., Abrahamowicz, Clampl and Ramsay (1992), Chui, Smith and Ward

(1980)). The weights are modeled using a feed-forward neural network (FFNN) with H hidden layers with N_l neurons in hidden layer l and a softmax activation function, equivalent to multinomial logistic weights, in its final layer. The model is

$$(8) \quad \begin{aligned} \pi_{ik}(\mathbf{x}_i, \mathcal{W}_i) &= \frac{\exp\{\gamma_{Hk}(\mathbf{x}_i, \mathcal{W}_i)\}}{\sum_{j=1}^K \exp\{\gamma_{Hj}(\mathbf{x}_i, \mathcal{W}_i)\}}, \\ \gamma_{lk}(\mathbf{x}_i, \mathcal{W}_i) &= W_{ilk0} + \sum_{j=1}^{N_l} W_{ilkj} \psi\{\gamma_{l-1,j}(\mathbf{x}_i, \mathcal{W}_i)\} \quad \text{for } l \in \{1, \dots, H\}, \\ \gamma_{0k}(\mathbf{x}_i, \mathcal{W}_i) &= W_{i0k0} + \sum_{j=1}^p W_{i0kj} x_{ij}, \end{aligned}$$

where $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$, $\mathcal{W}_i = \{W_{ilkj}\}$ are parameters to be estimated and ψ is the activation function. Activation functions are nonlinear transformations applied to each output element of a layer and are a key feature of neural networks that allows them to learn complex, nonlinear dependencies in the data. The SPQR methodology of [Xu and Reich \(2021\)](#) admits most of the commonly used activation functions, and we use the rectified linear unit (ReLU) ([Nair and Hinton \(2010\)](#)). FFNNs use optimization to obtain optimum values of $\pi_{ik}(\mathbf{x}_i, \mathcal{W}_i)$. In SPQR the FFNN minimizes the negative log-likelihood loss associated with the density in (7), using the process values evaluated at locations $\mathbf{s}_i$ as the response. Building on the universal approximation theorem for FFNNs ([Hornik, Stinchcombe and White \(1989\)](#)), [Xu and Reich \(2021\)](#) argue that, with $H = 1$ and large K and N_1 , the model in (8) can approximate any conditional density function that is smooth in its arguments.

Within this framework, approximating the conditional distributions is equivalent to estimating the weights $\mathcal{W}$. Unlike a typical statistical learning problem, observational data are not used to estimate $\mathcal{W}$. Rather, the weights are learned from training data generated from the PMM with parameters $\boldsymbol{\theta}^{\text{SPAT}} \sim p^*$ and then a realization from the process over sites $\mathbf{s}_i$ and $\mathbf{s}_{(i)}$ from the model conditioned on $\boldsymbol{\theta}^{\text{SPAT}}$. Specifically, we generate data at the observed spatial location with the same neighbor sets to be used in the analysis. We select the design distribution p^* with support covering the range of plausible values for $\boldsymbol{\theta}^{\text{SPAT}}$. Given these values, we generate $U(\mathbf{s})$ at $\mathbf{s} \in \{\mathbf{s}_i, \mathbf{s}_{(i)}\}$. The feature set $\mathbf{x}_i$ for modeling u_i at location $\mathbf{s}_i$ thus contains the spatial parameters $\boldsymbol{\theta}^{\text{SPAT}}$ and the process values $U(\mathbf{s}_{(i)})$ at the neighboring locations.

Therefore, all that is required to build the approximation is the ability to generate small datasets from the model. The size of the training data is effectively unlimited, meaning the approximation can be arbitrarily accurate. Once the weights have been learned, applying the FFNN to the approximate likelihood is straightforward, and the Vecchia approximation ensures that the computational burden increases linearly in the number of spatial locations. The proposed estimation approach is in the same vein as recent simulation based neural inference methods ([Gerber and Nychka \(2021\)](#), [Lenzi et al. \(2021\)](#), [Sainsbury-Dale, Zammit-Mangion and Huser \(2023\)](#)); an important distinction is that, instead of estimating model parameters using a neural network, our approach estimates a set of conditional densities which approximates the full model likelihood.

The weights in (7) are estimated separately for each location. That is, each component $f_i(u_i | \boldsymbol{\theta}^{\text{SPAT}}, u_{(i)})$ in (6) is modeled using its own FFNN, for $i = 2, \dots, n$. The model is fit using the R package SPQR ([Xu and Majumder \(2022\)](#)), and consequently, the fitting process is referred to as the SPQR approximation. The SPQR package supports hardware acceleration for systems with a CUDA-compatible NVIDIA graphical processing unit (GPU), which was used for all SPQR models in this paper and provided significant speedups for computation

Algorithm 1 Local SPQR approximation**Require:** Locations $\mathbf{s}_1, \dots, \mathbf{s}_n$ and corresponding sets of neighboring locations $\mathbf{s}_{(1)}, \dots, \mathbf{s}_{(n)}$ **Require:** Design distribution p^* , training sample size N $i \leftarrow 2$ **while** $i \leq n$ **do** $k \leftarrow 1$ **while** $k \leq N$ **do**Draw values of $\boldsymbol{\theta}_k^{\text{SPAT}} \sim p^*$ Generate $U_k(\mathbf{s})$ at $\mathbf{s} \in \{\mathbf{s}_i, \mathbf{s}_{(i)}\}$ given $\boldsymbol{\theta}_{2k}$ using (3)Define features $\mathbf{x}_{ik} = (\boldsymbol{\theta}_k^{\text{SPAT}}, u_{(i)k})$, where $u_{(i)k} = \{U_k(\mathbf{s}); \mathbf{s} \in \mathbf{s}_{(i)}\}$ $k \leftarrow k + 1$ **end while**solve $\hat{\mathcal{W}}_i \leftarrow \operatorname{argmax}_{\mathcal{W}} \prod_{k=1}^N f_i(u_{ik} | \mathbf{x}_{ik}, \mathcal{W})$ for $f_i(u_i | \mathbf{x}_i, \mathcal{W}_i)$ defined in (7) using

SPQR

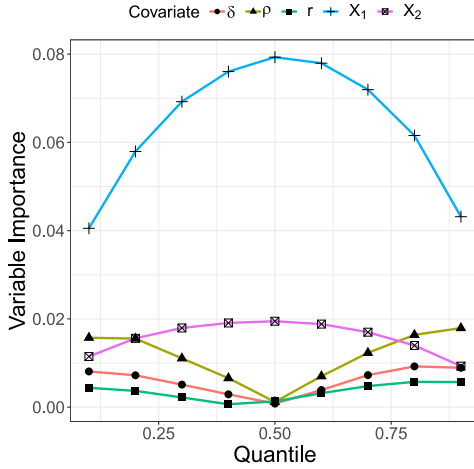
 $i \leftarrow i + 1$ **end while**

times. All computations were carried out on a mobile workstation with 11th Gen Intel Core i7-11800H processors (eight cores, 16 logical processors), 64 GB of RAM, and an NVIDIA T600 laptop GPU with CUDA support. Algorithm 1 outlines the local SPQR procedure.

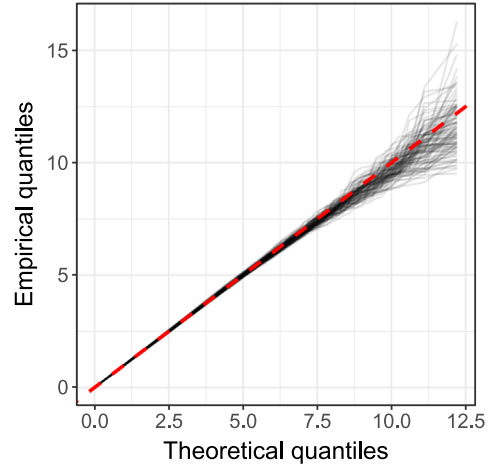
Given the approximate model in (5) for f_y with an SPQR approximation for f_u , a Bayesian analysis using MCMC methods is straightforward. We use Metropolis updates for both $\boldsymbol{\theta}^{\text{MARG}}$ and $\boldsymbol{\theta}^{\text{SPAT}}$. For a spatially-varying coefficient model with local GEV coefficients for location $\mathbf{s}_i$, we update parameters $\{\mu(\mathbf{s}_i), \sigma(\mathbf{s}_i), \xi(\mathbf{s}_i)\}$ as a block sequentially by site, and exploit the Vecchia approximation to use only terms in the likelihood corresponding to sites j such that $j \in \mathcal{N}_i$, that is, sites for which site i is included in the neighboring set. All Metropolis updates are tuned to give acceptance probability 0.4, and convergence is diagnosed based on the visual inspection of the trace plots. Additional computational details are given for specific analyses below, and MCMC code is provided in a GitHub repository—<https://github.com/reetamm/SPQR-for-spatial-extremes>.

5. Simulation study. We illustrate the potential of the proposed algorithm with a simulation study to evaluate the performance of the method on extreme value data in terms of both density estimation and parameter estimation. We consider the PMM as the underlying spatial process and an STVC model for the marginal GEV parameters. The goal of the study is to assess whether our approach can simultaneously model the marginal GEV distributions and the underlying extremal spatial process. We simulated data from the PMM, defined in Section 3, at $n = 50$ spatial locations distributed randomly on the unit square. To put the MSPs and GPs on the same scale, we assume common smoothness parameter $\alpha_R = \alpha_W = \alpha = 1$ and parameterize ρ_W and ρ_R to give the same effective range, that is, the distance at which the correlation of the GP reaches 0.05 ($h = \rho_W \log(20)$) and the χ -coefficient of the MSP reaches 0.05 ($h = \rho_R 4\Phi^{-1}(1 - 0.05/2)^2$ where Φ is the standard normal distribution function). This results in $\rho = \rho_W$ and $\rho_R = 0.19\rho$.

The locations are ordered by their distance from their origin. For the i th location $\mathbf{s}_i$, with $i > 1$, the Vecchia neighbor set $\mathbf{s}_{(i)}$ consists of the m nearest neighbors of $\mathbf{s}_i$ among the previous $i - 1$ locations. Up to 15 conditioning points are used in the Vecchia neighbor set, that is, $m = \min(i - 1, 15)$, where $m = |\mathcal{N}_i|$. We use the local SPQR model outlined in Algorithm 1 to model the conditional densities at each location, which employs stochastic gradient descent with the adaptive moment estimation (Adam) optimizer (Kingma and



(a) Variable importance of δ , ρ , r , and the two nearest neighbors.



(b) Q-Q plots on the exponential scale for checking goodness-of-fit.

FIG. 2. Model diagnostics for local SPQR fit on PMM data: VI plot of the five most important variables (left) and Q-Q plot on the exponential scale (right) for the local SPQR at site 25. The gray lines correspond to Q-Q plots of fits to different simulated datasets.

Ba (2014)). We compared multiple SPQR models as part of the density estimation process. Models were compared on the basis of the log-score and the Kullback–Leibler (KL) divergence between the estimated and true densities. Architectures with the lowest validation loss were chosen in each case. Our NN architecture for each SPQR model consists of two hidden layers with 30 and 15 neurons, 15 output nodes, a learning rate 0.001, batch size 100, and 50 epochs. We train the SPQR model with design distribution p^* , generating samples uniformly on $\rho \in (0, 0.5)$ and $\delta \in (0, 1)$.

We first evaluate the SPQR fits of the PMM full conditional distributions. Figure 2(a) plots variable importances for the local SPQR model at site 25. The VI plot identifies the features that are most important for explaining different aspects of the conditional distribution; the spatial parameters are found to be more important at extreme quantiles, while the process realizations at the Vecchia neighbor locations are more important closer to the median. The plots indicate that the conditional distributions at each location are sensitive to the spatial parameters; details of the VI metric used by SPQR are presented in the Supplementary Material (Majumder, Reich and Shaby (2024), Appendix A.7). To assess goodness of fit, we repeat the process of fitting the local SPQR model at site 25 for 100 independent datasets simulated from the PMM. Figure 2(b) is a Q-Q plot based on true and fitted values from the SPQR models, where each line corresponds to one of the 100 datasets. The Q-Q plot is presented on the exponential scale (Heffernan and Tawn (2001)) to verify whether the model can adequately capture tail behavior. The values fall along the $Y = X$ line, suggesting a good model fit. Computation time for local SPQR at sites with all 15 neighbors is approximately 22 minutes. The `doParallel` package in R was used to parallelize SPQR model fits and improve computation times.

We conduct a simulation study to explore how density-estimation errors propagate to parameter-estimation errors for spatially varying GEV parameters. Two scenarios are considered for the simulation studies, corresponding to $\delta = \{0.2, 0.8\}$. For both scenarios we assume the true values of $\rho = 0.15$ and $r = 0.80$, and simulate 100 datasets at the 50 locations. Each dataset consists of 50 independent (time) replicates. For spatial coordinates $\mathbf{s} = (s_1, s_2)$, the marginal GEV parameters are

$$\mu_0(\mathbf{s}) = \exp\{2 + \cos(2\pi s_1) + \cos(2\pi s_2)\},$$

$$\begin{aligned}\mu_1(\mathbf{s}) &= 1, \\ \sigma(\mathbf{s}) &= \exp\{\cos(2\pi s_2)\}, \\ \xi(\mathbf{s}) &= \frac{1}{2} \sin\left(\frac{\pi}{2}s_1\right).\end{aligned}$$

We model the marginal GEV parameters using an STVC model. At each location we assume the data to be GEV with parameters

$$Y_t(\mathbf{s}) \sim \text{GEV}(\mu_0(\mathbf{s}) + \mu_1(\mathbf{s})X_t^*, \sigma(\mathbf{s}), \xi(\mathbf{s})),$$

with $X_t^* = (t - 25.5)/10$, $t = 1 : 50$. The variable X_t^* represents changes in the location parameter over time.

The intercept process $\mu_0(\mathbf{s})$ is assigned a GP prior with a Matérn covariance function, and nugget effects allow local heterogeneity,

$$\begin{aligned}\mu_0(\mathbf{s}) &= \tilde{\mu}_0(\mathbf{s}) + e_0(\mathbf{s}), \\ e_0(\mathbf{s}) &\overset{\text{iid}}{\sim} \text{Normal}(0, v_{\mu_0}), \\ \tilde{\mu}_0(\mathbf{s}) &\sim \text{GP}(\beta_{\mu_0}, \tau_{\mu_0}^2 K(\mathbf{s}, \mathbf{s}'; \rho_{\mu_0}, \kappa_{\mu_0})), \\ \beta_{\mu_0} &\sim \text{Normal}(0, 10^2), \quad \tau_{\mu_0}^2, v_{\mu_0}^2 \overset{\text{iid}}{\sim} \text{IG}(0.1, 0.1), \\ \log \rho_{\mu_0} &\sim \text{Normal}(-1, 1), \quad \log \kappa_{\mu_0} \sim \text{Normal}(-2, 1),\end{aligned}$$

where $K(\mathbf{s}, \mathbf{s}'; \rho_{\mu_0}, \kappa_{\mu_0})$ is the Matérn correlation function with spatial range ρ_{μ_0} and smoothness parameter κ_{μ_0} , and $\text{IG}(\cdot, \cdot)$ is the inverse-Gamma distribution. The slope $\mu_1(\mathbf{s})$, the log-scale $\log \sigma(\mathbf{s})$, and the shape $\xi(\mathbf{s})$ are modeled similarly using GPs. The spatial parameters have priors $\delta \sim \text{Uniform}(0, 1)$, $\rho \sim \text{Uniform}(0.0, 0.5)$, and $r \sim \text{Uniform}(0, 1)$. Runtimes were approximately one minute for 1000 iterations of the MCMC.

Table 1 details coverage of the empirical 95% intervals for the posterior distribution of the marginal GEV parameters. Mean coverage across locations is near or at nominal level across different parameters and scenarios. Figure 3 plots the sampling distribution of the posterior mean estimator of δ for the two scenarios and provides empirical coverage of the 95% posterior interval. Both scenarios show low bias; coverage is 91% for the asymptotic independence scenario, and 97% for the asymptotic dependence scenario. Overall, the SPQR approach is able to distinguish between the two asymptotic regimes in the presence of spatially varying marginals.

Additional simulation studies are presented in the Supplementary Material (Majumder, Reich and Shaby (2024), Appendix B), which involve special cases of the PMM. The first considers a GP as the spatial process, and evaluates both the global and local SPQR methods. The second study is also for a PMM but in a non-STVC setting, where we explore a few additional scenarios, including negative shape parameters and missing and censored data. The final demonstrates the need of nonlinear NN layers by fitting a SPQR model without any hidden NN layers.

TABLE 1
Coverage (in %) for marginal GEV parameters under two scenarios based on MCMC simulations over 100 datasets. The three values represent the minimum, mean, and maximum coverage across the 50 study location

	μ_0	μ_1	σ	ξ
$\delta = 0.2$	(86, 92, 96)	(91, 96, 99)	(83, 91, 98)	(92, 96, 100)
$\delta = 0.8$	(85, 93, 100)	(90, 95, 100)	(86, 93, 98)	(90, 96, 100)

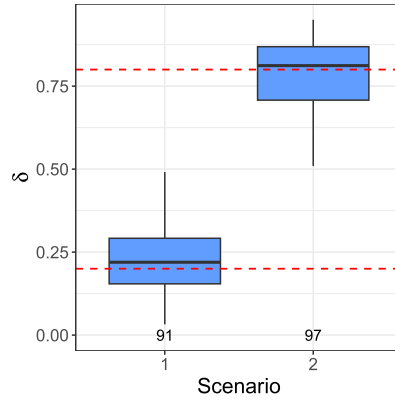


FIG. 3. Sampling distribution of the posterior mean for the asymptotic dependence parameter δ for two simulation scenarios. The horizontal dashed lines are true values, and the numbers along the bottom give the empirical coverage of the 95% intervals.

6. Analysis of extreme streamflow in the U.S.

6.1. Data description and exploratory analysis. We apply the proposed methods to model extreme streamflow from 1972–2021 at 487 stations across the U.S. with complete data. These locations are part of the USGS Hydro-Climatic Data Network (HCDN) 2009 (Lins (2012)) and are relatively unaffected by human activities. The data is downloaded using the dataRetrieval package in R (De Cicco et al. (2022)), and the code is made available in our GitHub repository.¹ Our goal is to identify regions within the U.S. where the distribution of extreme streamflow is changing over time. The annual maximum of daily streamflow is measured in m^3/s , and for each of the $T = 50$ years and $n = 487$ stations, the response $Y_t(\mathbf{s})$ is taken to be the logarithm of the annual maximum. The log-transformation was chosen as a Box–Cox transformation parameter after comparing parameter values between -2 and 2 on the basis of goodness of fit, profile likelihood values, and the stability of initial MLE estimates at the locations. Figure 4 plots the sample 0.9 quantile of the $T = 50$ observations at each station, which show considerable spatial variation.

In order to study the dependence structure of the process, especially at its extremes, we consider the conditional exceedance probability $\chi_u(h)$ of maximum streamflow at pairs of

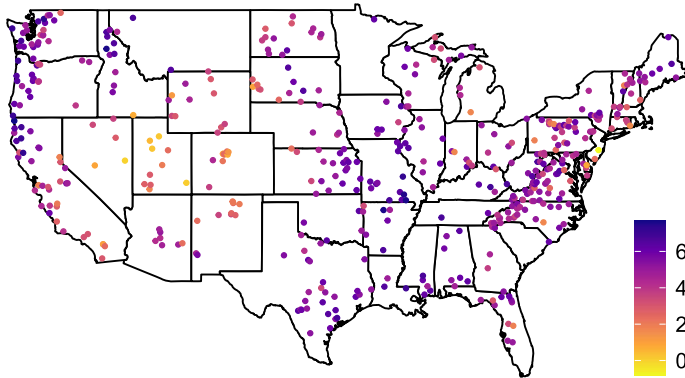


FIG. 4. HCDN annual maxima: Sample 0.9 quantile of the log-annual streamflow maxima $Y_t(\mathbf{s})$ (in m^3/s) at each of the 487 gauges.

¹<https://github.com/reetamm/SPQR-for-spatial-extremes>.

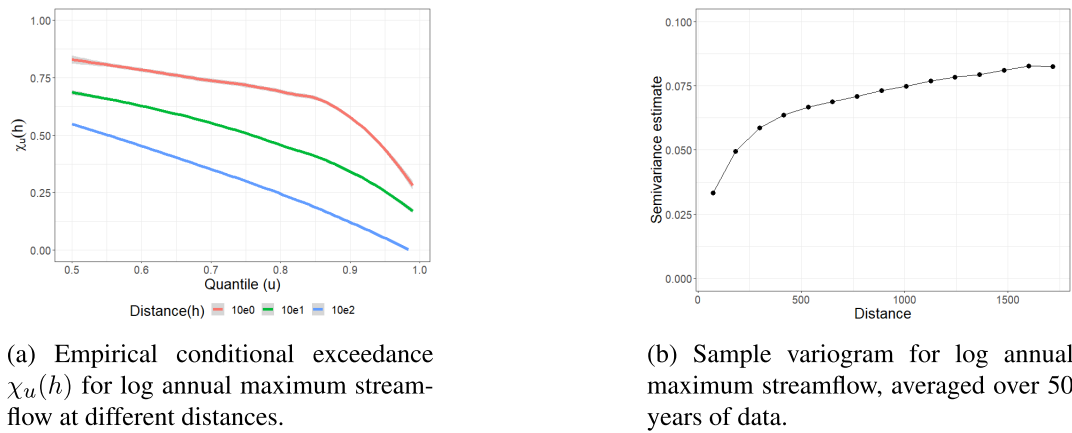


FIG. 5. Spatial behavior of log annual maximum streamflow in terms of the conditional exceedance and the variogram, units of km.

stations separated by a distance $h = \|\mathbf{s}_j - \mathbf{s}_k\|$ in kilometers (km). Figure 5(a) plots $\chi_u(h)$ for rank-standardized streamflow data as a function of u for different values of h . The rank standardization ensures a Uniform(0, 1) marginal distribution at each location. Stations farther away from each other can be seen to have less extremal dependence, with tail dependence approaching 0 for stations 1000 km apart. Figure 5(b) plots the mean of the annual variograms of the streamflow data. It shows a range of over 1500 km as well as the presence of a nugget effect. Both plots suggest that extremal streamflow is spatially dependent at distances of 1000 km or more, even after accounting for the spatial differences in the marginal distributions.

For the marginals at each location, we assume GEV distributions with STVC,

(9)
$$Y_t(\mathbf{s}) \sim \text{GEV}[\mu_0(\mathbf{s}) + \mu_1(\mathbf{s})X_t, \sigma(\mathbf{s}), \xi(\mathbf{s})],$$

where $X_t = (\text{year}_t - 1996.5)/10$ for $\text{year}_t = 1972 + t - 1$. This parameterization attempts to capture changes in the location parameter in the past 50 years due to changing climate; positive values of $\mu_1(\mathbf{s})$ would suggest an increase in the magnitude of the annual extremal streamflow. Figure 6 plots variograms for MLE estimates (estimated separately by location) of the GEV parameters at each location. All four GEV parameters show spatial dependency, which motivates the STVC specification.

The marginal GEV parameters for each location are assigned GP priors with a Matérn correlation functions and nugget effects to allow for local heterogeneity. The prior specification is similar to the one used in Section 5, except with the additional assumption of a common Matérn smoothness parameter was made for all four GP priors to improve MCMC convergence, that is, $\kappa_{\mu_0} = \kappa_{\mu_1} = \kappa_{\sigma} = \kappa_{\xi} = \kappa$. For the residual model, we use the PMM in

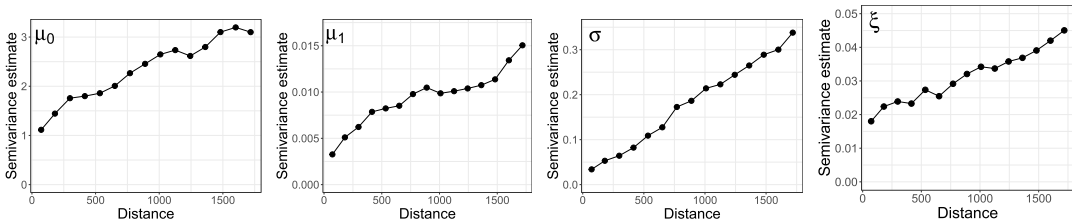


FIG. 6. Sample variogram plots for the MLE of GEV parameters for extreme streamflow at each location. Parameters are labeled on the top-left of each panel.

Section 3 for spatial dependence and assume independence across years. The specification includes a nugget based on the mean variogram (Figure 5(b)). As priors for the joint parameters θ^{SPAT} , we set $\delta, r \sim \text{Uniform}(0, 1)$, and $\rho \sim \text{Uniform}(0, 6251)$ measured in km.

6.2. Results. The local SPQR approximation for the log of annual streamflow maxima is thus based around $\theta^{\text{SPAT}} = (\delta, \rho, r)$, and models are fit using 200,000 synthetic observations at each of the 486 locations with neighbors. Once the local SPQR models have been fit, we run two MCMC chains for 30,000 iterations each, with two different starting values of δ . The first 10,000 iterations from each chain are discarded as burn-in; additional results are provided in the Supplementary Material (Majumder, Reich and Shaby (2024), Appendix C.2).

6.2.1. Parameter estimates. The posterior means (standard deviations) of the spatial parameters are $\hat{\delta} = 0.45$ (0.02), $\hat{\rho} = 807$ (45) km, and $\hat{r} = 0.92$ (0.004). The posterior of δ has a 95% interval of (0.40, 0.49), which puts the process in the asymptotic independence regime with high probability. The GEV Matérn smoothness parameter estimate is $\hat{\kappa} = 0.60$ (0.03), and the four range parameters (in km) are $\hat{\rho}_{\mu_0} = 12,435$ (10,645), $\hat{\rho}_{\mu_1} = 27,605$ (10,689), $\hat{\rho}_{\sigma} = 20,311$ (11,232), and $\hat{\rho}_{\xi} = 20,320$ (11,481). The STVC parameters are, therefore, much smoother over space than the year-to-year variation captured by ρ . The intercept is the most variable GEV parameter across space, and the slope is the least, having the smallest and largest range parameter estimates, respectively. It is likely that the intercept varies the most because the magnitude of streamflow at a station is dependent on very local features. Conversely, the slope parameter may vary smoothly because the drivers of change are regional rather than local in nature.

Figure 7 plots the spatial distribution of posterior means for the GEV parameters. The shape and scale parameters are negatively associated for large parts of the country, possibly a consequence of the constraints on the GEV parameters. The scale parameter $\sigma(\mathbf{s})$ is highest in the Arkansas–Rio Grande–Texas Gulf and the Missouri basin regions,² and the scale parameter is highest in the North Atlantic–Appalachian and the Columbia–Pacific Northwest regions. Areas with high estimates of $\mu_0(\mathbf{s})$ and $\xi(\mathbf{s})$ coincide with areas of high precipitation in the 1991–2020 U.S. Climate Normals,³ the current official baseline for describing average U.S. climate. This suggests that an association between precipitation and streamflow maxima. In Figure 7d we note that the posterior means of the shape parameters $\xi(\mathbf{s})$ are negative at all 487 stations on the log-transformed data. The use of the log-transform for the streamflow data leads to negative GEV shape parameter estimates, imposing a finite upper bound on the distribution, even on the original scale. However, we do not expect this to affect estimation of either the quantiles of the marginal distribution or the joint exceedance probabilities. To assess the effect of transforming the data, additional analysis was carried out on PMMs fitted to the original scale of the data as well as on the square root of streamflow. Results from both these analyses are provided in the Supplementary Material (Majumder, Reich and Shaby (2024), Appendix C.4).

Our primary interest, however, is in estimates of the location parameters across the U.S. Figure 7a plots the posterior mean of the intercept $\mu_0(\mathbf{s})$ of the location parameter and has a spatial distribution similar to Figure 4. Figure 7b plots the posterior mean of the slope $\mu_1(\mathbf{s})$ of the location parameters with respect to time across the U.S., and Figure 8 plots the posterior probability of the slope parameter being positive, $\Pr[\mu_1(\mathbf{s}) > 0]$. Positive slope estimates in Figure 7b indicate an increase in extreme streamflow over time, and high probabilities in Figure 8 indicate stronger evidence for the increase being significant. The majority of the

²DOI unified regions.

³U.S. Climate Normals, U.S. Climate Atlas

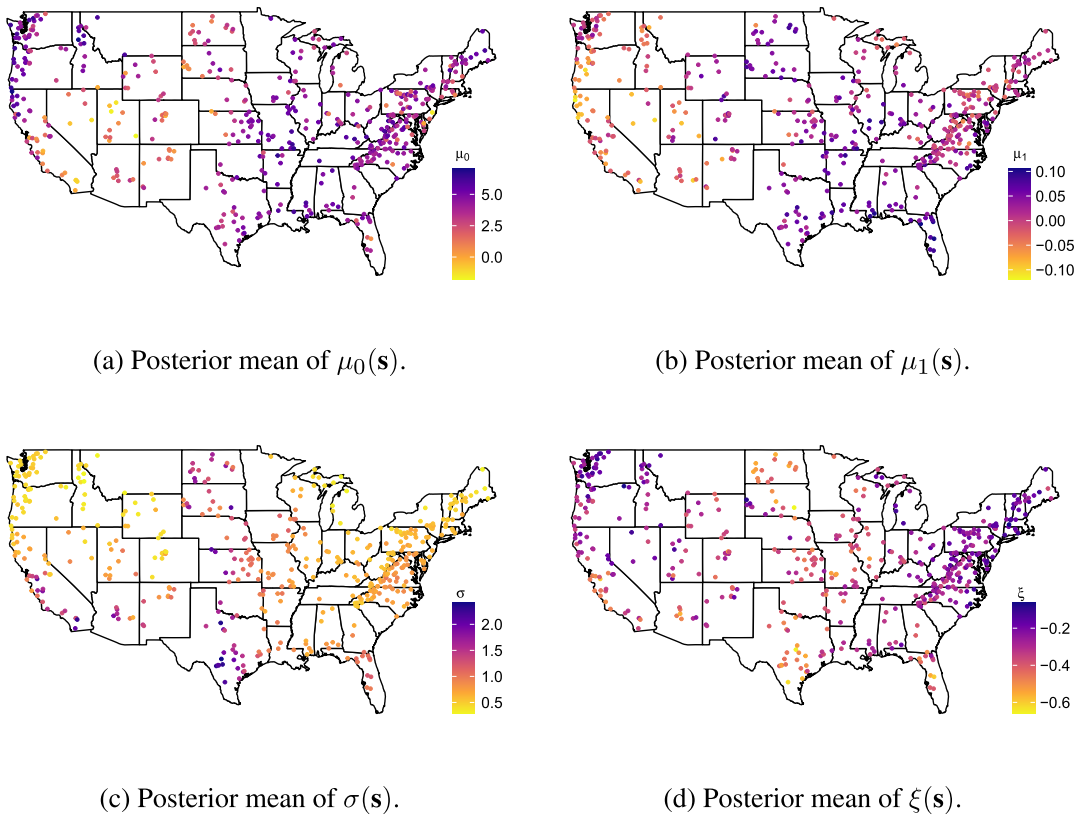


FIG. 7. HCDN GEV parameter estimates: Posterior means for 487 stations in the U.S.A. based on log-transformed data from 1972–2021.

positive slope parameters are concentrated in the Mississippi and Missouri basins and the Arkansas–Rio Grande–Texas Gulf regions. The North Atlantic–Appalachian region in the east has a large number of catchments with slope estimates near zero, but the majority of zero and negative slope estimates are concentrated around the Lower Colorado Basin, Columbia–Pacific Northwest, and the California–Great Basin regions. An exception is Washington, in the northwest, which has high estimates of the slope. The upper Colorado basin, which includes Wyoming, Colorado, and New Mexico, is of particular interest since HCDN stations

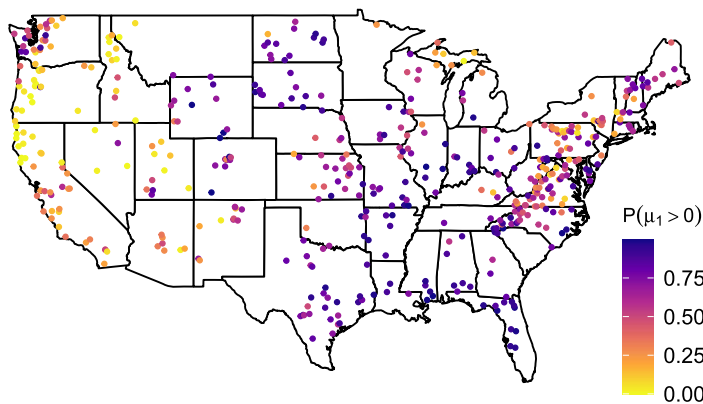


FIG. 8. Posterior of $\Pr[\mu_1(s) > 0]$ for GEV location parameters based on log-transformed data from 1972–2021.

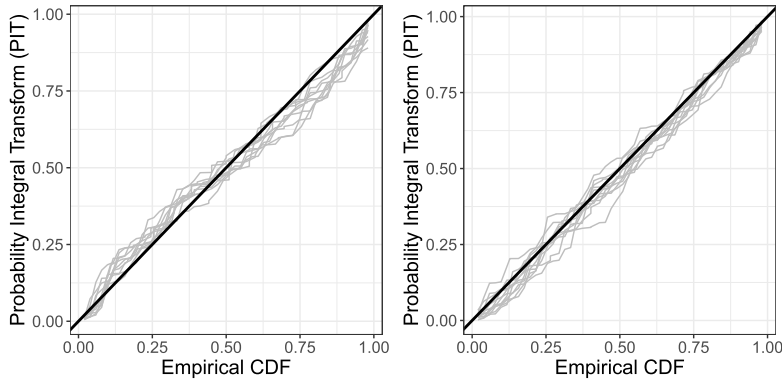


FIG. 9. PIT scores of marginal GEV fits for HCDN locations in Colorado (left) and New Mexico (right).

in the region have relatively low 0.9 quantile values in Figure 4 as well as low $\mu_0(\mathbf{s})$ estimates in Figure 7a, suggesting that extreme streamflow is starting to have large impacts in this region over time.

6.2.2. Regional joint exceedance behavior. To study the behavior of extreme streamflow jointly for multiple locations, we consider two clusters of HCDN stations in Colorado (CO) and in New Mexico (NM), comprising of 10 and 11 stations, respectively. Figure 9 plots the probability integral transform (PIT) scores of marginal GEV fits for locations within each cluster based on posterior means of GEV parameters, which suggest adequate marginal fits for both sets of locations. To quantify the effect of changing climate on extreme streamflow for each cluster, we look at the joint posterior probability of streamflow maxima exceeding the observed 0.90 quantile values, as shown in Figure 4, that is, $\Pr[Y_t(\mathbf{s}_i) > q_i, i = 1, \dots, n_i]$ for the sample 0.90 quantile q_i at location $\mathbf{s}_i$, where n_i is the number of stations within each cluster. Since our marginal models have STVC, we are able to calculate posterior probabilities for both 1972 and 2021. The probabilities are calculated based on 200 post burn-in MCMC samples from the posterior distribution of the parameters. For each MCMC sample, 20,000 observations are generated from the fitted model; the 1972 and 2021 probabilities are calculated based on 10,000 observations each.

In Colorado the 10 HCDN stations correspond to catchments with drainage ranging from 15.5 km² to 432.9 km². The cluster is well separated from other stations and has high posterior estimates of $\Pr[\mu_1(\mathbf{s}) > 0]$ (minimum = 0.10, mean = 0.57). The HCDN stations are all situated in the Upper Colorado Basin region, and the set of neighbors are spread across the Upper and Lower Colorado Basin regions. The joint exceedance probability for 1972 has a mean of 0.075 with an SD of 0.04. The joint exceedance probability for 2021 has a mean of 0.17 with an SD of 0.046. This corresponds to an increase of around 125% of the mean joint exceedance probability in the last 50 years. Note that if we assume independence across locations, the joint exceedance probability for the 10 locations would be approximately 10^{-10} . Finally, the probability that the joint exceedance in 2021 is higher than in 1972 is 0.90, providing strong evidence in favor of increased extremal streamflow in the area, possibly due to changing climate.

In New Mexico the 11 HCDN stations correspond to catchments with drainage ranging from 43.8 km² to 4804.9 km². The stations are all situated in the Upper Colorado Basin region, and the set of Vecchia neighbors are located across the Upper and Lower Colorado Basin as well as the California–Great Basin region. The catchments have a mix of high- and low-posterior estimates of $\Pr[\mu_1(\mathbf{s}) > 0]$ (minimum = 0.07, mean = 0.48). The joint exceedance probability for 1972 has a mean of 0.045 with an SD of 0.012. The joint exceedance

probability for 2021 has a mean of 0.053 with an SD of 0.017. This corresponds to an increase of around 18% of the mean joint exceedance probability in the last 50 years. The probability that the the joint exceedance in 2021 is higher than in 1972 is 0.695, which is lower than the result for Colorado but still significantly higher than the independent scenario. [Majumder, Reich and Shaby \(\(2024\), Appendix C.2\)](#) contains additional results for both clusters.

6.2.3. Model comparison and model fit. We compared the PMM with three spatial processes—a GP, an MSP, and the Huser–Wadsworth (HW) process. In each case inference is carried out by using SPQR for density estimation and using MCMC afterward for parameter estimation. This allows for a comparison of the appropriateness of different spatial processes for our application. For all three competing models, we used the same neural network architecture in our local SPQR fits as the PMM.

Table 2 lists estimates and standard errors from leave-one-out cross-validation (LOO-CV) ([Vehtari, Gelman and Gabry \(2016\)](#)) and the Watanabe–Akaike information criterion (WAIC) ([Watanabe \(2010\)](#)) for the four models. Based on both metrics, the PMM has the lowest values and thus the best model fit, followed by the HW model. The MSP and the GP are noticeably worse than the PMM and HW models, but their LOO-CV and WAIC estimates are relatively close to each other. This suggests that models, which assume asymptotic (in)dependence (e.g., the MSP and the GP), are likely to have worse fits when the dependence structure is complex. An intercomparison of the posterior means of the slope parameter from each of these models is presented in the Supplementary Material ([Majumder, Reich and Shaby \(2024\), Appendix C.3](#)).

Finally, we evaluate the fit when using the PMM as the underlying spatial process by comparing estimates of $\chi_u(h)$ based on the posterior distribution with empirical estimates obtained from the rank-standardized observed data. For this we choose three regions within the U.S.A., the first two of which are the CO and NM clusters studied in Section 6.2.2. Both clusters have a total of 34 locations (including the Vecchia neighbor locations). The third area consists of HCDN locations in Oregon (OR) and with its Vecchia neighbors, accounting for a total of 56 locations in Oregon, Washington, Nevada, and California. This region provides a contrast to the CO and NM clusters; locations are farther apart and have low estimates of the slope and scale but high estimates of the shape parameter. For each of the three clusters, we generated realizations of the spatial process based 200 samples from the posterior and estimated $\chi_u(h)$ for each sample.

Figure 10 plots the empirical and posterior estimates of $\chi_u(h)$ for the three regions. The top row plots the upper tail coefficient for high quantiles, for locations $h = 10$ km apart for CO and NM, and $h = 100$ km apart for OR. The bottom row plots estimates of $\chi_u(h)$ for $u = 0.90$. In each panel the bold line represents the empirical estimate, and the dashed line and band represent the mean and 95% interval of the posterior estimates, respectively. We note that both the model-based and empirical estimates of the upper tail coefficient behaves

TABLE 2
Estimates and standard errors (in parentheses) from leave-one-out cross validation (LOO-CV) and the Watanabe–Akaike information criterion (WAIC) for comparing the process mixture model (PMM), the Huser–Wadsworth model (HW), the max-stable process model (MSP), and the Gaussian process model (GP). Lower values indicate a better fit

	PMM	HW	MSP	GP
LOO-CV	29,108 (540)	29,708 (544)	32,058 (583)	33,842 (561)
WAIC	29,559 (549)	30,193 (565)	33,441 (552)	34,440 (585)

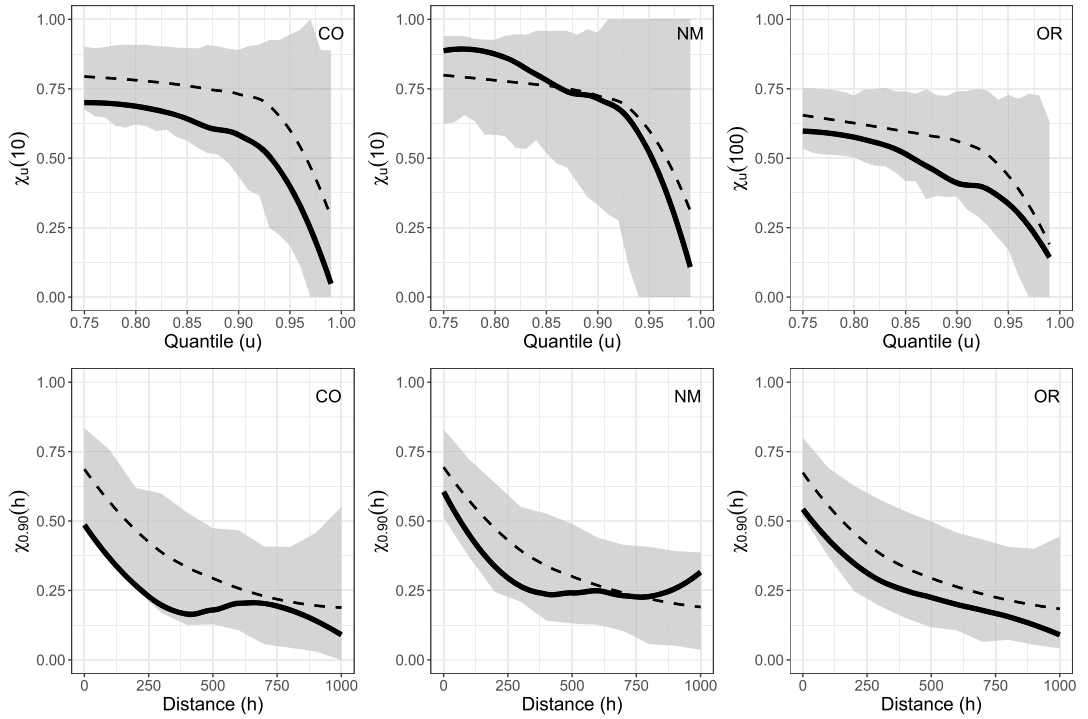


FIG. 10. Estimates of $\chi_u(h)$ for three different regions: Empirical estimates of the upper tail coefficient in bold, compared against estimates based on the posterior distribution. The dashed lines and the bands correspond to the mean and 95% interval, respectively.

similarly for the three regions and are consistent with the empirical estimate that is obtained from data for the entire U.S.A. (Figure 5(a)). This suggests that the stationarity assumption for the dependence structure in our model is appropriate. The 95% intervals in the top row includes 0 for all three regions, reflecting the asymptotic independence corresponding to our estimate of δ . Both the empirical and posterior estimates show behavior similar to Figure 1(a) for the asymptotic independent case. The tail coefficients can also be seen to decrease as distance increases, similar to Figure 1(b). We noted similar behavior for other regions within the country. While the fitted model tends to overestimate $\chi_u(h)$ for several regions, it is able to capture the behavior of extremal dependence over large distances and at high quantile levels.

7. Discussion. In this paper we proposed a process mixture model (PMM) for spatial extremes, where the marginal distributions at different spatial locations are GEV and their spatial dependence is captured using a convex combination of a GP and an MSP. The PMM extends Huser and Wadsworth (2019) and is flexible enough to accommodate missingness and censoring as well as STVC for the marginal GEV distributions. We approximated the joint likelihood for the spatial model using a Vecchia approximation. We used the density regression model proposed in Xu and Reich (2021) to approximate this likelihood, whose weights are modeled using a feed-forward neural network and learned using synthetic data generated from a design distribution. Parameter estimation for the model is carried out using MCMC.

We used the PMM to analyze changes in annual maximum streamflow within the U.S. over the past 50 years. For this study we used the annual maximum streamflow measured at 487 stations in the USGS Hydro-Climatic Data Network. The posterior means of the location parameter have nonzero slope estimates in several parts of the country. We noted a

high concentration of positive slope estimates in the Mississippi and Missouri basins and the Arkansas–Rio Grande–Texas Gulf regions, indicating that extremal streamflow has increased in those areas over the last 50 years.

Future work will focus on the theoretical properties of this model. While it is straightforward to derive an analytical expression for $\chi_u(h)$ for the trivial case of a shared MSP and has been provided in the Supplementary Material (Majumder, Reich and Shaby (2024), Appendix A.4), that is, for $R(s) = R$, obtaining an analytical expression for the general case is more challenging. This would also enable us to investigate the properties of $\chi_u(h)$ for the PMM for different values of δ and as $h \rightarrow \infty$. We would like to further investigate improvements to the computational aspects of this model and identify reasonable plug-and-play settings for local and global SPQR approximations. Finally, we would also like to extend this model to provide climate-informed estimates by regressing the spatiotemporal variability of the EVA parameters onto large-scale climate drivers from GCM output with spatially-varying regression coefficients for local calibration. Another area of future work is to extend the model to accommodate more complex spatial dependence structures. Recent work on spatial extremes has incorporated graphical models as additional information for computing distances (Engelke and Hitz (2020)). While our work considers only spatial coordinates, it is possible to incorporate additional distance measures in the form of river network information to reflect the physical structure of watersheds and the so-called “river distance” between stations (Asadi, Davison and Engelke (2015)). While the stream network information is not readily available for these data, incorporating this network structure might improve spatial modeling. Spatial models on stream networks have been developed for both max-stable (Asadi, Davison and Engelke (2015)) and Gaussian (Santos-Fernandez et al. (2022b)) processes, and so it should be possible to incorporate these features into the PMM.

Acknowledgments. The authors thank Prof. Sankarasubramanian Arumugam of North Carolina State University for discussion of the data and scope of the project.

Funding. This work was supported by grants from the Southeast National Synthesis Wildfire and the United States Geological Survey’s National Climate Adaptation Science Center (G21AC10045), the National Science Foundation (CBET2151651, DMS2152887, and DMS2001433), and the National Institutes of Health (R01ES031651-01).

SUPPLEMENTARY MATERIAL

Appendices (DOI: [10.1214/23-AOAS1847SUPPA](https://doi.org/10.1214/23-AOAS1847SUPPA); .pdf). The Supplementary Material consists of three appendices. Appendix A goes over the some properties of the PMM, and an overview of the variable importance measure used in the text. Appendix B presents supplementary simulation studies detailing the performance of the PMM in various density estimation and parameter estimation scenarios. Appendix C consists of additional results from the HCDN data analysis, including MCMC convergence, model comparison and model fit results, and selected results from analyzing the extremal streamflow data in its original scale.

MCMC code for the HCDN streamflow data (DOI: [10.1214/23-AOAS1847SUPPB](https://doi.org/10.1214/23-AOAS1847SUPPB); .zip). The folder contains code pertaining to the case study presented in Section 6. Users can replicate our results by running `SPQR_fit.R`, which carries out the density estimation, and `simEVP_local_parallel_HUC02.R`, which carries out parameter estimation. The input data has already been provided for convenience; it can also be generated from scratch using `download_annual_max.R`. Most of the remaining files contain code for exploratory data analysis as well as generating the various plots that we have presented. Some require intermediate files which we have been unable to provide due to data size constraints, but they

can all be generated by the user, mostly as the full output of `SPQR_fit.R`. A vignette has also been provided in the form of `Demo.html` where the density estimation and parameter estimation has been demonstrated for a section of North East US.

REFERENCES

- ABRAHAMOWICZ, M., CLAMPL, A. and RAMSAY, J. O. (1992). Nonparametric density estimation for censored survival data: Regression-spline approach. *Canad. J. Statist.* **20** 171–185.
- ARCHFIELD, S. A., HIRSCH, R. M., VIGLIONE, A. and BLÖSCHL, G. (2016). Fragmented patterns of flood change across the United States. *Geophys. Res. Lett.* **43** 10–232.
- ASADI, P., DAVISON, A. C. and ENGELKE, S. (2015). Extremes on river networks. *Ann. Appl. Stat.* **9** 2023–2050.
- BLÖSCHL, G., HALL, J., VIGLIONE, A., PERDIGÃO, R. A., PARAJKA, J., MERZ, B., LUN, D., ARHEIMER, B., ARONICA, G. T. et al. (2019). Changing climate both increases and decreases European river floods. *Nature* **573** 108–111.
- BOPP, G. P., SHABY, B. A. and HUSER, R. (2021). A hierarchical max-infinitely divisible spatial model for extreme precipitation. *J. Amer. Statist. Assoc.* **116** 93–106.
- BROWN, B. M. and RESNICK, S. I. (1977). Extreme values of independent stochastic processes. *J. Appl. Probab.* **14** 732–739.
- CASTRUCCIO, S., HUSER, R. and GENTON, M. G. (2016). High-order composite likelihood inference for max-stable distributions and processes. *J. Comput. Graph. Statist.* **25** 1212–1229.
- CHUI, C., SMITH, P. and WARD, J. (1980). Degree of L_p approximation by monotone splines. *SIAM J. Math. Anal.* **11** 436–447.
- COLES, S., BAWA, J., TRENNER, L. and DORAZIO, P. (2001). *An Introduction to Statistical Modeling of Extreme Values* **208**. Springer.
- CONDON, L., GANGOPADHYAY, S. and PRUITT, T. (2015). Climate change and non-stationary flood risk for the upper Truckee River basin. *Hydrol. Earth Syst. Sci.* **19** 159–175.
- DATTA, A., BANERJEE, S., FINLEY, A. O. and GELFAND, A. E. (2016). Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *J. Amer. Statist. Assoc.* **111** 800–812.
- DAWDY, D. R., GRIFFIS, V. W. and GUPTA, V. K. (2012). Regional flood-frequency analysis: How we got here and where we are going. *J. Hydrol. Eng.* **17** 953–959.
- DE CICCIO, L. A., LORENZ, D., HIRSCH, R. M., WATKINS, W. and JOHNSON, M. (2022). dataRetrieval: R packages for discovering and retrieving water data available from U.S. federal hydrologic web services, Reston, VA. <https://doi.org/10.5066/P9X4L3GE>
- DE HAAN, L. (1984). A spectral representation for max-stable processes. *Ann. Probab.* **12** 1194–1204.
- DE HAAN, L. and FERREIRA, A. (2006). *Extreme Value Theory: An Introduction*. Springer, Berlin.
- ENGELKE, S. and HITZ, A. S. (2020). Graphical models for extremes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 871–932. <https://doi.org/10.1111/rssb.12355>
- ERHARDT, R. J. and SMITH, R. L. (2012). Approximate Bayesian computing for spatial extremes. *Comput. Statist. Data Anal.* **56** 1468–1481.
- FRANÇOIS, B., SCHLEF, K., WI, S. and BROWN, C. (2019). Design considerations for riverine floods in a changing climate—A review. *J. Hydrol.* **574** 557–573.
- FRANKS, S. W. (2002). Identification of a change in climate state using regional flood data. *Hydrol. Earth Syst. Sci.* **6** 11–16.
- GERBER, F. and NYCHKA, D. (2021). Fast covariance parameter estimation of spatial Gaussian process models using neural networks. *Stat* **10** e382.
- GREENBERG, D., NONNENMACHER, M. and MACKE, J. (2019). Automatic posterior transformation for likelihood-free inference. In *Proceedings of the 36th International Conference on Machine Learning* (K. Chaudhuri and R. Salakhutdinov, eds.). *Proceedings of Machine Learning Research* **97** 2404–2414.
- GUINNESS, J. (2018). Permutation and grouping methods for sharpening Gaussian process approximations. *Technometrics* **60** 415–429. <https://doi.org/10.1080/00401706.2018.1437476>
- HAZRA, A., HUSER, R. and BOLIN, D. (2021). Realistic and fast modeling of spatial extremes over large geographical domains. <https://doi.org/10.48550/ARXIV.2112.10248>
- HEFFERNAN, J. E. and TAWN, J. A. (2001). Extreme value analysis of a large designed experiment: A case study in bulk carrier safety. *Extremes* **4** 359–378.
- HIRABAYASHI, Y., MAHENDRAN, R., KOIRALA, S., KONOSHIMA, L., YAMAZAKI, D., WATANABE, S., KIM, H. and KANAE, S. (2013). Global flood risk under climate change. *Nat. Clim. Change* **3** 816–821.
- HIRSCH, R. M. (2011). A perspective on nonstationarity and water management 1. *J. Am. Water Resour. Assoc.* **47** 436–446.

- HIRSCH, R. M. and RYBERG, K. R. (2012). Has the magnitude of floods across the USA changed with global CO₂ levels? *Hydrol. Sci. J.* **57** 1–9.
- HORNIK, K., STINCHCOMBE, M. and WHITE, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Netw.* **2** 359–366.
- HUSER, R. and DAVISON, A. C. (2014). Space–time modelling of extreme events. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 439–461.
- HUSER, R., DAVISON, A. C. and GENTON, M. G. (2016). Likelihood estimators for multivariate extremes. *Extremes* **19** 79–103.
- HUSER, R., DOMBRY, C., RIBATET, M. and GENTON, M. G. (2019). Full likelihood inference for max-stable data. *Stat* **8** e218.
- HUSER, R., STEIN, M. L. and ZHONG, P. (2022). Vecchia likelihood approximation for accurate and fast inference in intractable spatial extremes models. Preprint. Available at [arXiv:2203.05626](https://arxiv.org/abs/2203.05626).
- HUSER, R. and WADSWORTH, J. L. (2019). Modeling spatial processes with unknown extremal dependence class. *J. Amer. Statist. Assoc.* **114** 434–444.
- JÄRVENPÄÄ, M., GUTMANN, M. U., VEHTARI, A. and MARTTINEN, P. (2021). Parallel Gaussian process surrogate Bayesian inference with noisy likelihood evaluations. *Bayesian Anal.* **16** 147–178.
- JOE, H. (1997). *Multivariate Models and Multivariate Dependence Concepts*, 1st ed. CRC Press/CRC, Boca Raton.
- KABLUCHKO, Z., SCHLATHER, M. and DE HAAN, L. (2009). Stationary max-stable fields associated to negative definite functions. *Ann. Probab.* **37** 2042–2065.
- KATZFUSS, M. and GUINNESS, J. (2021). A general framework for Vecchia approximations of Gaussian processes. *Statist. Sci.* **36** 124–141.
- KINGMA, D. P. and BA, J. (2014). Adam: A method for stochastic optimization. Preprint. Available at [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
- KNOX, J. C. (1993). Large increases in flood magnitude in response to modest changes in climate. *Nature* **361** 430–432.
- KOBYZEV, I., PRINCE, S. J. D. and BRUBAKER, M. A. (2021). Normalizing flows: An introduction and review of current methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **43** 3964–3979. <https://doi.org/10.1109/TPAMI.2020.2992934>
- KUNDZEWICZ, Z. W., KANAE, S., SENEVIRATNE, S. I., HANDMER, J., NICHOLLS, N., PEDUZZI, P., MECHLER, R., BOUWER, L. M., ARNELL, N. et al. (2014). Flood risk and climate change: Global and regional perspectives. *Hydrol. Sci. J.* **59** 1–28.
- KUNDZEWICZ, Z. W., KRYSANOVA, V., DANKERS, R., HIRABAYASHI, Y., KANAE, S., HATTERMANN, F. F., HUANG, S., MILLY, P. C., STOFFEL, M. et al. (2017). Differences in flood hazard projections in Europe—their causes and consequences for decision making. *Hydrol. Sci. J.* **62** 1–14.
- KUNKEL, K. E., KARL, T. R., SQUIRES, M. F., YIN, X., STEGALL, S. T. and EASTERLING, D. R. (2020). Precipitation extremes: Trends and relationships with average precipitation and precipitable water in the contiguous United States. *J. Appl. Meteorol. Climatol.* **59** 125–142.
- LEDFORD, A. W. and TAWN, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika* **83** 169–187.
- LEDFORD, A. W. and TAWN, J. A. (1997). Modelling dependence within joint tail regions. *J. Roy. Statist. Soc. Ser. B* **59** 475–499.
- LENZI, A., BESSAC, J., RUDI, J. and STEIN, M. L. (2021). Neural networks for parameter estimation in intractable models. Preprint. Available at [arXiv:2107.14346](https://arxiv.org/abs/2107.14346).
- LI, L., HOLBROOK, A., SHAHBABA, B. and BALDI, P. (2019). Neural network gradient Hamiltonian Monte Carlo. *Comput. Statist.* **34** 281–299. <https://doi.org/10.1007/s00180-018-00861-z>
- LIMA, C. H., LALL, U., TROY, T. and DEVINENI, N. (2016). A hierarchical Bayesian GEV model for improving local and regional flood quantile estimates. *J. Hydrol.* **541** 816–823.
- LINS, H. F. (2012). USGS hydro-climatic data network 2009 (HCDN-2009). *US Geological Survey Fact Sheet* **3047**.
- MAJUMDER, R., REICH, B. J. and SHABY, B. A. (2024). Supplement to “Modeling extremal streamflow using deep learning approximations and a flexible spatial process.” <https://doi.org/10.1214/23-AOAS1847SUPPA>, <https://doi.org/10.1214/23-AOAS1847SUPPB>
- MEEHL, G. A., ZWIERS, F., EVANS, J., KNUTSON, T., MEARN, L. and WHETTON, P. (2000). Trends in extreme weather and climate events: Issues related to modeling extremes in projections of future climate change. *Bull. Am. Meteorol. Soc.* **81** 427–436.
- MERZ, B., AERTS, J., ARNBJERG-NIELSEN, K., BALDI, M., BECKER, A., BICHET, A., BLÖSCHL, G., BOUWER, L. M., BRAUER, A. et al. (2014). Floods and climate: Emerging perspectives for flood risk assessment and management. *Nat. Hazards Earth Syst. Sci.* **14** 1921–1942.

- MILLY, P., BETANCOURT, J., FALKENMARK, M., HIRSCH, R. M., KUNDZEWICZ, Z. W., LETTENMAIER, D. P. and STOFFER, R. J. (2008). Stationarity is dead: Whither water management? *Earth* **4**.
- MILLY, P. C., BETANCOURT, J., FALKENMARK, M., HIRSCH, R. M., KUNDZEWICZ, Z. W., LETTENMAIER, D. P., STOFFER, R. J., DETTINGER, M. D. and KRYSANOVA, V. (2015). On critiques of “Stationarity is dead: Whither water management?”. *Water Resour. Res.* **51** 7785–7789.
- MILLY, P. C., DUNNE, K. A. and VECCHIA, A. V. (2005). Global pattern of trends in streamflow and water availability in a changing climate. *Nature* **438** 347–350.
- MORRIS, S. A., REICH, B. J. and THIBAUD, E. (2019). Exploration and inference in spatial extremes using empirical basis functions. *J. Agric. Biol. Environ. Stat.* **24** 555–572.
- NAIR, V. and HINTON, G. E. (2010). Rectified linear units improve restricted Boltzmann machines. In *Proceedings of the 27th International Conference on International Conference on Machine Learning, ICML'10* 807–814. Omnipress, Madison, WI, USA.
- PADOAN, S. A., RIBATET, M. and SISSON, S. A. (2010b). Likelihood-based inference for max-stable processes. *J. Amer. Statist. Assoc.* **105** 263–277.
- PAPAMAKARIOS, G., NALISNICK, E., REZENDE, D. J., MOHAMED, S. and LAKSHMINARAYANAN, B. (2021). Normalizing flows for probabilistic modeling and inference. *J. Mach. Learn. Res.* **22** 1–64.
- PENROSE, M. D. (1992). Semi-min-stable processes. *Ann. Probab.* **20** 1450–1463. <https://doi.org/10.1214/aop/1176989700>
- PRICE, L. F., DROVANDI, C. C., LEE, A. and NOTT, D. J. (2018). Bayesian synthetic likelihood. *J. Comput. Graph. Statist.* **27** 1–11.
- RASMUSSEN, C. E. (2003). Gaussian processes to speed up hybrid Monte Carlo for expensive Bayesian integrals. In *Seventh Valencia International Meeting, Dedicated to Dennis V. Lindley* 651–659. Oxford Univ. Press, London.
- REICH, B. J. and SHABY, B. A. (2012b). A hierarchical max-stable spatial model for extreme precipitation. *Ann. Appl. Stat.* **6** 1430–1451. <https://doi.org/10.1214/12-AOAS591>
- REICH, B. J., SHABY, B. A. and COOLEY, D. (2013). A hierarchical model for serially-dependent extremes: A study of heat waves in the western US. *J. Agric. Biol. Environ. Stat.* **19** 119–135.
- RIBATET, M., COOLEY, D. and DAVISON, A. C. (2012). Bayesian inference from composite likelihoods, with an application to spatial extremes. *Statist. Sinica* **22** 813–845.
- SAINSBURY-DALE, M., ZAMMIT-MANGION, A. and HUSER, R. (2023). Fast optimal estimation with intractable models using permutation-invariant neural networks.
- SALAS, J. D. and OBEYSEKERA, J. (2014). Revisiting the concepts of return period and risk for nonstationary hydrologic extreme events. *J. Hydrol. Eng.* **19** 554–568.
- SANG, H. and GENTON, M. G. (2014b). Tapered composite likelihood for spatial max-stable models. *Spat. Stat.* **8** 86–103.
- SANTOS-FERNANDEZ, E., VER HOEF, J. M., PETERSON, E. E., MCGREE, J., ISAAK, D. J. and Mengersen, K. (2022b). Bayesian spatio-temporal models for stream networks. *Comput. Statist. Data Anal.* **170** 107446.
- SCHLATHER, M. (2002). Models for stationary max-stable random fields. *Extremes* **5** 33–44.
- SHARMA, A., WASKO, C. and LETTENMAIER, D. P. (2018). If precipitation extremes are increasing, why aren't floods? *Water Resour. Res.* **54** 8545–8551.
- SMITH, R. L. (1990). Max-stable processes and spatial extremes. Unpublished manuscript.
- SRAJ, M., VIGLIONE, A., PARAJKA, J. and BLÖSCHL, G. (2016). The influence of non-stationarity in extreme hydrological events on flood frequency estimation. *J. Hydrol. Hydromech.* **64** 426–437.
- STEIN, M. L., CHI, Z. and WELTY, L. J. (2004b). Approximating likelihoods for large spatial data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66** 275–296.
- TAWN, J. A. (1990). Modelling multivariate extreme value distributions. *Biometrika* **77** 245–253.
- VECCHIA, A. V. (1988b). Estimation and model identification for continuous spatial processes. *J. Roy. Statist. Soc. Ser. B* **50** 297–312.
- VEHTARI, A., GELMAN, A. and GABRY, J. (2016). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **27** 1413–1432. <https://doi.org/10.1007/s11222-016-9696-4>
- VOGEL, R. M., YAINDL, C. and WALTER, M. (2011). Nonstationarity: Flood magnification and recurrence reduction factors in the United States. *J. Am. Water Resour. Assoc.* **47** 464–474.
- WADSWORTH, J. L. (2015). On the occurrence times of componentwise maxima and bias in likelihood inference for multivariate max-stable distributions. *Biometrika* **102** 705–711. [MRMR3394285 https://doi.org/10.1093/biomet/asv029](https://doi.org/10.1093/biomet/asv029)
- WADSWORTH, J. L. and TAWN, J. A. (2012b). Dependence modelling for spatial extremes. *Biometrika* **99** 253–272.
- WADSWORTH, J. L. and TAWN, J. A. (2014). Efficient inference for spatial extreme value processes associated to log-Gaussian random functions. *Biometrika* **101** 1–15. [MRMR3180654 https://doi.org/10.1093/biomet/ast042](https://doi.org/10.1093/biomet/ast042)

- WALTER, M. (2010). Increasing trends in peak flows in the northeastern United States and their impacts on design. Ph.D. thesis Tufts Univ.
- WANG, H. and LI, J. (2018b). Adaptive Gaussian process approximation for Bayesian inference with expensive likelihood functions. *Neural Comput.* **30** 3072–3094.
- WANG, Y. and STOEV, S. A. (2010). On the structure and representations of max-stable processes. *Adv. in Appl. Probab.* **42** 855–877. <https://doi.org/10.1239/aap/1282924066>
- WATANABE, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J. Mach. Learn. Res.* **11** 3571–3594.
- WILKINSON, R. (2014). Accelerating ABC methods using Gaussian processes. In *Artificial Intelligence and Statistics* 1015–1023. PMLR.
- WINSEMIUS, H. C., JONGMAN, B., VELDKAMP, T. I., HALLEGATTE, S., BANGALORE, M. and WARD, P. J. (2018). Disaster risk, climate change, and poverty: Assessing the global exposure of poor people to floods and droughts. *Environ. Dev. Econ.* **23** 328–348.
- XU, S. and MAJUMDER, R. (2022). SPQR: Semi-Parametric Quantile Regression. R package version 0.1.0.
- XU, S. G. and REICH, B. J. (2021). Bayesian nonparametric quantile process regression and estimation of marginal quantile effects. *Biometrics* **00** 1–14. <https://doi.org/10.1111/biom.13576>