Journal of the American Statistical Association

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/uasa20

Distributed Inference for Spatial Extremes Modeling in High Dimensions

Emily C. Hector & Brian J. Reich

To cite this article: Emily C. Hector & Brian J. Reich (13 Apr 2023): Distributed Inference for Spatial Extremes Modeling in High Dimensions, Journal of the American Statistical Association, DOI: 10.1080/01621459.2023.2186886

To link to this article: https://doi.org/10.1080/01621459.2023.2186886

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2023, VOL. 00, NO. 0, 1–12
https://doi.org/10.1080/01621459.2023.2186886

Taylor & Francis
Taylor & Francis Group

THEORY AND METHODS

Distributed Inference for Spatial Extremes Modeling in High Dimensions

Emily C. Hector and Brian J. Reich

Department of Statistics, North Carolina State University, Raleigh, NC

ABSTRACT

Extreme environmental events frequently exhibit spatial and temporal dependence. These data are often modeled using Max Stable Processes (MSPs) that are computationally prohibitive to fit for as few as a dozen observations. Supposed computationally-efficient approaches like the composite likelihood remain computationally burdensome with a few hundred observations. In this article, we propose a spatial partitioning approach based on local modeling of subsets of the spatial domain that delivers computationally and statistically efficient inference. Marginal and dependence parameters of the MSP are estimated locally on subsets of observations using censored pairwise composite likelihood, and combined using a modified generalized method of moments procedure. The proposed distributed approach is extended to estimate inverted MSP models, and to estimate spatially varying coefficient models to deliver computationally efficient modeling of spatial variation in marginal parameters. We demonstrate consistency and asymptotic normality of estimators, and show empirically that our approach leads to statistically efficient estimation of model parameters. We illustrate the flexibility and practicability of our approach through simulations and the analysis of streamflow data from the U.S. Geological Survey. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received May 2022
Accepted February 2023

KEYWORDS
Bias-variance tradeoff;
Brown-Resnick process;
Divide-and-conquer;
Scalable computing

1. Introduction

The modeling of spatial extremes using Max-Stable Processes (MSP) is theoretically and computationally challenging in high spatial dimensions. The main technical challenge lies in adequately capturing spatial dependence using low-dimensional marginal projections of the joint distribution of the spatial extreme outcomes while controlling the computational burden of the analysis as the dimension of these marginal distributions increases (Huser and Wadsworth 2022). To balance these two fundamental necessities, we propose a data partitioning approach that leverages recent advances in divide-and-conquer techniques for dependent outcomes and delivers three new tools for analysis of spatial extremes: (i) a censored pairwise likelihood approach for analysis of spatial extremes when the MSP model is only valid for outcomes above a threshold, (ii) a computationally and statistically efficient divide-and-conquer meta-estimator that integrates censored pairwise likelihood information from the whole spatial domain, and (iii) a flexible analytic toolbox for spatially varying coefficient MSP models in high dimensions.

The MSP models pointwise maxima over infinitely many independent realizations of a spatial process, and provides a flexible modeling class for spatial extremes (de Haan 1984). The extremal tail dependence is specified by an exponent function for which many models have been proposed, such as Smith (1990), Tawn (1990), Schlather (2002), Kabluchko, Schlather, and de Haan (2009), Buishand, de Haan, and Zhou (2008), and Opitz (2013). Theoretical assumptions for MSPs are difficult to satisfy in practice: extreme events are by definition rare and there are often not enough replicates to justify the theoretical approximation to maxima over infinitely many observations. When pointwise maxima are taken over a small to moderate number of replicates, the MSP fit is poor (Huang et al. 2016). A viable solution, censored likelihoods model the dependence between observations above a threshold (Thibaud, Mutzner, and Davison 2013; Huser and Davison 2014). This approach uses the partial information available on the extremal coefficients from points below the threshold and has been used for spatial and nonspatial applications (Ledford and Tawn 1996; Smith, Tawn, and Coles 1997; Bortot, Coles, and Tawn 2000; Coles 2001; Wadsworth and Tawn 2014; Thibaud and Opitz 2015; Huser and Davison 2014).

The computational cost of pairwise censored likelihood methods remains high, and the analysis of extreme values on large spatial domains persists as an open problem. The analytic form of (censored) MSP densities is computationally intractable for all but trivially small spatial fields. Only a few models for the exponent function have computationally tractable bi- or tri-variate densities (Schlather 2002; Kabluchko, Schlather, and de Haan 2009), with higher-order densities typically impossibly complex. The primary difficulty lies in computing the exploding number of partial derivatives of the exponent function. For example, the Brown-Resnick and other max-stable processes have dth order density consisting of B_d terms with B_d the dth Bell number (Wadsworth and Tawn 2014). This has led to the predominant use of the composite likelihood (CL) (Lindsay 1988). The core philosophy of the CL approach is to construct

CONTACT Emily Hector ehector@ncsu.edu Department of Statistics, North Carolina State University, Raleigh, NC.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

marginal likelihoods on subsets of data and integrate them using working independence assumptions; thus, the CL is not a proper likelihood, but a product of proper likelihoods. The most widely used form of the CL, the pairwise CL, assumes pairwise working independence and replaces the likelihood with the product of bivariate likelihoods. The seminal work of Padoan, Ribatet, and Sisson (2010) cemented the pairwise CL as a practical method for inference with MSPs by formally defining the procedure and examining its theoretical properties. Since then, the pairwise and triplewise CL have played a prominent role in computationally attractive methods for inference with MSPs (Genton, Ma, and Sang 2011; Davison and Gholamrezaee 2012; Huser and Davison 2013; Sang and Genton 2014; Castruccio, Huser, and Genton 2016; Huser and Genton 2016).

The pairwise CL is attractive because it offers a tradeoff between statistical efficiency and computational speed. Moreover, the maximum CL estimator is consistent and asymptotically normal under mild regularity conditions (Padoan, Ribatet, and Sisson 2010). The pairwise CL still suffers from loss of efficiency that is particularly evident for large dimensions (Huser, Davison, and Genton 2016). In addition, for d observation locations, the pairwise CL evaluates bivariate densities for $\binom{d}{2} = O(d^2)$ pairs of observations, which may become computationally burdensome for $d \gtrsim 10^2$. A common solution is to use tapered pairwise CL, which only includes t pairs up to a predefined distance. The number of pairs of observations is then $O(td)$, which may become computationally burdensome for $d \gtrsim 10^3$.

Spatially-varying coefficient models that allow marginal parameters to vary by observation location are essential for modeling the spatial distribution of extreme events. While spatially varying coefficient MSP model fitting tools are available in R packages (Ribatet 2015), these have not been investigated in the CL literature except under a working independence model (Sass, Li, and Reich 2021), presumably due to the tremendous computational burden of estimating a large number of parameters with pairwise CL. A method that simultaneously analyzes spatially varying coefficients and asymptotic spatial dependence is desirable to provide more efficient parameter estimation and estimates of joint exceedences across multiple spatial locations. A new strategy capable of handling spatially varying coefficients that makes use of dependence between some pairs of observations is desired.

We propose a new local model building approach for spatial extreme value analysis that constructs censored pairwise CL on subsets of spatial observations and integrates these dependent CLs using a modified Generalized Method of Moments (GMM) objective function (Hansen 1982). The resulting integrated censored pairwise CL estimator is statistically and computationally efficient. Our approach hinges on two key observations for the construction of the GMM weight matrix: (i) the optimal choice of the weight matrix is the sample covariance matrix of the pairwise composite score functions, which yields an estimator with variance at least as small as any other estimator constructed from the same pairwise score functions; (ii) the variability in the sample covariance weight matrix introduces finite sample bias in the integrated estimator. To tradeoff between the desire for both optimal efficiency and reduced bias, we propose a new weighting matrix that strikes a balance between these two goals, and show how the resulting estimator can be estimated using a computationally appealing meta-estimator implemented in the MapReduce paradigm. We extend this approach to inverted MSPs and spatially varying marginal regression models for added modeling flexibility. We show through simulations that our approach's tremendous computational advantage enables MSP and inverted MSP inference with potentially thousands of spatially dependent extreme value observations.

We review the MSP construction in Section 2. In Section 3, we describe the proposed data partitioning and censored pairwise CL integration approach. Spatially-varying coefficient models are considered in Section 4. The finite sample performance of the proposed estimator is investigated through simulations in Section 5. An analysis of flood frequency data from the U.S. Geological Survey is presented in Section 6. Derivations, theorem conditions, proofs, extension to subasymptotic models including the inverted MSP, additional simulation and data analysis results and an R package are available in the supplementary materials.

2. Problem Set-Up

2.1. The Max-Stable Process

Let $\mathcal{S} \subset \mathbb{R}^2$ a spatial domain and $Y_r(\boldsymbol{s})$ the outcome value at location $\boldsymbol{s} \in \mathcal{S}$ for replicate $r \in \{1, \ldots, m\}$. Assume that $Y(\boldsymbol{s})$ is the block-maximum, that is, $Y(\boldsymbol{s}) = \max\{Y_1(\boldsymbol{s}), \ldots, Y_m(\boldsymbol{s})\}$. Considering the joint distribution of the point-wise maximum of the m realizations at all locations in $\mathcal{S}$ gives the random field $\mathcal{Y} = \{Y(\boldsymbol{s}); \boldsymbol{s} \in \mathcal{S}\}$. Under regularity conditions, $\mathcal{Y}$ can be well approximated by a max-stable process (MSP) for large m; see also the excellent reviews of Ribatet (2017) and Davison, Huser, and Thibaud (2019). For simplicity of exposition, we describe our approach in the context of the MSP model, but note that the approach can be extended to alternative dependence models described in detail in the supplementary material. Notably, since the MSP is an asymptotically justified model, the MSP approximation may not always hold in practice, especially in data scarce settings. Section 3.2 proposes a thresholding approach to mitigate this issue; the supplementary materials develops further modeling strategies.

Assuming the process is max-stable, then the marginal distribution of $Y(\boldsymbol{s})$ is the Generalized Extreme Value (GEV) distribution $\text{GEV}\{\mu(\boldsymbol{s}), \sigma(\boldsymbol{s}), \xi(\boldsymbol{s})\}$, where $\mu(\boldsymbol{s})$ is the location, $\sigma(\boldsymbol{s}) > 0$ is the scale, and $\xi(\boldsymbol{s})$ is the shape. The GEV parameters can vary spatially to capture local differences in the magnitude of extremes. The MSP can be written equivalently as $Y(\boldsymbol{s}) = \mu(\boldsymbol{s}) + (\sigma(\boldsymbol{s})/\xi(\boldsymbol{s}))[X(\boldsymbol{s})^{\xi(\boldsymbol{s})} - 1]$, where $\mathcal{X} = \{X(\boldsymbol{s}); \boldsymbol{s} \in \mathcal{S}\}$ is a MSP with unit Fréchet margins, $X(\boldsymbol{s}) \sim \text{GEV}(1, 1, 1)$. The three GEV parameters explain spatial variation in the marginal distribution, whereas the spatial dependence of $\mathcal{X}$ explains residual variation. For example, if $Y(\boldsymbol{s})$ is the annual maximum ($m = 365$) of daily precipitation at $\boldsymbol{s}$, then $\mu(\boldsymbol{s})$, $\sigma(\boldsymbol{s})$ and $\xi(\boldsymbol{s})$ determine the distribution of the annual maximum across years at $\boldsymbol{s}$, whereas the spatial dependence of $\mathcal{X}$ determines the likelihood that two locations will simultaneously experience an above average rainfall amount in a given year.

The finite-dimensional distribution function of any MSP at d locations $\mathcal{D} = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_d\}$ has the form $\text{Prob}\{X(\boldsymbol{s}_i) < x_i, i = 1, \ldots, d\} = \exp\{-V(x_1, \ldots, x_d)\}$ for some exponent function V that satisfies $V(\lambda x_1, \ldots, \lambda x_d) = V(x_1, \ldots, x_d)/\lambda$ for any $x_1, \ldots, x_d, \lambda > 0$. Under the assumption that $X(\mathcal{D}) = \{X(\boldsymbol{s}_1), \ldots, X(\boldsymbol{s}_d)\}$ has unit Fréchet marginal distributions, then the exponent function must satisfy $V(x_1, \ldots, x_d) = 1/x_j$ if $x_i = \infty$ for all $i \neq j$. Of the many possibilities for the exponent function, we choose the Brown-Resnick model (Brown and Resnick 1977; Kabluchko, Schlather, and de Haan 2009) because it gives a stationary process, provides flexibility in modeling the smoothness of $\mathcal{X}$ across space and it can capture independence at large distances. Let Φ the standard normal distribution function. The exponent function that defines the joint distribution function of the pair $X(\boldsymbol{s}_i)$ and $X(\boldsymbol{s}_j)$ is $V(x_i, x_j) = x_i^{-1}\Phi\{a_{ij}/2 - a_{ij}^{-1}\log(x_i/x_j)\} + x_j^{-1}\Phi\{a_{ij}/2 - a_{ij}^{-1}\log(x_j/x_i)\}$, where $a_{ij} = \left\{2\gamma(\boldsymbol{s}_i - \boldsymbol{s}_j)\right\}^{1/2}$ and γ is a semivariogram. Following Huser and Davison (2013), we use the isotropic semivariogram $\gamma(\boldsymbol{s}_i - \boldsymbol{s}_j) = (||\boldsymbol{s}_i - \boldsymbol{s}_j||/\phi)^\alpha$ defined by spatial range $\phi > 0$ and smoothness $\alpha \in [0, 2]$, and denote $V(x_i, x_j; \alpha, \phi) \equiv V(x_i, x_j)$. Our objectives are to estimate the GEV parameters $\mu(\boldsymbol{s})$, $\sigma(\boldsymbol{s})$ and $\xi(\boldsymbol{s})$ and the spatial dependence parameters ϕ and α when the number d of observation locations is large.

2.2. Model and Existing Approaches

We consider the setting with n independent replicates of $Y(\mathcal{D})$ denoted by $Y_1(\mathcal{D}), \ldots, Y_n(\mathcal{D})$, where $Y_i(\mathcal{D}) = \{Y_i(\boldsymbol{s}_1), \ldots, Y_i(\boldsymbol{s}_d)\}^\top$ and $\mathcal{D} = \{\boldsymbol{s}_j\}_{j=1}^d$ the set of observation locations. Correspondingly, we have n independent replicates of $X(\mathcal{D})$ denoted by $X_1(\mathcal{D}), \ldots, X_n(\mathcal{D})$, where $X_i(\mathcal{D}) = \{X_i(\boldsymbol{s}_1), \ldots, X_i(\boldsymbol{s}_d)\}^\top$ is related to $Y_i(\mathcal{D})$, $i = 1, \ldots, n$, through

$$Y_i(\boldsymbol{s}_j) = \mu_i(\boldsymbol{s}_j) + \frac{\sigma_i(\boldsymbol{s}_j)}{\xi_i(\boldsymbol{s}_j)}\left\{X_i(\boldsymbol{s}_j)^{\xi_i(\boldsymbol{s}_j)} - 1\right\}, \tag{1}$$

for $j = 1, \ldots, d$. Let $\boldsymbol{z}_i(\cdot)$ be q explanatory variables observed at locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_d$, $i \in \{1, \ldots, n\}$. For $\boldsymbol{z}_{i1}, \boldsymbol{z}_{i2}, \boldsymbol{z}_{i3} \subseteq \boldsymbol{z}_i$ of respective dimensions q_1, q_2, q_3, we posit the model

$$\mu_i(\boldsymbol{s}; \boldsymbol{\beta}_1) = \boldsymbol{z}_{i1}(\boldsymbol{s})^\top\boldsymbol{\beta}_1, \quad \sigma_i(\boldsymbol{s}; \boldsymbol{\beta}_2) = \exp\{\boldsymbol{z}_{i2}(\boldsymbol{s})^\top\boldsymbol{\beta}_2\},$$
$$\xi_i(\boldsymbol{s}; \boldsymbol{\beta}_3) = \boldsymbol{z}_{i3}(\boldsymbol{s})^\top\boldsymbol{\beta}_3. \tag{2}$$

A more flexible spatially-varying coefficient model is introduced in Section 4, where $\boldsymbol{z}_{i1}(\cdot)$ and $\boldsymbol{z}_{i2}(\cdot)$ correspond to radial basis functions. Let $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top, \boldsymbol{\beta}_3^\top)^\top$. To facilitate estimation of α, ϕ, we propose the reparametrization $\omega = \log\{\alpha/(2 - \alpha)\}$, $\zeta = \log(\phi)$, or equivalently $\alpha = 2\exp(\omega)/\{1 + \exp(\omega)\}$, $\phi = \exp(\zeta)$, and let $\boldsymbol{\theta} = (\omega, \zeta, \boldsymbol{\beta}^\top)^\top \in \mathbb{R}^p$. The analytic goal is to estimate and make inference on $\boldsymbol{\theta}$.

When $d = 2$ or 3, estimation and inference on $\boldsymbol{\theta}$ using maximum likelihood is feasible since the full likelihood of the Brown-Resnick process has a closed form following Huser and Davison (2013) and Ribatet (2017); see also the supplementary materials. For $d > 3$, however, the full likelihood generally becomes computationally burdensome to evaluate. For general MSPs, Castruccio, Huser, and Genton (2016) stated that full likelihood inference seemed limited to $d = 12$ or 13 by then-current technologies. More recently, Huser et al. (2019) proposed an expectation-maximization algorithm for full likelihood inference for $d > 13$. Their approach, however, remains computationally prohibitive for large d due to the evaluation of multivariate Gaussian probabilities, with computation time of 19.8 hr when $d = 20$. Some approaches, such as the combined score equation of Wang et al. (2021), make simplifying assumptions on the working spatial dependence structure to reduce the computational burden of estimation. Some Bayesian approaches have been proposed for large d, although these remain computationally burdensome; see for example Stephenson et al. (2015) whose model on $d = 17{,}000$ sites takes approximately 50 hr to run. Lenzi et al. (2021) uses simulated data to train a deep convolutional neural network to learn a map between data and parameters. Their approach remains at the proof-of-concept stage; one potential bottleneck is the need to simulate large amounts of data for a broad set of potential parameter values, a problem further exacerbated with spatially-varying coefficient models.

Composite likelihood (CL) (Lindsay 1988) has therefore become the method of choice to overcome the computational burden of full likelihood inference. Denote $\{\mathcal{D}_1, \ldots, \mathcal{D}_K\}$ a collection of subsets of $\mathcal{D}$ such that $\mathcal{D} = \cup_{k=1}^K \mathcal{D}_k$ and $\mathcal{D}_j \cap \mathcal{D}_k$ not necessarily empty for $j \neq k$. Denoting $y_i(\mathcal{D}) = \{y_i(\boldsymbol{s}_j)\}_{j=1}^d$, the log CL assumes working independence between observations in different sets $\mathcal{D}_k$ and takes the form

$$\mathcal{CL}\{\boldsymbol{\theta}; y_i(\mathcal{D}), i = 1, \ldots, n\} = \log\prod_{i=1}^n\prod_{k=1}^K f\{y_i(\mathcal{D}_k); \boldsymbol{\theta}\}$$
$$= \sum_{i=1}^n\sum_{k=1}^K \log f\{y_i(\mathcal{D}_k); \boldsymbol{\theta}\},$$

where $f\{y_i(\mathcal{D}_k); \boldsymbol{\theta}\}$ is the multivariate marginal density of $y_i(\mathcal{D}_k) = \{y_i(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D}_k\}$. The pairwise CL has cardinality $|\mathcal{D}_k| = 2$ of $\mathcal{D}_k$, $K = \binom{d}{2}$, and has been widely used in spatial extreme value analysis; see for example the review of Davison, Padoan, and Ribatet (2012) and references in Section 1. See also Huser, Davison, and Genton (2016), who conducted a thorough empirical study of bias and variance with various CL-based methods.

The pairwise CL remains computationally burdensome when d is large. This difficulty stems from the need to evaluate analytically complex bivariate likelihoods at all pairs of observations. The pairwise CL is not scalable and alternative strategies are required.

3. A Spatial Partitioning Approach

3.1. Partitioning the Spatial Domain

Following the spirit of the CL, we propose a partition of the spatial domain $\mathcal{D}$ into K disjoint regions $\mathcal{D}_1, \ldots, \mathcal{D}_K$ such that $\cup_{k=1}^K \mathcal{D}_k = \mathcal{D}$ and denote by d_k the number of observation locations in $\mathcal{D}_k$. To facilitate estimation of $\boldsymbol{\theta}$ in each subset $\mathcal{D}_k$, we partition $\mathcal{D}$ such that d_k is relatively small, for example, $d_k = 25$; see the supplementary materials for an example partition of a 400-dimensional domain. While disjoint regions $\mathcal{D}_1, \ldots, \mathcal{D}_K$ are not technically required, overlapping regions may increase dependence between regions and incur numerical instability at the integration step; this is discussed further in Section 3.3. The literature is rich with methods for choosing

partitions for Gaussian processes; see Heaton et al. (2019). With MSPs, due to the range parameter ϕ, we generally recommend regions of similar size d_k based on nearest locations with d_k large enough to allow for a range of distances. Many of the spatial partitioning approaches reviewed in Heaton et al. (2019) assume independence between $\mathcal{D}_k$, which we do not. The local likelihood approach for threshold exceedances of Castro-Camilo and Huser (2020) resembles the first step of our approach, but the models bear substantial differences and the authors focus on dependence parameter estimation.

3.2. Local Likelihood Specification

Let $k \in \{1,\dots,K\}$. Since d_k may still be large enough to render full likelihood estimation of $\boldsymbol{\theta}$ in $\mathcal{D}_k$ intractable, we propose to estimate $\boldsymbol{\theta}$ in $\mathcal{D}_k$ using the pairwise CL. Inference with pairwise CL for max-stable processes has been ubiquitous since the seminal paper of Padoan, Ribatet, and Sisson (2010), but this approach assumes the MSP is an appropriate model for $Y_i(\boldsymbol{s})$, $i = 1,\dots,n$, which may not hold in practice if the block maxima were taken over blocks with small size m. Huser and Davison (2014) observed that the MSP defined in Section 2.1 also models extremes of individual observations, and used a censored likelihood approach for bivariate extremes in the CL framework to overcome this difficulty.

Inspired by their approach, consider two locations $\boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathcal{D}_k$ and denote $y_{ij} = y_i(\boldsymbol{s}_j)$, $x_{ij} = x_i(\boldsymbol{s}_j)$, $j = 1,2$. Let $u_1 = u(\boldsymbol{s}_1), u_2 = u(\boldsymbol{s}_2)$ be sufficiently high thresholds such that $f(y_{i1}, y_{i2}; \boldsymbol{\theta})$ is a valid model for $y_{i1}, y_{i2} \in \{y_i(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D}_k\}$ when $y_{i1} > u_1, y_{i2} > u_2$, $i = 1,\dots,n$. Here, $f(y_{i1}, y_{i2}; \boldsymbol{\theta}) = f(x_{i1}, x_{i2}; \alpha, \phi) J(y_{i1}; \boldsymbol{\beta}) J(y_{i2}; \boldsymbol{\beta})$ is the bivariate max-stable density obtained from the MSP defined in Section 2.1, with Jacobians $J(y_{i1}; \boldsymbol{\beta})$, $J(y_{i2}; \boldsymbol{\beta})$ and $f(x_{i1}, x_{i2}; \alpha, \phi) = \partial^2 \exp\{-V(x_{i1}, x_{i2}; \alpha, \phi)\}/(\partial x_{i1} \partial x_{i2})$ given in the supplementary materials. We remark that, when insufficient data are available, it may not be possible to select a threshold high enough for the MSP approximation to be valid while simultaneously observing outcomes above this threshold. In this case, we may require alternative subasymptotic models; these are considered in the supplementary material, where we also detail the inverted MSP.

Let $u_{ij} = \{1 + \xi_i(\boldsymbol{s}_j)(u_j - \mu_i(\boldsymbol{s}_j)/\sigma_i(\boldsymbol{s}_j)\}^{1/\xi_i(\boldsymbol{s}_j)}$, $j = 1,2$. The likelihood contribution $g(y_{i1}, y_{i2}; \boldsymbol{\theta}, u_1, u_2)$ of the pair (y_{i1}, y_{i2}) for $\boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathcal{D}_k$ is

$$\begin{cases} f(x_{i1}, x_{i2}; \alpha, \phi) J(y_{i1}; \boldsymbol{\beta}) J(y_{i2}; \boldsymbol{\beta}), & y_{i1} > u_1, y_{i2} > u_2, \\ \left[\frac{\partial}{\partial x_{i1}} \exp\{-V(x_{i1}, u_{i2}; \alpha, \phi)\}\right] J(y_{i1}; \boldsymbol{\beta}), & y_{i1} > u_1, y_{i2} \le u_2, \\ \left[\frac{\partial}{\partial x_{i2}} \exp\{-V(u_{i1}, x_{i2}; \alpha, \phi)\}\right] J(y_{i2}; \boldsymbol{\beta}), & y_{i1} \le u_1, y_{i2} > u_2, \\ \exp\{-V(u_{i1}, u_{i2}; \alpha, \phi)\}, & y_{i1} \le u_1, y_{i2} \le u_2. \end{cases}$$

Define $\mathcal{P}_k = \{(\boldsymbol{s}_1, \boldsymbol{s}_2) : \boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathcal{D}_k, \boldsymbol{s}_1 \neq \boldsymbol{s}_2\}$. Using the censored likelihood pairs, the log censored CL (CCL) in $\mathcal{D}_k$ takes the form $\mathcal{CCL}_k(\boldsymbol{\theta}; u) = (1/n) \sum_{i=1}^n \sum_{\mathcal{P}_k} \log g(y_{i1}, y_{i2}; \boldsymbol{\theta}, u_1, u_2)$. Clearly, letting $u_1, u_2 \to -\infty$ recovers the uncensored CL. Here, the CCL uses all locations in region $\mathcal{D}_k$. In practice, pairs may be weighted to achieve higher statistical efficiency, although d_k should be small enough that only marginal efficiency gains can be expected from this approach. We obtain the censored composite score function:

$$\begin{aligned} \Psi_k(\boldsymbol{\theta}) &= \frac{1}{n} \sum_{i=1}^n \boldsymbol{\psi}_{ik}(\boldsymbol{\theta}) \\ &= \frac{1}{n} \left(\frac{\partial}{\partial \omega}, \frac{\partial}{\partial \zeta}, \frac{\partial}{\partial \boldsymbol{\beta}_1}, \frac{\partial}{\partial \boldsymbol{\beta}_2}, \frac{\partial}{\partial \boldsymbol{\beta}_3} \right)^\top \mathcal{CCL}_k(\boldsymbol{\theta}) \in \mathbb{R}^p, \quad (3) \end{aligned}$$

with specific form given in the supplementary materials. Solving $\Psi_k(\boldsymbol{\theta}) = \mathbf{0}$ yields the maximum CCL estimator (MCCLE) $\widehat{\boldsymbol{\theta}}_k = (\widehat{\omega}_k, \widehat{\zeta}_k, \widehat{\boldsymbol{\beta}}_{1k}, \widehat{\boldsymbol{\beta}}_{2k}, \widehat{\boldsymbol{\beta}}_{3k})^\top$ of $\boldsymbol{\theta}$. We denote by $\boldsymbol{i}_k(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}[\nabla_{\boldsymbol{\theta}}\{\Psi_k(\boldsymbol{\theta})\}]$ and $\boldsymbol{c}_k(\boldsymbol{\theta}) = \text{var}_{\boldsymbol{\theta}}\{\sqrt{n}\Psi_k(\boldsymbol{\theta})\}$ the sensitivity and variability matrices, respectively, of $\Psi_k(\boldsymbol{\theta})$. Under mild regularity conditions (C1), it follows from Lindsay (1988), Padoan, Ribatet, and Sisson (2010), Huser and Davison (2014) that $\sqrt{n}(\widehat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}\{\mathbf{0}, \boldsymbol{i}_k^\top(\boldsymbol{\theta}_0) \boldsymbol{c}_k^{-1}(\boldsymbol{\theta}_0) \boldsymbol{i}_k(\boldsymbol{\theta}_0)\}$, where $\boldsymbol{\theta}_0 = (\omega_0, \zeta_0, \boldsymbol{\beta}_0^\top)^\top$ is the true value of $\boldsymbol{\theta}$ such that $E_{\boldsymbol{\theta}}\{\Psi_k(\boldsymbol{\theta})\}$ is uniquely zero at $\boldsymbol{\theta}_0$.

3.3. Censored Composite Likelihood Integration

Suppose we have successfully obtained the K MCCLEs $\{\widehat{\boldsymbol{\theta}}_k\}_{k=1}^K$ for regions $\mathcal{D}_k$, $k = 1,\dots,K$. We now wish to integrate these local estimators into one unified estimator of $\boldsymbol{\theta}$ over all K regions. Huser and Davison (2013) show that the MCCLEs are less efficient than even trivariate CL estimators, let alone maximum likelihood estimators. Thus, it is desirable to incorporate as much spatial dependence for improved statistical efficiency without incurring an undesirable computational burden. An efficient model integration procedure should therefore leverage the dependence between the K MCCLEs, but this dependence is difficult to estimate directly because we do not have replicates of these estimators. Alternative bootstrap-type procedures to estimate this dependence are computationally costly. To overcome this difficulty, we integrate the censored composite score functions in equation (3) rather than the MCCLEs.

Define the stacking operation $\{\boldsymbol{a}_k\}_{k=1}^K = (\boldsymbol{a}_1^\top, \dots, \boldsymbol{a}_K)^\top \in \mathbb{R}^{\sum_{k=1}^K b_k}$ and $\{\boldsymbol{A}_k\}_{k=1}^K = (\boldsymbol{A}_1^\top, \dots, \boldsymbol{A}_K^\top)^\top \in \mathbb{R}^{\sum_{k=1}^K b_{1k} \times b_2}$ for $\boldsymbol{a}_k \in \mathbb{R}^{b_k}$ and $\boldsymbol{A}_k \in \mathbb{R}^{b_{1k} \times b_2}$. Define the stacked CCL kernel and score functions $\boldsymbol{\psi}_{i,\text{all}}(\boldsymbol{\theta}) = \{\boldsymbol{\psi}_{ik}(\boldsymbol{\theta})\}_{k=1}^K \in \mathbb{R}^{Kp}$ and $\Psi_{all}(\boldsymbol{\theta}) = \{\Psi_k(\boldsymbol{\theta})\}_{k=1}^K \in \mathbb{R}^{pK}$, respectively. Denote the sensitivity and variability matrices of $\Psi_{all}(\boldsymbol{\theta})$ by $\boldsymbol{i}(\boldsymbol{\theta}) = \{\boldsymbol{i}_k(\boldsymbol{\theta})\}_{k=1}^K \in \mathbb{R}^{Kp \times p}$ and $\boldsymbol{c}(\boldsymbol{\theta}) = \text{var}_{\boldsymbol{\theta}}\{\sqrt{n}\Psi_{all}(\boldsymbol{\theta})\} \in \mathbb{R}^{Kp \times Kp}$, respectively. A key insight is that the stacked censored composite score function $\Psi_{all}(\boldsymbol{\theta})$ over-identifies $\boldsymbol{\theta}$: there are more estimating equations than there are dimensions on $\boldsymbol{\theta}$. Hansen's (1982) Generalized Method of Moments (GMM) minimizes a quadratic form of the over-identifying moment conditions:

$$\begin{aligned} \widehat{\boldsymbol{\theta}}_{\text{GMM}} &= \arg\min_{\boldsymbol{\theta}} n \Psi_{all}^\top(\boldsymbol{\theta}) \boldsymbol{W} \Psi_{all}(\boldsymbol{\theta}) \\ &= \arg\min_{\boldsymbol{\theta}} n \sum_{k,k'=1}^K \Psi_k(\boldsymbol{\theta}) \, (\boldsymbol{W})_{k,k'} \, \Psi_{k'}(\boldsymbol{\theta}), \quad (4) \end{aligned}$$

where $(\boldsymbol{W})_{k,k'}$ denotes the rows and columns of $\boldsymbol{W}$ corresponding to subsets $\mathcal{D}_k$ and $\mathcal{D}_{k'}$, respectively, for any positive semidefinite weight matrix $\boldsymbol{W}$. This approach has been successfully employed by others (Hector and Song 2021) although never with a MSP or censored (composite) likelihood, and

has connections to weighted CL (Sang and Genton 2014; Castruccio, Huser, and Genton 2016). Under mild regularity conditions (C1) and (C2), $\widehat{\boldsymbol{\theta}}_{\mathrm{GMM}}$ is a consistent estimator of $\boldsymbol{\theta}$ and asymptotically normally distributed as $n \to \infty$: $\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\mathrm{GMM}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}\{0, \Omega(\boldsymbol{\theta}_0)\boldsymbol{c}(\boldsymbol{\theta}_0)\Omega^\top(\boldsymbol{\theta}_0)\}$, where $\Omega(\boldsymbol{\theta}_0) = -\{\boldsymbol{i}^\top(\boldsymbol{\theta}_0)\boldsymbol{W}\boldsymbol{i}(\boldsymbol{\theta}_0)\}^{-1}\boldsymbol{i}^\top(\boldsymbol{\theta}_0)\boldsymbol{W}$. From the presence of $\boldsymbol{c}(\boldsymbol{\theta})$ in the asymptotic variance of $\widehat{\boldsymbol{\theta}}_{\mathrm{GMM}}$, dependence between the K spatial subsets $\{\mathcal{D}_k\}_{k=1}^K$ is incorporated in the evaluation of the estimator's uncertainty. Thus, the GMM estimator is not evaluated under working independence assumptions, and the uncertainty quantification of $\widehat{\boldsymbol{\theta}}_{\mathrm{GMM}}$ is robust to the form of the between-subset dependence.

Following Hansen (1982), the most efficient choice of $\boldsymbol{W}$ is clearly $\boldsymbol{c}^{-1}(\boldsymbol{\theta}_0)$, which minimizes the diagonal of $\Omega(\boldsymbol{\theta}_0)\boldsymbol{c}(\boldsymbol{\theta}_0)\Omega^\top(\boldsymbol{\theta}_0)$. This choice is equivalent to using all the dependence between spatial subsets $\{\mathcal{D}_k\}_{k=1}^K$. On the other hand, when d is large, for example, $d \gtrsim 100$, estimation of $\boldsymbol{c}(\boldsymbol{\theta}_0)$ based on the sample covariance matrix $\boldsymbol{C}(\boldsymbol{\theta}) = (1/n)\sum_{i=1}^n \boldsymbol{\psi}_{i,\mathrm{all}}(\boldsymbol{\theta})\boldsymbol{\psi}_{i,\mathrm{all}}^\top(\boldsymbol{\theta})$ may introduce bias into $\widehat{\boldsymbol{\theta}}_{\mathrm{GMM}}$ due to variability in the estimation of $\boldsymbol{c}(\boldsymbol{\theta}_0)$. To overcome this difficulty, we use the fact that, jointly, $\sqrt{n}(\widehat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_0)_{k=1}^K \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \boldsymbol{s}^\top(\boldsymbol{\theta}_0)\boldsymbol{c}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{s}(\boldsymbol{\theta}_0))$ (Hector and Song 2021), and so, marginally, $\sqrt{n}(\widehat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}\{\boldsymbol{0}, \boldsymbol{i}_k^\top(\boldsymbol{\theta}_0)\boldsymbol{w}_k(\boldsymbol{\theta}_0)\boldsymbol{i}_k(\boldsymbol{\theta}_0)\}$, with $\boldsymbol{w}_k(\boldsymbol{\theta}) = \{\boldsymbol{c}^{-1}(\boldsymbol{\theta})\}_{k,k}$. This motivates our choice of a weight matrix $\boldsymbol{W}$ that mitigates the effect of covariance between subsets on the finite-sample bias of the GMM estimator. We propose $\boldsymbol{W}(\boldsymbol{\theta}) = \mathrm{diag}\{\boldsymbol{W}_k(\boldsymbol{\theta})\}_{k=1}^K$, with $\boldsymbol{W}_k(\boldsymbol{\theta}) = \{\boldsymbol{C}^{-1}(\boldsymbol{\theta})\}_{k,k} \in \mathbb{R}^{p\times p}$ to obtain the GMM estimator,

$$\widehat{\boldsymbol{\theta}}^{\star}_{\mathrm{GMM}} = \arg\min_{\boldsymbol{\theta}} n \sum_{k=1}^K \Psi_k^\top(\boldsymbol{\theta})\boldsymbol{W}_k(\boldsymbol{\theta})\Psi_k(\boldsymbol{\theta}). \tag{5}$$

Theorem 1. Under regularity conditions (C1) and (C2), the GMM estimator $\widehat{\boldsymbol{\theta}}^{\star}_{\mathrm{GMM}}$ in (5) satisfies $\sqrt{n}(\widehat{\boldsymbol{\theta}}^{\star}_{\mathrm{GMM}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}[\boldsymbol{0}, \{\boldsymbol{h}(\boldsymbol{\theta}_0)\}^{-1}\boldsymbol{g}(\boldsymbol{\theta}_0)\{\boldsymbol{h}^\top(\boldsymbol{\theta}_0)\}^{-1}]$ as $n \to \infty$,

$$\boldsymbol{h}(\boldsymbol{\theta}) = \sum_{k=1}^K \boldsymbol{i}_k^\top(\boldsymbol{\theta})\boldsymbol{w}_k(\boldsymbol{\theta})\boldsymbol{i}_k(\boldsymbol{\theta}),$$

$$\boldsymbol{g}(\boldsymbol{\theta}) = \sum_{k,k'=1}^K \boldsymbol{i}_k^\top(\boldsymbol{\theta})\boldsymbol{w}_k(\boldsymbol{\theta})\{\boldsymbol{c}(\boldsymbol{\theta})\}_{k,k'}\boldsymbol{w}_{k'}(\boldsymbol{\theta})\boldsymbol{i}_{k'}(\boldsymbol{\theta}).$$

The proof follows immediately from Hansen (1982). The proposed GMM estimator in (5) thus possesses required statistical properties for inference while benefiting from a substantial reduction in computation time. We remark that invertibility of $\boldsymbol{C}(\boldsymbol{\theta})$ may be unstable when regions $\mathcal{D}_1, \ldots, \mathcal{D}_K$ are allowed to overlap, motivating our preference for disjoint subsets. In this case, the Moore–Penrose generalized inverse may be used.

3.4. Implementation: A Meta-Estimator

The iterative minimization in (5) remains computationally burdensome for large d because the censored composite score function of d_k pairs must be evaluated at each iteration of the minimization. Fortunately, this iterative procedure may be altogether bypassed through the closed-form meta-estimator derived by Hector and Song (2021):

$$\widehat{\boldsymbol{\theta}}_m = \left\{\sum_{k=1}^K \boldsymbol{I}_k^\top(\widehat{\boldsymbol{\theta}}_c)\boldsymbol{W}_k(\widehat{\boldsymbol{\theta}}_c)\boldsymbol{I}_k(\widehat{\boldsymbol{\theta}}_c)\right\}^{-1} \sum_{k=1}^K \boldsymbol{I}_k^\top(\widehat{\boldsymbol{\theta}}_c)\boldsymbol{W}_k(\widehat{\boldsymbol{\theta}}_c)\boldsymbol{I}_k(\widehat{\boldsymbol{\theta}}_c)\widehat{\boldsymbol{\theta}}_k \tag{6}$$

where $\boldsymbol{I}_k(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}}\Psi_k(\boldsymbol{\theta})$ denotes the sample sensitivity matrix, and $\widehat{\boldsymbol{\theta}}_c$ is a suitable consistent estimator of $\boldsymbol{\theta}$ specified as follows. The estimation of $\boldsymbol{I}_k(\boldsymbol{\theta})$ and $\boldsymbol{W}_k(\boldsymbol{\theta})$ by $\boldsymbol{I}_k(\widehat{\boldsymbol{\theta}}_c)$ and $\boldsymbol{W}_k(\widehat{\boldsymbol{\theta}}_c)$, respectively, requires careful consideration. These matrices may be estimated by plugging in the MCCLEs, that is, using $\boldsymbol{I}_k(\widehat{\boldsymbol{\theta}}_k)$ and

$$\left\{\left(\frac{1}{n}\sum_{i=1}^n \{\boldsymbol{\psi}_{ik}(\widehat{\boldsymbol{\theta}}_k)\}_{k=1}^K \left[\{\boldsymbol{\psi}_{ik}(\widehat{\boldsymbol{\theta}}_k)\}_{k=1}^K\right]^\top\right)^{-1}\right\}_{k,k},$$

but these estimators may have high variability depending on the performance of the MCCLEs in each subset. A better estimator can be constructed from the average of the MCCLEs: $\widehat{\boldsymbol{\theta}}_c = (1/K)\sum_{k=1}^K \widehat{\boldsymbol{\theta}}_k$. This leads to the following distributed procedure:

1. Partition the spatial domain $\mathcal{D}$ into K disjoint regions $\mathcal{D}_1, \ldots, \mathcal{D}_K$.
2. For $k = 1, \ldots, K$, estimate $\widehat{\boldsymbol{\theta}}_k$ in subset $\mathcal{D}_k$ using the CCL. This step can be performed in parallel on K nodes to accelerate computation.
3. Compute the average of the MCCLEs, $\widehat{\boldsymbol{\theta}}_c = \sum_{k=1}^K \widehat{\boldsymbol{\theta}}_k/K$, on the main computing node.
4. For $k = 1, \ldots, K$, evaluate and return $\boldsymbol{\psi}_{ik}(\widehat{\boldsymbol{\theta}}_c)$ and $\boldsymbol{I}_k(\widehat{\boldsymbol{\theta}}_c)$ to the main computing node. This step can be performed in parallel on K nodes to accelerate computation.
5. Form $\boldsymbol{\psi}_{i,\mathrm{all}}(\widehat{\boldsymbol{\theta}}_c) = \{\boldsymbol{\psi}_{ik}(\widehat{\boldsymbol{\theta}}_c)\}_{k=1}^K$ and compute $\boldsymbol{C}(\widehat{\boldsymbol{\theta}}_c) = (1/n)\sum_{i=1}^n \boldsymbol{\psi}_{i,\mathrm{all}}(\widehat{\boldsymbol{\theta}}_c)\boldsymbol{\psi}_{i,\mathrm{all}}^\top(\widehat{\boldsymbol{\theta}}_c)$, $\boldsymbol{W}_k(\widehat{\boldsymbol{\theta}}_c) = \{\boldsymbol{C}^{-1}(\widehat{\boldsymbol{\theta}}_c)\}_{k,k}$, and $\widehat{\boldsymbol{\theta}}_m$ in (6).

This distributed approach to estimation of $\boldsymbol{\theta}$ requires two rounds of communication between distributed nodes and the main computing node: the first to return $\widehat{\boldsymbol{\theta}}_k$, and the second to return $\boldsymbol{\psi}_{ik}(\widehat{\boldsymbol{\theta}}_c)$ and $\boldsymbol{I}_k(\widehat{\boldsymbol{\theta}}_c)$. The derivative $\boldsymbol{I}_k(\boldsymbol{\theta})$ can be estimated as the sum of the sample covariance of bivariate censored score functions for each pair of observations in $\mathcal{D}_k$. This results in a flexible and computationally efficient procedure. Inferential properties of the estimator $\widehat{\boldsymbol{\theta}}_m$ in (6) are shown in Theorem 2.

Theorem 2. Under conditions (C1) and (C2), the proposed estimator $\widehat{\boldsymbol{\theta}}_m$ in (6) is consistent and asymptotically normally distributed as $n \to \infty$:

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0\right) \xrightarrow{d} \mathcal{N}\left[\boldsymbol{0}, \{\boldsymbol{h}(\boldsymbol{\theta}_0)\}^{-1}\boldsymbol{g}(\boldsymbol{\theta}_0)\left\{\boldsymbol{h}^\top(\boldsymbol{\theta}_0)\right\}^{-1}\right],$$

where the asymptotic covariance of $\widehat{\boldsymbol{\theta}}_m$ in (6) can be consistently estimated by $\boldsymbol{J}^{-1}(\widehat{\boldsymbol{\theta}}_c) = n^{-1}\{\boldsymbol{H}(\widehat{\boldsymbol{\theta}}_c)\}^{-1}\boldsymbol{G}(\widehat{\boldsymbol{\theta}}_c)\{\boldsymbol{H}^\top(\widehat{\boldsymbol{\theta}}_c)\}^{-1}$ with

$$\boldsymbol{H}(\boldsymbol{\theta}) = \sum_{k=1}^K \boldsymbol{I}_k^\top(\boldsymbol{\theta})\boldsymbol{W}_k\boldsymbol{I}_k(\boldsymbol{\theta}),$$

$$\boldsymbol{G}(\boldsymbol{\theta}) = \sum_{k,k'=1}^K \boldsymbol{I}_k^\top(\boldsymbol{\theta})\boldsymbol{W}_k^{-1}(\boldsymbol{\theta})\,[\boldsymbol{C}(\boldsymbol{\theta})]_{k,k'}\,\boldsymbol{W}_{k'}^{-1}(\boldsymbol{\theta})\boldsymbol{I}_{k'}(\boldsymbol{\theta}). \tag{7}$$

The proof of Theorem 2 is a special case of the proofs given in the supplementary materials for the spatially-varying coefficient model (see Section 4). From the form of the asymptotic covariance matrices in Theorems 1 and 2, the meta-estimator $\widehat{\boldsymbol{\theta}}_m$ is asymptotically equivalent to the GMM estimator $\widehat{\boldsymbol{\theta}}^{\star}_{\text{GMM}}$, so that $\widehat{\boldsymbol{\theta}}_m$ loses no statistical efficiency but possesses a substantial computational advantage. Moreover, $\boldsymbol{J}(\boldsymbol{\theta})$ can be computed in a distributed fashion using the quantities returned from the distributed nodes at the second round of communication. In the supplementary material, we find the distribution of $(\widehat{\alpha}_m, \widehat{\boldsymbol{\phi}}_m, \widehat{\boldsymbol{\beta}}_m^\top)^\top$ through the Delta method, and show that the computational complexity of our approach is $O(K^3 + d^2_{\max})$, where $d_{\max} = \max_{k=1,\ldots,K} d_k$, compared to $O(td)$ for the tapered pairwise CL that uses the nearest t pairs, and $O(d)$ for other approaches (Huser, Stein, and Zhong 2022).

4. Extension to Spatially Varying Coefficients

4.1. Local Model and Likelihood Specification

When the spatial variation in $\mu_i(\boldsymbol{s})$ and/or $\sigma_i(\boldsymbol{s})$ is of interest, we propose a spatially varying coefficient model (Hastie and Tibshirani 1993) for added modeling flexibility. We prefer the varying coefficient model over nonparametric kernel smoothing for its ability to fit in our divide-and-conquer framework; see Davison and Ramesh (2002) for a univariate (i.e., nonspatial) nonparametric kernel smoothing approach. Suppose $\mu_i(\cdot)$ depends on $z_{i1,t}(\cdot)$, the tth covariate in $\boldsymbol{z}_{i1}(\cdot) \in \mathbb{R}^{q_1}$ for replicate i, through some unknown function $b_{t,\mu}(\cdot)$, $t = 1,\ldots,q_1$: $\mu_i(\boldsymbol{s}) = \sum_{t=1}^{q_1} z_{i1,t}(\boldsymbol{s}) b_{t,\mu}(\boldsymbol{s})$. Let $\{\phi_{jt,k}(\cdot)\}_{j,k=1}^{\infty,K}$ be radial basis functions of the functional space to which $b_{t,\mu}(\cdot)$ belongs, $t = 1,\ldots,q_1$. Within each spatial subset $\mathcal{D}_k$, we approximate $b_{t,\mu}(\mathcal{D}_k)$ by a finite linear combination of the basis functions, that is, $b_{t,\mu}(\mathcal{D}_k) \approx \sum_{j=1}^{J_{1tk}} \eta_{1jtk}\phi_{jt,k}(\mathcal{D}_k)$, where J_{1tk} is the number of basis functions for the tth covariate function and $\boldsymbol{\eta}_{1k} = \{\eta_{1jtk}\}_{j,t=1}^{J_{1tk},q_1}$ is the unknown $p_{\boldsymbol{\eta}_{1k}}$-dimensional parameter of interest, $p_{\boldsymbol{\eta}_{1k}} = \sum_{t=1}^{q_1} J_{1tk}$ fixed. Defining $\widetilde{z}_{i1,jt}(\boldsymbol{s}) = z_{i1,t}(\boldsymbol{s})\phi_{jt,k}(\boldsymbol{s})$ for $j = 1,\ldots,J_{1tk}$, $t = 1,\ldots,q_1$, $\boldsymbol{s} \in \mathcal{D}_k$, and substituting into the mean model yields $\mu_i(\boldsymbol{s}) \approx \sum_{t=1}^{q_1}\sum_{j=1}^{J_{1tk}} \widetilde{z}_{i1,jt}(\boldsymbol{s})\eta_{1jtk}$ for $\boldsymbol{s} \in \mathcal{D}_k$.

Suppose we propose a similar varying coefficient model expansion for $\log\{\sigma_i(\boldsymbol{s})\}$ with parameter vector $\boldsymbol{\eta}_{2k} = \{\eta_{2jtk}\}_{j,t=1}^{J_{2tk},q_2} \in \mathbb{R}^{p_{\boldsymbol{\eta}_{2k}}}$, where J_{2tk} is the number of basis functions for the tth covariate function in the finite linear approximation of $\log\{\sigma_i(\cdot)\}$ in $\mathcal{D}_k$ and $p_{\boldsymbol{\eta}_{2k}} = \sum_{t=1}^{q_2} J_{2tk}$ is fixed, $k = 1,\ldots,K$. Let $\boldsymbol{\eta}_k = (\boldsymbol{\eta}_{1k}^\top, \boldsymbol{\eta}_{2k}^\top)^\top \in \mathbb{R}^{p_k}$, with $p_k = p_{\boldsymbol{\eta}_{1k}} + p_{\boldsymbol{\eta}_{2k}}$ fixed. Due to the difficulty in estimating $\xi(\boldsymbol{s})$, we maintain the model $\xi_i(\boldsymbol{s};\boldsymbol{\beta}_3) = \boldsymbol{z}_{i3}(\boldsymbol{s})^\top\boldsymbol{\beta}_3$ proposed in Section 2.2. The MCCLE of $\boldsymbol{\theta}_k = (\omega, \zeta, \boldsymbol{\eta}_k^\top, \boldsymbol{\beta}_3^\top)^\top$ in subset $\mathcal{D}_k$ can be computed as in Section 3.2 and is denoted by $\widehat{\boldsymbol{\theta}}_k = (\widehat{\omega}_k, \widehat{\zeta}_k, \widehat{\boldsymbol{\eta}}_k^\top, \widehat{\boldsymbol{\beta}}_{3k}^\top)^\top \in \mathbb{R}^{2+p_k+q_3}$, with $\widehat{\boldsymbol{\eta}}_k^\top = (\widehat{\boldsymbol{\eta}}_{1k}^\top, \widehat{\boldsymbol{\eta}}_{2k}^\top)$. Note that we are not assuming that $\boldsymbol{\eta}_{1k} \equiv \boldsymbol{\eta}_1$, $\boldsymbol{\eta}_{2k} \equiv \boldsymbol{\eta}_2$: we model these marginal parameters separately for each subset to retain the spatial variation of the relationship between $\mu_i(\boldsymbol{s})$, $\sigma_i(\boldsymbol{s})$ and $\boldsymbol{z}_{i1}(\boldsymbol{s})$, $\boldsymbol{z}_{i2}(\boldsymbol{s})$. This results in heterogeneous marginal parameters. On the other hand, ω, ζ and $\boldsymbol{\beta}_3$ are assumed homogeneous across all subsets.

Let $\widehat{\omega}_c = (1/K)\sum_{k=1}^K \widehat{\omega}_k$, $\widehat{\zeta}_c = (1/K)\sum_{k=1}^K \widehat{\zeta}_k$, $\widehat{\boldsymbol{\beta}}_{3c} = (1/K)\sum_{k=1}^K \widehat{\boldsymbol{\beta}}_{3k}$, $\widehat{\boldsymbol{\eta}}_c = (\widehat{\boldsymbol{\eta}}_1^\top, \ldots, \widehat{\boldsymbol{\eta}}_K^\top)^\top \in \mathbb{R}^p$ with $p = \sum_{k=1}^K p_k$, and $\boldsymbol{\eta} = (\boldsymbol{\eta}_1^\top, \ldots, \boldsymbol{\eta}_K^\top)^\top$. Under conditions (C1), $\widehat{\boldsymbol{\theta}}_c = (\widehat{\omega}_c, \widehat{\zeta}_c, \widehat{\boldsymbol{\eta}}_c^\top, \widehat{\boldsymbol{\beta}}_{3c}^\top) \in \mathbb{R}^{2+p+q_3}$ is a consistent estimator of $\boldsymbol{\theta} = (\omega, \zeta, \boldsymbol{\eta}^\top, \boldsymbol{\beta}_3^\top)^\top$.

4.2. Integration Procedure

The goal of the integration procedure is to update the K estimators $\widehat{\boldsymbol{\eta}}_k$, $k = 1,\ldots,K$, for the heterogeneous parameters and to combine the K estimators $(\widehat{\omega}_k, \widehat{\zeta}_k, \widehat{\boldsymbol{\beta}}_{3k}^\top)^\top$, $k = 1,\ldots,K$, for the homogeneous parameters while leveraging dependence between subsets as in Sections 3.3 and 3.4. This will yield an integrated estimator of $\boldsymbol{\theta}$ that retains the spatial variation of the relationship between $\mu_i(\boldsymbol{s})$, $\sigma_i(\boldsymbol{s})$ and $\boldsymbol{z}_{i1}(\boldsymbol{s})$, $\boldsymbol{z}_{i2}(\boldsymbol{s})$. This integration can be derived with modification of the framework developed in Section 3. Denote

$$\widetilde{\Psi}_{1k}(\boldsymbol{\theta}_k) = \frac{1}{n}\sum_{i=1}^n \widetilde{\boldsymbol{\psi}}_{i1k}(\boldsymbol{\theta}_k) = \frac{1}{n}\left(\frac{\partial}{\partial\omega}, \frac{\partial}{\partial\zeta}\right)^\top \mathcal{CCL}_k(\boldsymbol{\theta}) \in \mathbb{R}^2,$$

$$\widetilde{\Psi}_{2k}(\boldsymbol{\theta}_k) = \frac{1}{n}\sum_{i=1}^n \widetilde{\boldsymbol{\psi}}_{i2k}(\boldsymbol{\theta}_k) = \frac{1}{n}\left(\frac{\partial}{\partial\boldsymbol{\eta}_{1k}}, \frac{\partial}{\partial\boldsymbol{\eta}_{2k}}\right)^\top \mathcal{CCL}_k(\boldsymbol{\theta}) \in \mathbb{R}^{p_k},$$

$$\widetilde{\Psi}_{3k}(\boldsymbol{\theta}_k) = \frac{1}{n}\sum_{i=1}^n \widetilde{\boldsymbol{\psi}}_{i3k}(\boldsymbol{\theta}_k) = \frac{1}{n}\frac{\partial}{\partial\boldsymbol{\beta}_3}\mathcal{CCL}_k(\boldsymbol{\theta}) \in \mathbb{R}^{q_3}.$$

Let $\widetilde{\boldsymbol{\psi}}_{ik}(\boldsymbol{\theta}_k) = \{\widetilde{\boldsymbol{\psi}}_{ijk}(\boldsymbol{\theta}_k)\}_{j=1}^3$, $\widetilde{\boldsymbol{\psi}}_{i,\text{all}}(\boldsymbol{\theta}) = \{\widetilde{\boldsymbol{\psi}}_{ik}(\boldsymbol{\theta}_k)\}_{k=1}^K$, $i = 1,\ldots,n$ and $\widetilde{\Psi}_k(\boldsymbol{\theta}_k) = \{\widetilde{\Psi}_{jk}(\boldsymbol{\theta}_k)\}_{j=1}^3$, $k = 1,\ldots,K$. Let $\widetilde{\Psi}_{\text{all}}(\boldsymbol{\theta}) = \{\widetilde{\Psi}_k(\boldsymbol{\theta}_k)\}_{k=1}^K$. Define the estimated sample covariance matrix of $\widetilde{\boldsymbol{\psi}}_{i,\text{all}}(\boldsymbol{\theta})$ as $\widetilde{\boldsymbol{C}}(\boldsymbol{\theta}) = (1/n)\sum_{i=1}^n \widetilde{\boldsymbol{\psi}}_{i,\text{all}}(\boldsymbol{\theta})\widetilde{\boldsymbol{\psi}}_{i,\text{all}}^\top(\boldsymbol{\theta}) \in \mathbb{R}^{(2K+p+q_3K)\times(2K+p+q_3K)}$.

Let $\lambda_1, \lambda_2 \geq 0$ denote two tuning parameters, $\boldsymbol{\lambda}_{\eta_j} \in \mathbb{R}^{(2+p+q_3)\times(2+p+q_3)}$ denote a diagonal matrix with λ_j's in the positions for $\boldsymbol{\eta}_j$ and 0 elsewhere, $j = 1, 2$. We define two sensitivity matrices: $\boldsymbol{I}_k(\boldsymbol{\theta}_k) = \nabla_{\boldsymbol{\theta}_k}\widetilde{\Psi}_k^\top(\boldsymbol{\theta}_k) \in \mathbb{R}^{(2+p_k+q_3)\times(2+p_k+q_3)}$ and $\widetilde{\boldsymbol{I}}_k(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}}\widetilde{\Psi}_k^\top(\boldsymbol{\theta}_k) \in \mathbb{R}^{(2+p_k+q_3)\times(2+p+q_3)}$, $k = 1,\ldots,K$. By construction of $\boldsymbol{\theta}$, $\widetilde{\boldsymbol{I}}_k(\boldsymbol{\theta}_k)$ is obtained from $\boldsymbol{I}_k(\boldsymbol{\theta})$ by adding rows of 0's for parameters in $\boldsymbol{\theta}$ that are not in $\boldsymbol{\theta}_k$. Define $\widetilde{\boldsymbol{W}}_k(\boldsymbol{\theta}) = \{\widetilde{\boldsymbol{C}}^{-1}(\boldsymbol{\theta})\}_{k,k}$. The varying-coefficient model meta-estimator is given by

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_{Vm} &= (\widehat{\omega}_{Vm}, \widehat{\zeta}_{Vm}, \widehat{\boldsymbol{\eta}}_{Vm}^\top, \widehat{\boldsymbol{\beta}}_{3Vm}^\top)^\top \\ &= \left\{\sum_{k=1}^K \widetilde{\Pi}_k(\widehat{\boldsymbol{\theta}}_c) + \boldsymbol{\lambda}_{\eta_1} + \boldsymbol{\lambda}_{\eta_2}\right\}^{-1} \sum_{k=1}^K \Pi_k(\widehat{\boldsymbol{\theta}}_c)\widehat{\boldsymbol{\theta}}_k, \qquad (8)\end{aligned}$$

with $\widetilde{\Pi}_k(\boldsymbol{\theta}) = \widetilde{\boldsymbol{I}}_k^\top(\boldsymbol{\theta})\widetilde{\boldsymbol{W}}_k(\boldsymbol{\theta})\widetilde{\boldsymbol{I}}_k(\boldsymbol{\theta})$ and $\Pi_k(\boldsymbol{\theta}) = \widetilde{\boldsymbol{I}}_k^\top(\boldsymbol{\theta})\widetilde{\boldsymbol{W}}_k(\boldsymbol{\theta})\boldsymbol{I}_k(\boldsymbol{\theta}_k)$.

Let $\widetilde{\boldsymbol{i}}_k = E_{\boldsymbol{\theta}}[\nabla_{\boldsymbol{\theta}}\{\widetilde{\Psi}_k(\boldsymbol{\theta}_k)\}]$ and $\widetilde{\boldsymbol{c}}(\boldsymbol{\theta}) = \text{var}_{\boldsymbol{\theta}}\{\sqrt{n}\widetilde{\Psi}_{\text{all}}(\boldsymbol{\theta})\}$. Denote by $\boldsymbol{\theta}_0$ the unique $\mathbf{0}$ of $E_{\boldsymbol{\theta}}\{\widetilde{\Psi}_{\text{all}}(\boldsymbol{\theta})\}$, the true value of $\boldsymbol{\theta}$.

Theorem 3. Suppose conditions (C1) and (C2) with $\Psi_k(\boldsymbol{\theta})$, $\Psi_{\text{all}}(\boldsymbol{\theta})$, $\boldsymbol{i}_k(\boldsymbol{\theta})$, $\boldsymbol{c}(\boldsymbol{\theta})$ replaced with $\widetilde{\Psi}_k(\boldsymbol{\theta}_k)$, $\widetilde{\Psi}_{\text{all}}(\boldsymbol{\theta})$, $\widetilde{\boldsymbol{i}}_k(\boldsymbol{\theta})$ and $\widetilde{\boldsymbol{c}}(\boldsymbol{\theta})$ are satisfied. When $\lambda_1, \lambda_2 \to 0$, $\widehat{\boldsymbol{\theta}}_{Vm}$ in (8) is a consistent estimator of $\boldsymbol{\theta}_0$. If λ_1, λ_2 are $o(n^{-1/2})$, then $\widehat{\boldsymbol{\theta}}_{Vm}$ is asymptotically

normally distributed, with asymptotic covariance consistently estimated by $\boldsymbol{J}^{-1}(\widehat{\boldsymbol{\theta}}_c) = n^{-1}\{\boldsymbol{H}(\widehat{\boldsymbol{\theta}}_c)\}^{-1}\boldsymbol{G}(\widehat{\boldsymbol{\theta}}_c)\{\boldsymbol{H}^\top(\widehat{\boldsymbol{\theta}}_c)\}^{-1}$,

$$\boldsymbol{H}(\boldsymbol{\theta}) = \sum_{k=1}^{K} \widetilde{\Pi}_k(\boldsymbol{\theta}) + \boldsymbol{\lambda}_{\boldsymbol{\eta}_1} + \boldsymbol{\lambda}_{\boldsymbol{\eta}_2},$$

$$\boldsymbol{G}(\boldsymbol{\theta}) = \sum_{k,k'=1}^{K} \widetilde{\boldsymbol{I}}_k^\top(\boldsymbol{\theta})\widetilde{\boldsymbol{W}}_k(\boldsymbol{\theta})\{\widetilde{\boldsymbol{C}}(\boldsymbol{\theta})\}_{k,k'}\widetilde{\boldsymbol{W}}_{k'}(\boldsymbol{\theta})\widetilde{\boldsymbol{I}}_{k'}(\boldsymbol{\theta}).$$

The asymptotic covariance of $(\widehat{\alpha}_{Vm}, \widehat{\phi}_{Vm}, \widehat{\boldsymbol{\eta}}_{Vm}^\top, \widehat{\boldsymbol{\beta}}_{3Vm}^\top)^\top$ can be recovered using the Delta method. In the supplementary materials, we recover estimates of $\mu_i(\boldsymbol{s})$ and $\sigma_i(\boldsymbol{s})$ and their standard errors and propose a generalized cross-validation statistic for selecting λ_1, λ_2.

Due to the heterogeneous nature of $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2$, the term $\boldsymbol{\lambda}_{\boldsymbol{\eta}_1} + \boldsymbol{\lambda}_{\boldsymbol{\eta}_2}$ induces smoothing of $\widehat{\mu}_i(\boldsymbol{s})$ and $\widehat{\sigma}_i(\boldsymbol{s})$ only within each subset $\mathcal{D}_k$. As a result, these may exhibit discontinuities at the boundaries of the subsets $\mathcal{D}_k$. Heaton et al. (2019) also observed this phenomenon for the local approximate Gaussian process and noted that these discontinuities are typically small enough so as to be undetectable in visual representations. If spatial smoothness of $\widehat{\mu}_i(\boldsymbol{s})$ and $\widehat{\sigma}_i(\boldsymbol{s})$ is critical, spatial interpolation may be used in post-processing. Alternatively, it is possible to use overlapping regions and to combine resulting estimates using a weighted average over the region overlaps to obtain smoother estimates at boundary points. This approach will retain some discontinuities but reduce their size. Recently, Manschot and Hector (2022) eliminated these discontinuities in the longitudinal outcome setting using a constrained generalized method of moments estimator, but the extension to the spatial domain is nontrivial due to the two-dimensional boundary between regions $\mathcal{D}_1, \ldots, \mathcal{D}_K$.

Table 1. Simulation metrics for Settings I and II in the first set of simulations.
(a) Setting I: u_j the 80% quantile, $(\alpha, \beta_3) = (0.8, 0.2)$.

Metric	K	α	ϕ	$\beta_{1,1}$	$\beta_{1,2}$	β_2	β_3
BIAS $\times 10^3$	16	0.41	94.19	−1.82	−2.97	4.76	0.26
	10	−0.36	110.48	−1.53	−3.27	4.76	0.09
	8	−0.88	120.11	−1.76	−3.56	5.65	−0.34
	1	−2.07	85.12	−1.62	−2.56	4.07	−0.86
ASE$\times 10^2$	16	0.93	107.59	1.63	1.63	4.65	2.31
	10	1.49	111.66	1.60	1.61	4.71	2.36
	8	1.43	110.91	1.55	1.57	4.69	2.36
	1	2.14	113.47	1.56	1.56	4.96	2.56
CP	16	0.95	0.96	0.95	0.95	0.94	0.93
	10	0.94	0.95	0.94	0.94	0.95	0.94
	8	0.93	0.95	0.95	0.94	0.95	0.94
	1	0.93	0.96	0.96	0.94	0.95	0.94

(b) Setting II: u_j the 90% quantile, $(\alpha, \beta_3) = (1, 0.2)$.

Metric	K	α	ϕ	$\beta_{1,1}$	$\beta_{1,2}$	β_2	β_3
BIAS $\times 10^3$	16	2.75	85.88	−5.41	−6.72	5.57	1.81
	10	2.18	103.13	−5.18	−7.87	7.20	1.03
	8	1.06	120.41	−5.83	−8.58	8.81	0.47
	1	−1.81	73.49	−5.56	−6.65	8.14	−1.42
ASE$\times 10^2$	16	1.40	117.93	3.36	3.36	6.60	3.06
	10	2.35	123.88	3.24	3.28	6.63	3.11
	8	2.23	122.57	3.10	3.13	6.57	3.11
	1	3.15	125.56	2.93	2.94	6.73	3.30
CP	16	0.93	0.94	0.95	0.95	0.93	0.93
	10	0.91	0.95	0.95	0.94	0.93	0.94
	8	0.92	0.95	0.95	0.94	0.93	0.93
	1	0.94	0.94	0.95	0.95	0.94	0.94

5. Simulations

We investigate the finite sample performance of the proposed meta-estimator $(\widehat{\alpha}_m, \widehat{\phi}_m, \widehat{\boldsymbol{\beta}}_m^\top)^\top$ derived from (6) and $(\widehat{\alpha}_{Vm}, \widehat{\phi}_{Vm}, \widehat{\boldsymbol{\eta}}_{Vm}^\top, \widehat{\boldsymbol{\beta}}_{3Vm}^\top)^\top$ derived from (8). Throughout, $\mathcal{D}$ consists of a square grid of evenly spaced locations. The Brown-Resnick processes $\{\mathcal{X}\}_{i=1}^n$ are independently simulated using the `SpatialExtremes` (Ribatet 2015) R package with unit Fréchet margins and values of α and ϕ specified below. Then $\{\mathcal{Y}_i\}_{i=1}^n$ are computed following the relationship in (1) with values of $\mu_i(\boldsymbol{s})$, $\sigma_i(\boldsymbol{s})$ and $\xi_i(\boldsymbol{s})$ specified below. All simulations are run on a standard Linux cluster with CCL analyses performed in parallel across K CPUs with 1GB of RAM. Standard errors and confidence intervals are calculated using the asymptotic normality results in Theorems 2 and 3. An additional simulation for the inverted MSP is given in the supplementary materials.

In the first set of simulations, we consider a $d = 400$-dimensional square spatial domain $\mathcal{D} = [1, 20]^2 = \{\boldsymbol{s}_j\}_{j=1}^{400}$, $\boldsymbol{s}_j \in \mathbb{R}^2$, with $n = 1000$ and consider a simple model from Section 2.2: $\mu_i(\boldsymbol{s}; \boldsymbol{\beta}_1) = \boldsymbol{s}^\top\boldsymbol{\beta}_1$, $\sigma_i(\boldsymbol{s}; \boldsymbol{\beta}_2) = \exp(\beta_2)$, $\xi_i(\boldsymbol{s}; \boldsymbol{\beta}_3) = \beta_3$, with $\phi = 10$, $\boldsymbol{\beta}_1 = (\beta_{11}, \beta_{12})^\top = (0.5, 0.5)^\top$, $\beta_2 = 1.5$. We evaluate the performance of $\widehat{\boldsymbol{\theta}}_m$ in (6) and its covariance with $\boldsymbol{J}^{-1}(\widehat{\boldsymbol{\theta}}_c)$, with $\mathcal{D}$ evenly partitioned based on nearest locations into $K = 16, 10, 8, 1$ square regions of size $d_k = 25, 40, 50, 400$, $k = 1, \ldots, K$, respectively. In Setting I, threshold u_j is the 80% quantile of $\{y_i(\boldsymbol{s}_j)\}_{i=1}^n$ and $(\alpha, \beta_3) = (0.8, 0.2)$; in Setting II, u_j is the 90% quantile and $(\alpha, \beta_3) = (1, 0.2)$. Table 1 reports the Asymptotic Standard Error (ASE), bias (BIAS) and 95% confidence interval coverage (CP) averaged across 500 simulations.

Since outcomes are simulated directly from the MSP, the model is appropriate for both thresholds. The effect of increasing the threshold is therefore to reduce the effective sample size. This is evident in the increased standard errors for larger threshold. The BIAS is well controlled for all parameter estimates. For example, in Setting I the BIAS (Monte Carlo standard error) for $\widehat{\alpha}_m$ and $\widehat{\phi}_m$ is 0.41×10^{-3} (0.95×10^{-2}) and 9.4×10^{-2} (1.1), respectively, for $K = 16$. Confidence interval coverage is appropriate for all settings. Across all settings, the ASE tends to increase as K decreases for all parameter estimates except $\widehat{\boldsymbol{\beta}}_1$. This reflects the fact that estimation of location parameters $\boldsymbol{\beta}_1$ is easiest. Mean elapsed times are reported in Table 2 and highlight the significant computational gain of our partitioning approach and its weak scalability. A third setting with u_j the 90% quantile and $(\alpha, \beta_3) = (1, -0.2)$ in the supplementary materials agrees with the findings from Settings I and II. We investigate the role of the threshold in balancing bias and variance when too small a threshold is used with an additional simulation in the supplementary materials.

To emphasize the importance of accounting for spatial dependence using the weight matrix $\boldsymbol{W}_k$, we compare our

Table 2. Mean elapsed time in minutes for the first set of simulations.

	$K = 16$	$K = 10$	$K = 8$	$K = 1$
Setting I	11.2	16.1	22.4	873.6
Setting II	13.7	18.6	23.6	845.7

approach in Setting I to the estimator $\widehat{\boldsymbol{\theta}}_c$ that averages the block MCCLEs. BIAS, averaged across 500 simulations and reported in the supplementary material, is larger for $\widehat{\boldsymbol{\theta}}_c$ than for our distributed estimator, suggesting our estimator has better point estimation properties. In the supplementary material, we also compare graphically the empirical standard errors, that is, an estimate of the true standard deviation of the estimators, and show that our estimator is generally more efficient: averaging across parameters for $K = 16$, the empirical standard error of the averaged MCCLE is 12% larger than that of our distributed estimator. The primary challenge in using $\widehat{\boldsymbol{\theta}}_c$ for inference, however, lies in the estimation of its variability. We consider two approaches for estimating the variance of $\widehat{\boldsymbol{\theta}}_c$: (i) as the average of the MCCLE variances divided by K and (ii) via the bootstrap. We report ASE and CP, averaged across 500 simulations, of $\widehat{\boldsymbol{\theta}}_c$ for approaches (i) and (ii) in the supplementary material. Using approach (i) vastly underestimates the ASE resulting in CP below the nominal 0.95 level. Thus, accounting for spatial dependence between blocks using the weight matrix $\boldsymbol{W}_k$ is not only important for improved statistical efficiency but crucial for correct inference. Moreover, mean elapsed times for approach (i) are 10.7, 15.6, and 21.8 min for $K = 16, 10, 8$, respectively. This approach provides very little computation gain over our approach, since our meta-estimator only incurs an additional inversion at a computational cost of $O(K^3)$. Using approach (ii), we estimate the variance of $\widehat{\boldsymbol{\theta}}_c$ using the sample variance of 100 nonparametric bootstrap samples (of the n independent replicates, with replacement) of the averaged block MCCLE. Due to the computational burden, we only consider $K = 16$. Parallelizing across K CPUs, mean elapsed time is 184 min, or 16 times slower than our distributed approach. CP of $\widehat{\boldsymbol{\theta}}_c$ reaches nominal levels for all parameters except the shape parameter β_3 (CP of 0.88). It is possible that resampling is unstable in the tails, leading to this undercoverage. We thus deem our distributed approach far superior to simply averaging the MCCLEs: it achieves correct statistical inference with very little additional computational cost over the calculation of the block MCCLE's, and a substantial computational advantage over the bootstrap.

We also compare our approach in Setting I to the tapered pairwise CL estimator that uses 5%, 10%, 20%, and 30% of nearest locations. We report BIAS, ASE and CP of the tapered CL estimator averaged across 500 simulations in the supplementary materials. BIAS, ASE, and CP between our distributed approach and the tapered pairwise CL are mostly similar with minor differences attributable to Monte Carlo error, with two exceptions: BIAS of the shape parameter β_3 and ASE of the smoothness parameter α are 7 and 1.7 times larger, respectively, for the tapered CL estimator that uses the nearest 20 locations than for our distributed approach that uses $d_k = 25$. Mean elapsed times of the tapered CL estimator are 85.7, 153.9, 377.4, and 462 min using 5%, 10%, 20%, and 30% of nearest locations, respectively. Notably, the tapered CL estimator that uses the nearest 20 locations is approximately eight times slower than our distributed approach that uses $d_k = 25$. Thus, our distributed approach has a substantial advantage over the tapered pairwise CL estimator since it delivers equivalent or superior inference in substantially faster computing time.

In the supplementary materials, a second set of simulations with $d = 900$ investigates the strong scalability of our proposed distributed approach. Computation time of the first and second set of simulations is also displayed graphically in the supplementary materials with a comparison to the theoretical computing time.

In the third set of simulations, we consider a $d = 400$-dimensional square spatial domain $\mathcal{D} = [1, 20]^2 = \{\boldsymbol{s}_j\}_{j=1}^{400}$, $\boldsymbol{s}_j = (s_{1,j}, s_{2,j}) \in \mathbb{R}^2$, with $n = 2000$ and a spatially-varying coefficient model from Section 4.1 with $z_{i1}(\boldsymbol{s}) = z_{i2}(\boldsymbol{s}) = 1$ (i.e., $t = 1$ and subscript t is omitted): $\mu_i(\boldsymbol{s}) = b_1(\boldsymbol{s})$, $\sigma_i(\boldsymbol{s}) = \exp(b_2(\boldsymbol{s}))$, and $\xi_i(\boldsymbol{s}; \boldsymbol{\beta}_3) = \beta_3$. We consider two settings for $b_1(\boldsymbol{s})$ and $b_2(\boldsymbol{s})$, $\boldsymbol{s} = (s_1, s_2)$. In Setting I, $b_1(\boldsymbol{s}) = (s_1^4 + s_2^4 + s_1 s_2)/d^2$ and $b_2(\boldsymbol{s}) = (s_1^2 + s_2^2)^{1/2}/10$. In Setting II, we let $b_1(\boldsymbol{s})$ and $b_2(\boldsymbol{s})$ be random draws from a Gaussian random field with Matérn covariance structure: two observations $b_j(\boldsymbol{s}_1)$ and $b_j(\boldsymbol{s}_2)$, $j \in \{1, 2\}$, separated by a Euclidean distance of t have covariance $t^5 K_5(t)/\{2^4\Gamma(5)\}$, where Γ is the Gamma function and K_5 is the modified Bessel function of the second kind. In both settings, we partition $\mathcal{D}$ evenly based on nearest locations into $K = 16$ square regions of size $d_k = 25$, $k = 1, \ldots, K$ and approximate $b_1(\boldsymbol{s})$ and $b_2(\boldsymbol{s})$ using the same basis function expansion as follows. In each subset $k \in \{1, \ldots, K\}$, we specify knot locations $\{\boldsymbol{\kappa}_{jk}\}_{j=1}^{10}$ at 10 locations chosen by minimizing a geometric space-filling criterion (Royle and Nychka 1998); an illustration of the geometric space-filling criterion is provided in the supplementary materials. We approximate $b_1(\boldsymbol{s})$ and $b_2(\boldsymbol{s})$ by linear combinations of Gaussian radial spline basis functions $\{C(||\boldsymbol{s} - \boldsymbol{\kappa}_{jk}||)\}_{j=1}^{10}$, $\boldsymbol{s} \in \mathcal{D}_k$, where $C(0) = 1$, $C(b) = \exp(-0.05b^2)$ for $b > 0$. Formally, we let $\phi_{j,k}(\boldsymbol{s}) \equiv \phi_j(\boldsymbol{s})$ and for $\boldsymbol{s} = (s_1, s_2) \in \mathcal{D}_k$, we define $\phi_1(\boldsymbol{s}) = 1$, $\phi_2(\boldsymbol{s}) = s_1$, $\phi_3(\boldsymbol{s}) = s_2$, $\phi_j(\boldsymbol{s}) = C(||\boldsymbol{s} - \boldsymbol{\kappa}_{j-3k}||)$, $j = 4, \ldots, 13$, and $\widetilde{z}_{i1,j}(\boldsymbol{s}) = \widetilde{z}_{i2,j}(\boldsymbol{s}) = \phi_j(\boldsymbol{s})$. Finally, we approximate $b_1(\boldsymbol{s}) \approx \sum_{j=1}^{13} \widetilde{z}_{i1,j}(\boldsymbol{s})\eta_{1,j,k}$, $b_2(\boldsymbol{s}) \approx \sum_{j=1}^{13} \widetilde{z}_{i2,j}(\boldsymbol{s})\eta_{2,j,k}$, $\boldsymbol{s} \in \mathcal{D}_k$, for some unknown parameters $\{\eta_{1,j,k}, \eta_{2,j,k}\}_{j=1}^{13}$. We estimate $\boldsymbol{\theta} = (\omega, \zeta, \boldsymbol{\eta}^\top, \beta_3)^\top$ with $\widehat{\boldsymbol{\theta}}_{Vm}$ in (8) and its covariance with $\boldsymbol{J}^{-1}(\widehat{\boldsymbol{\theta}}_c)$ in Theorem 3, with $\lambda_1, \lambda_2 \in \{0, 0.05, 0.1\}$. The partition of $\mathcal{D}$ with $K = 16$ gives $\boldsymbol{\theta} \in \mathbb{R}^{419}$, where $419 = 2 + 13 \times K + q_3$. We set u_j to the 85% quantile of $\{y_i(\boldsymbol{s}_j)\}_{i=1}^n$. True parameter values of α, ϕ, β_3 are set to $1, 10, 0.2$, respectively. Define the absolute error deviation, its average and its maximum as

$$\text{AED}_j(\boldsymbol{s}) = \left|\widehat{b}_j(\boldsymbol{s}) - b_j(\boldsymbol{s})\right| / \left[\max\left\{b_j(\mathcal{D})\right\} - \min\left\{b_j(\mathcal{D})\right\}\right],$$

Table 3. Varying coefficient model simulation results for $n = 2000$, $d = 400$, $K = 16$.

(a) Setting I: $b_1(\boldsymbol{s}) = (s_1^4 + s_2^4 + s_1 s_2)/d^2, b_2(\boldsymbol{s}) = (s_1^2 + s_2^2)^{1/2}/10$.

Parameter	BIAS × 10^2	ASE×10	aAED	mAED	CP
α	−0.40	0.08	–	–	0.93
ϕ	2.40	7.10	–	–	0.96
$b_1(\boldsymbol{s})$	–	–	0.23	1.2	0.94
$b_2(\boldsymbol{s})$	–	–	0.016	0.044	0.92
β_3	−0.86	0.18	–	–	0.86

(b) Setting II: $b_1(\boldsymbol{s}), b_2(\boldsymbol{s})$ drawn from Gaussian random fields.

Parameter	BIAS × 10^2	ASE×10	aAED	mAED	CP
α	−0.75	0.08	–	–	0.81
ϕ	2.30	7.10	–	–	0.96
$b_1(\boldsymbol{s})$	–	–	0.036	0.43	0.90
$b_2(\boldsymbol{s})$	–	–	0.011	0.038	0.91
β_3	−0.53	0.18	–	–	0.91

NOTE: Simulation metrics for $b_1(\boldsymbol{s})$ and $b_2(\boldsymbol{s})$ are averaged over observation locations.

$$\text{aAED}_j = \sum_{s\in\mathcal{D}} \text{AED}_j(s)/d, \quad \text{mAED}_j = \max_{s\in\mathcal{D}} \text{AED}_j(s), \quad j = 1, 2,$$

respectively. We report the BIAS, ASE, and CP for estimates of α, ϕ, β_3 and the aAED_j, mAED_j and CP_j of $b_j(s)$, $j = 1, 2$, with optimal λ_1, λ_2 selected using the generalized cross-validation statistic, averaged across 100 simulations for Settings I and II in Table 3. Selected values of λ_1, λ_2 are reported in the supplementary material.

In Setting I, BIAS, aAED, and mAED are appropriately small to suggest good point estimation, and CP of estimates of $\alpha, \phi, b_1(s)$ and $b_2(s)$ is appropriate. In Setting I, β_3 is disappointingly undercovered (CP of 86%). In Setting II, BIAS, aAED and mAED again suggest good point estimation, with slight undercoverage of parameters. The undercoverage in Settings I and II is potentially due to several factors. The performance of the GMM is known to deteriorate as the dimension of $\widetilde{\Psi}_{\text{all}}(\boldsymbol{\theta})$ increases relative to n. In both settings, $\widetilde{\Psi}_{\text{all}}(\boldsymbol{\theta}) \in \mathbb{R}^{464}$, where $464 = \sum_{k=1}^{K}(2 + p_k + q_3) \approx n/4$. Thus, $\widetilde{\boldsymbol{C}}(\widehat{\boldsymbol{\theta}}_c)$ may yield a poor estimate of the covariance of $\widetilde{\Psi}_{\text{all}}(\boldsymbol{\theta})$, affecting the estimation of the covariance of $\widehat{\boldsymbol{\theta}}_{Vm}$. Estimation of the shape parameter can be difficult because the bounds of the parameter space depend on observed values of $\mathcal{Y}$, potentially explaining the undercoverage of β_3. Undercoverage of α in Setting II may be due to the roughness of the location and scale parameters when simulated from the Gaussian process, which may confound the smoothness of the spatial dependence. Given the difficulty of Setting II, the CP for $b_1(s)$ and $b_2(s)$ is surprisingly good. Mean elapsed time, including cross-validation over the grid of (λ_1, λ_2) values, is 6 and 7.1 hr in Settings I and II, respectively. In the supplementary materials, plots of the estimated $\mu(s), \sigma(s)$ show a slight discontinuity between blocks, as discussed in Section 4.2.

6. Analysis of Extreme Streamflow Across the United States

To illustrate the proposed method, we analyze monthly measurements of streamflow from 1950 to 2020 at 702 locations across the United States as shown in Figure 1; locations are indexed by longitude and latitude, with longitude range −124.39 to −67.94 and latitude range 27.05 to 48.82. These locations are part of the USGS Hydro-Climatic Data Network 2009 (Lins 2012) and are selected because of their long record and because they are relatively unaffected by human activities. The locations are partitioned into $K = 12$ blocks based on the USGS watershed boundary regions (Figure 1). The response for month i at location s, $y_i(s)$, is the monthly maximum of the daily streamflow measurements. Streamflow has strong seasonality (see supplementary materials), so covariates for the GEV location ($z_{i1}(s)$) and log scale ($z_{i2}(s)$) include an intercept and four Fourier basis functions (two sine and two cosine) of the observation month to capture seasonality. The effects of these covariates are allowed to vary spatially following Section 4. The shape parameter ξ is assumed constant across space and time.

While the data are block maxima, the block size of a month may be insufficient to assume the data follow a max-stable process. Therefore, we analyze threshold exceedances using the censored MSP likelihood and compare this with a censored inverted MSP fit at the end of this section. To account for local heterogeneity we standardize the data at each site by subtracting the site's sample median and dividing by the difference of the site's 95% and 5% quantiles; all plots are made on the original data scale. We use roughly one spatial basis function per 20 locations in each block, giving between 2 and 5 basis functions per block. The basis functions are the same Gaussian kernel functions as in Section 5. We take the threshold $u(s)$ at a location s to be the level q sample quantile of the observations at the location. We fit the spatial model for several q and compare the results. Because many sites have a large number of zeros, we consider only $q \in \{85\%, 90\%, 95\%\}$. As shown in the supplementary materials, the fitted values and goodness of fit diagnostics are similar for all three thresholds so here we present results for $q = 85\%$. For $q = 85\%$, tuning parameters $\lambda_1 = 0.001, \lambda_2 = 0.090$ are selected from the ranges $\lambda_1 \in [0, 0.003], \lambda_2 \in [0, 0.15]$ via the generalized cross-validation statistic. The range of values for λ_1 and λ_2 was selected by first choosing a coarse grid, and then refining the grid to observe an initial decrease followed by an increase in the generalized cross-validation statistic. In general, practitioners should use a fine enough grid to find a small value of the generalized cross-validation statistic. In practice, the number of values over which to tune is constrained by computing time.

We use a probability integral transform plot to evaluate the fit of the model with $q = 85\%$. For each observation we compute $U_i(s) = \hat{F}_i\{y_i(s); s\}$, where $\hat{F}_i(y; s)$ is the fitted marginal GEV distribution function at site s and time i. Assuming the model fits well, the distribution of the $U_i(s)$ should be approximately Uniform(0,1). Figure 2 shows that this is the case for the fitted

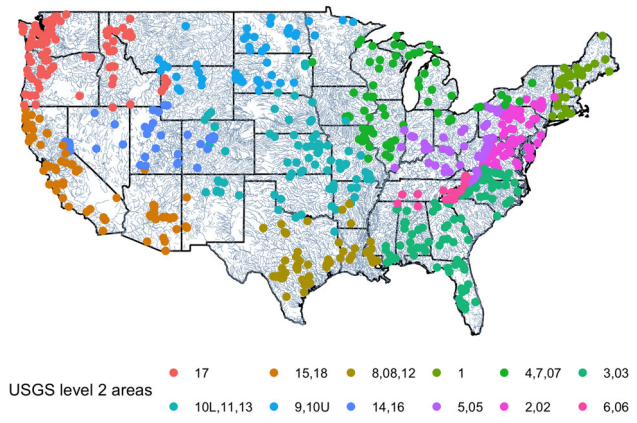

Figure 1. Partitioning of 702 spatial locations into $K = 12$ blocks based on USGS watershed boundary regions level 2 hydrologic unit codes (HUC02).

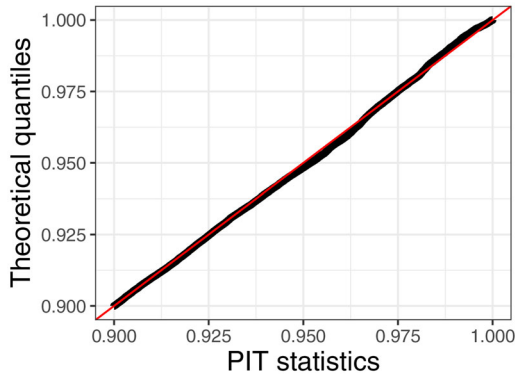

Figure 2. Probability integral transform plot with threshold set to the $q = 85\%$ quantile.

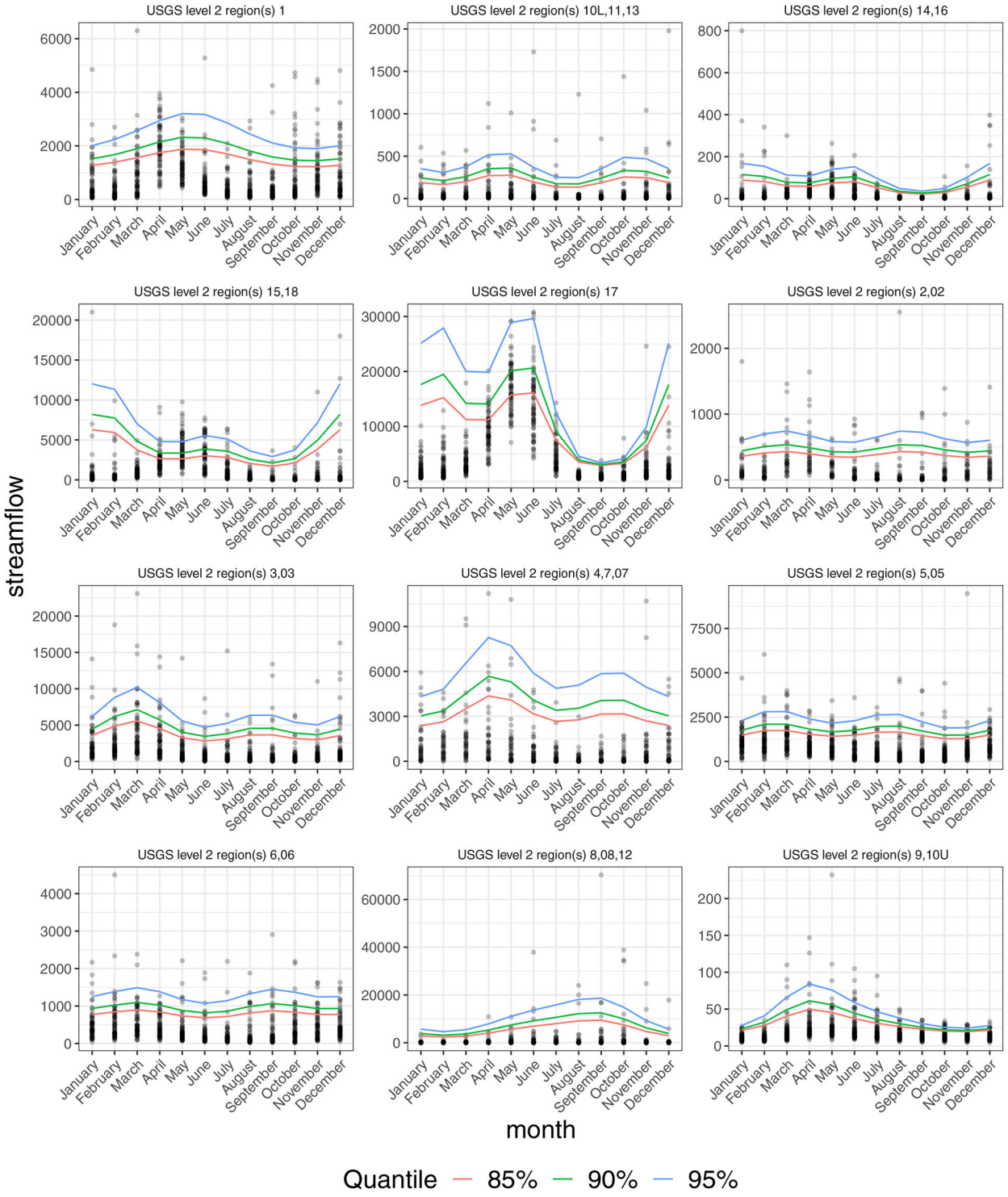

Figure 3. Streamflow (cfs) observations (dots) versus fitted quantiles (lines) for one randomly selected station in each of the $K = 12$ blocks with threshold set to the $q = 85\%$ quantile.

model. Probability integral transform plots in the supplementary materials for one randomly selected station in each of the $K = 12$ blocks with threshold set to the $q = 85\%$ quantile show that the model fits well at individual sites as well.

Figures in the supplementary materials map the estimated values of $\mu_i(\boldsymbol{s})$ and $\sigma_i(\boldsymbol{s})$. The location parameter varies considerably over space, with highest values in the Pacific Northwest and Missouri. The scale varies more by season, most notably in the Northern Plains. The GEV shape parameter (standard error) is estimated to be $\widehat{\xi}_{Vm} = 0.31$ (0.0049) giving a heavy right-tailed distribution. The estimated spatial dependence parameters are $\widehat{\alpha}_{Vm} = 0.73$ (0.0051) and $\widehat{\phi}_{Vm} = 56.3$ kilometers (1.98). Seasonal variation in these figures is largely obscured by spatial variation, so Figure 3 plots the data (pooled across years) versus fitted GEV quantiles for one randomly-selected station in each of the K blocks. The sites have prominent and varied seasonal patterns, illustrating the difficulty in modeling extremes over a large and heterogeneous region. For example, streamflow peaks

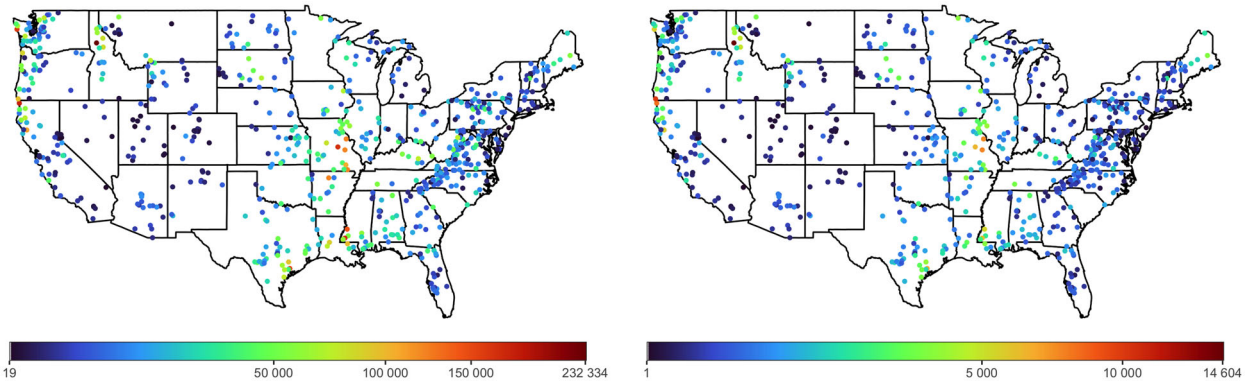

Figure 4. Estimated 50-year return levels (left) and standard errors (right) (cfs) with optimal smoothing parameters and threshold set to the $q = 85\%$ quantile.

in the spring for USGS level 2 regions 10L, 11, and 13, fall for USGS level 2 regions 14 and 16 and winter for USGS level 2 regions 2 and 02; the fitted model generally captures these disparate trends.

While the fitted model includes seasonality, the results can also be used to estimate the distribution and return level of the annual maximum. Let $\hat{F}_i(y;\boldsymbol{s})$ be the fitted GEV distribution function for month i at location $\boldsymbol{s}$, then the distribution function of the annual maximum is estimated as $\hat{F}(y;\boldsymbol{s}) = \prod_{i=1}^{12} \hat{F}_i(y;\boldsymbol{s})$ (noting that the fitted GEV distribution does not change by year). Inverting $\hat{F}(y;\boldsymbol{s})$ gives the estimated quantile function and thus the r-year return level, that is, the $1 - 1/r$ quantile of the annual maximum. Figure 4 plots the estimated 50-year return level and its standard error. The return level is maximized for stations at the mouth of the Mississippi River, in Southern Missouri and the Pacific Northwest. These stations also have high sample quantiles (see supplementary materials) but the fitted return levels are more stable and smooth across space.

The MSP model assumes nearby sites are asymptotically dependent, which may be questionable given the small block size. Therefore, we also fit the asymptotically-independent censored inverted MSP (iMSP) model described in the supplementary material. At quantile $q = 85\%$, tuning parameters $\lambda_1 = 0$, $\lambda_2 = 0.07$ are selected from the same ranges as the MSP model. The spatially varying coefficient marginal distribution model and spatial partitioning are the same as the MSP model. Following Hansen (1982) and Andrews (1999), the quadratic form $\sum_{k=1}^{K} \Psi_k^\top(\widehat{\boldsymbol{\theta}}_{Vm}) \boldsymbol{W}_k(\widehat{\boldsymbol{\theta}}_{Vm}) \Psi_k(\widehat{\boldsymbol{\theta}}_{Vm})$ can be used to assess model fit, similar to a log-likelihood. The quadratic form is 1006 units higher for the iMSP model (9008.254) than the MSP model (8002.017) over the entire spatial domain using $q = 85\%$, indicating the MSP model provides a better overall fit than the iMSP model; note that these models have the same number of parameters. The estimated iMSP spatial dependence parameters are $\widehat{\alpha}_{Vm} = 0.87$ (0.0066), $\widehat{\phi}_{Vm} = 217$ kilometers (5.39). Conditional exceedance probability plots in the supplementary materials give similar results for lower thresholds and, as expected, the MSP model shows stronger dependence for high thresholds. Despite these differences, the GEV parameter estimates are similar for the two models. The estimated GEV shape parameter is $\widehat{\xi}_{Vm} = 0.25$ (0.0043) for the iMSP fit, which is slighly lower than the estimate ($\widehat{\xi}_{Vm} = 0.31$) under the MSP model. Estimated values of $\mu_i(\boldsymbol{s})$ and $\sigma_i(\boldsymbol{s})$, estimated returns, streamflow versus fitted quantiles and probability integral transform plots for the iMSP model with $q = 85\%$, $q = 90\%$ and $= 95\%$ are very similar for the two models and are provided in the supplementary materials.

7. Discussion

The GMM suffers from well-known variance under-estimation when the sample size n is small; see Hansen, Heaton, and Yaron (1996) and others in the same issue. The difficulty primarily stems from evaluating $\boldsymbol{C}(\boldsymbol{\theta})$ at a consistent estimator whose variability is not accounted for in $\boldsymbol{J}(\boldsymbol{\theta})$. This issue is mitigated by the use of CCL, but the rarity of extreme events may prohibit the use of the asymptotic covariance formula in Theorems 2 and 3. As mentioned throughout this article, the choice K should be large enough that block MCCLEs are computationally fast to obtain and finite-sample bias is minimal, but small enough that a range of distances are available in each block for estimation of the dependence parameters, and in particular the range parameter. In practice, we have found that $d_k = 25$ locations per block performs well. See Section 3.1 for a discussion on the choice of K.

Supplementary Materials

Derivations, theorem conditions and proofs, extension to sub-asymptotic models including the inverted MSP, a computational complexity analysis, additional simulation and data analysis results and an R package are available in the supplementary material.

Acknowledgments

The authors thank Dr. Sankarasubramanian Arumugam of North Carolina State University for providing the streamflow data, and the reviewers and associate editor for their valuable feedback that led to a great improvement in the manuscript.

Disclosure Statement

The authors report there are no competing interests to declare.

Funding

This work was supported by grants from the National Science Foundation (DMS2152887, CBET2151651) and the National Institutes of Health (R01ES031651-01).

References

Andrews, D. W. (1999), "Consistent Moment Selection Procedures for Generalized Method of Moments Estimation," *Econometrica*, 67, 543–564. [11]

Bortot, P., Coles, S. G., and Tawn, J. A. (2000), "The Multivariate Gaussian Tail Model: An Application to Oceanographic Data," *The Annals of Applied Statistics*, 49, 31–49. [1]

Brown, B. M., and Resnick, S. I. (1977), "Extreme Values of Independent Stochastic Processes," *Journal of Applied Probability*, 14, 732–739. [3]

Buishand, T. A., de Haan, L., and Zhou, C. (2008), "On Spatial Extremes: With Application to a Rainfall Problem," *The Annals of Applied Statistics*, 2, 624–642. [1]

Castro-Camilo, D., and Huser, R. (2020), "Local Likelihood Estimation of Complex Tail Dependence Structures, Applied to u.s. Precipitation Extremes," *Journal of the American Statistical Association*, 115, 1037–1054. [4]

Castruccio, S., Huser, R., and Genton, M. G. (2016), "High-Order Composite Likelihood Inference for Max-Stable Distributions and Processes," *Journal of Computational and Graphical Statistics*, 25, 1212–1229. [2,3,5]

Coles, S. G. (2001), *An Introduction to Statistical Modeling of Extreme Values*, London: Springer. [1]

Davison, A. C., and Gholamrezaee, M. M. (2012), "Geostatistics of Extremes," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 468, 581–608. [2]

Davison, A. C., Huser, R., and Thibaud, E. (2019), *Spatial Extremes*, Boca Raton, FL: CRC Press. [2]

Davison, A. C., Padoan, S., and Ribatet, M. (2012), "Statistical Modeling of Spatial Extremes," (with Discussion), *Statistical Science*, 27, 161–186. [3]

Davison, A. C., and Ramesh, N. I. (2002). "Local Likelihood Smoothing of Sample Extremes," *Journal of the Royal Statistical Society*, Series B, 62, 191–208. [6]

de Haan, L. (1984), "A Spectral Representation for Max-Stable Processes," *The Annals of Probability*, 12, 1194–1204. [1]

Genton, M. G., Ma, Y., and Sang, H. (2011), "On the Likelihood Function of Gaussian Max-Stable Processes," *Biometrika*, 98, 481–488. [2]

Hansen, L. P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054. [2,4,5,11]

Hansen, L. P., Heaton, J., and Yaron, A. (1996), "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics*, 14, 262–280. [11]

Hastie, T., and Tibshirani, R. (1993), "Varying-Coefficient Models," *Journal of the Royal Statistical Society*, Series B, 55, 757–796. [6]

Heaton, M. J., Datta, A., Finley, A. O., Furrer, R., Guinness, J., Guhaniyogi, R., Gerber, F., Gramacy, R. B., Hammerling, D., Katzfuss, M., Lindgren, F., Nychka, D. W., Sun, F., and Zammit-Mangion, A. (2019), "A Case Study Competition Among Methods for Analyzing Large Spatial Data," *Journal of Agricultural, Biological and Environmental Statistics*, 24, 398–425. [4,7]

Hector, E. C., and Song, P. X.-K. (2021), "A Distributed and Integrated Method of Moments for High-Dimensional Correlated Data Analysis," *Journal of the American Statistical Association*, 116, 805–818. [4,5]

Huang, W. K., Stein, M. L., McInerney, D. J., Sun, S., and Moyer, E. J. (2016), "Estimating Changes in Temperature Extremes from Millennial-Scale Climate Simulations using Generalized Extreme Value (GEV) Distributions," *Advances in Statistical Climatology, Meteorology and Oceanography*, 2, 79–103. [1]

Huser, R., and Davison, A. C. (2013), "Composite Likelihood Estimation for the Brown-Resnick Process," *Biometrika*, 100, 511–518. [2,3,4]

——— (2014), "Space–Time Modelling of Extreme Events," *Journal of the Royal Statistical Society*, Series B, 76, 439–461. [1,4]

Huser, R., Davison, A. C., and Genton, M. G. (2016). Likelihood estimators for multivariate extremes. *Extremes*, 19(1):79–103. [2,3]

Huser, R., Dombry, C., Ribatet, M., and Genton, M. G. (2019), "Full Likelihood Inference for Max-Stable Data," *Stat*, 8, e218. [3]

Huser, R., and Genton, M. G. (2016), "Non-stationary Dependence Structures for Spatial Extremes," *Journal of Agricultural, Biological and Environmental Statistics*, 21, 470–491. [2]

Huser, R., Stein, M. L., and Zhong, P. (2022), "Vecchia Likelihood Approximation for Accurate and Fast Inference in Intractable Spatial Extremes Models," arXiv, arXiv:2203.05626. [6]

Huser, R., and Wadsworth, J. L. (2022), "Advaces in Statistical Modeling of Spatial Extremes," *WIREs Computational Statistics*, 14, e1537. [1]

Kabluchko, Z., Schlather, M., and de Haan, L. (2009), "Stationary Max-Stable Fields Associated to Negative Definite Functions," *The Annals of Probability*, 37, 2042–2065. [1,3]

Ledford, A. W., and Tawn, J. A. (1996), "Statistics for Near Independence in Multivariate Extreme Values," *Biometrika*, 83, 169–187. [1]

Lenzi, A., Bessac, J., Rudi, J., and Stein, M. L. (2021), "Neural Networks for Parameter Estimation in Intractable Models," arXiv, arXiv:2107.14346. [3]

Lindsay, B. G. (1988), "Composite Likelihood Methods," *Contemporary Mathematics*, 80, 221–239. [1,3,4]

Lins, H. F. (2012), "USGS Hydro-Climatic Data Network 2009 (HCDN-2009)," *US Geological Survey Fact Sheet*, 3047, 4. [9]

Manschot, C., and Hector, E. C. (2022), "Functional Regression with Intensively Measured Longitudinal Outcomes: A New Lens through Data Partitioning," arXiv, arXiv:2207.13014. [7]

Opitz, T. (2013), "Extremal t Processes: Elliptical Domain of Attraction and a Spectral Representation," *Journal of Multivariate Analysis*, 122, 409–413. [1]

Padoan, S., Ribatet, M., and Sisson, S. (2010), "Likelihood-based Inference for Max-Stable Processes," *Journal of the American Statistical Association*, 105, 263–277. [2,4]

Ribatet, M. (2015), *SpatialExtremes: Modelling Spatial Extremes*. R package version 2.0-2. [2,7]

——— (2017), *Nonlinear and Stochastic Climate Dynamics*, Cambridge: Cambridge University Press. [2,3]

Royle, J. A., and Nychka, D. (1998), "An Algorithm for the Construction of Spatial Coverage Designs with Implementation in SPLUS," *Computers & Geosciences*, 24, 479–488. [8]

Sang, H., and Genton, M. G. (2014), "Tapered Composite Likelihood for Spatial Max-Stable Models," *Spatial Statistics*, 8, 86–103. [2,5]

Sass, D., Li, B., and Reich, B. J. (2021), "Flexible and Fast Spatial Return Level Estimation via a Spatially Fused Penalty," *Journal of Computational and Graphical Statistics*, 30, 1124–1142. [2]

Schlather, M. (2002), "Models for Stationary Max-Stable Random Fields," *Extremes*, 5, 33–44. [1]

Smith, R. L. (1990), "Max-Stable Processes and Spatial Extremes," Unpublished manuscript. [1]

Smith, R. L., Tawn, J. A., and Coles, S. G. (1997), "Markov Chain Models for Threshold Exceedances," *Biometrika*, 84, 249–268. [1]

Stephenson, A. G., Shaby, B. A., Reich, B. J., and Sullivan, A. L. (2015), "Estimating Spatially Varying Severity Thresholds of a Forest Fire Danger Rating System Using Max-Stable Extreme-Event Modeling," *Journal of Applied Meteorology and Climatology*, 54, 395–407. [3]

Tawn, J. A. (1990), "Modelling Multivariate Extreme Value Distributions," *Biometrika*, 77, 245–253. [1]

Thibaud, E., Mutzner, R., and Davison, A. C. (2013), "Threshold Modeling of Extreme Spatial Rainfall," *Water Resources Research*, 49, 4633–4644. [1]

Thibaud, E., and Opitz, T. (2015), "Efficient Inference and Simulation for Elliptical Pareto Processes," *Biometrika*, 102, 855–870. [1]

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. [1]

Wang, Z., Jiang, Y., Wan, H., Yan, J., and Zhang, X. (2021), "Toward Optimal Fingerprinting in Detection and Attribution of Changes in Climate Extremes," *Journal of the American Statistical Association*, 116, 1–13. [3]