Received: 21 February 2020 | Revised: 9 August 2020 | Accepted: 13 August 2020

DOI: 10.1002/env.2656

RESEARCH ARTICLE

WILEY

Simultaneous autoregressive models for spatial extremes

Miranda J. Fix[1] | **Daniel S. Cooley**[2] | **Emeric Thibaud**[3]

[1]Department of Biostatistics, University of Washington, Seattle, Washington

[2]Department of Statistics, Colorado State University, Fort Collins, Colorado

[3]Institute of Mathematics, Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland

Correspondence
Miranda J. Fix, Department of Biostatistics, University of Washington, Seattle, WA.
Email: miranda.fix@gmail.com

Funding information
National Institute of Environmental Health Sciences, Grant/Award Number: T32ES015459; National Science Foundation, Grant/Award Numbers: DMS-1243102, DMS-1811657

Abstract

Motivated by the widespread use of large gridded data sets in the atmospheric sciences, we propose a new model for extremes of areal data that is inspired by the simultaneous autoregressive (SAR) model in classical spatial statistics. Our extreme SAR model extends recent work on transformed-linear operations applied to regularly varying random vectors, and is unique among extremes models in being directly analogous to a classical linear model. An additional appeal is its simplicity; given a proximity matrix W, spatial dependence is described by a single parameter ρ. We develop an estimation method that minimizes the discrepancy between the tail pairwise dependence matrix (TPDM) for the fitted model and the estimated TPDM. Applying this method to simulated data demonstrates that it is able to produce good estimates of extremal spatial dependence even in the case of model misspecification, and additionally produces reasonable estimates of uncertainty. We also apply the method to gridded precipitation observations for a study region over northeast Colorado, and find that a single-parameter extreme SAR model paired with a neighborhood structure which accounts for longer range dependence effectively models spatial dependence in these data.

KEYWORDS

areal data, regular variation, tail dependence, threshold exceedance

1 | INTRODUCTION

Natural hazards, such as floods and heatwaves, arise as extremes of physical processes that are inherently spatial. Figure 1 shows gridded daily total precipitation output on July 23, 1991 for a study region over northeast Colorado; we will apply our method to this data in Section 4. The figure clearly shows the spatial extent of this event: Several cells experience what are extreme precipitation amounts for this region. When estimating joint risk over a region due to extreme events, for example, for planning or insurance purposes, it is essential to account for spatial dependence.

In classical spatial statistics (e.g., Cressie, 1993), two common approaches are geostatistical models and areal data models. Geostatistical models are process models designed for point-referenced data, where we assume we have sampled d of an infinite number of potential spatial locations. Spatial dependence is modeled via a covariance function, and is a function of distance in the case of isotropy. In areal data models, the entire domain is partitioned into d regions and values are associated with these regions. Areal data models are not process models but rather finite-dimensional multivariate models incorporating spatial (neighborhood) structure. The simultaneous autoregressive (SAR; Whittle, 1954) and conditional autoregressive (CAR; Besag, 1974) are two well-known areal models, and modeling their simple spatial structure often requires estimation of only one parameter.

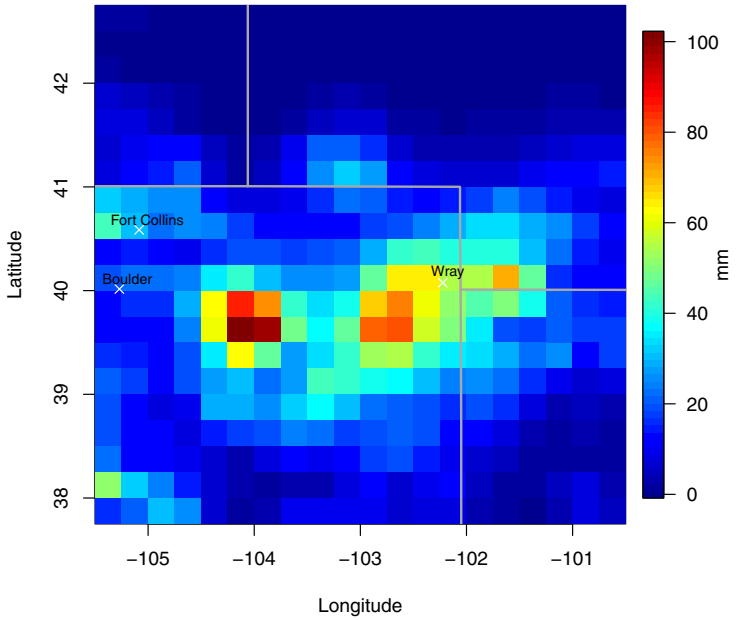

FIGURE 1 Gridded daily precipitation data from the Climate Prediction Center (CPC) unified gauge-based analysis on July 23, 1991 over a study region containing northeast Colorado. Gray lines indicate state borders, and white x's mark three cities in Colorado

Methods for extremes differ from standard statistical methods in that they employ models specifically suited for the distribution's tail and estimation utilizes only a small subset of extreme data. Classical spatial statistics methods are closely tied to mean-square prediction and Gaussian distributions and are best suited for describing central tendencies. Consequently, there has been much interest in developing spatial models appropriate for extremes (see, e.g., Davison, Huser, & Thibaud, 2019, for an overview). Existing models for spatial extremes, such as max-stable process models, tend to be geostatistical in nature. Full inference for max-stable and similar models has often been limited to a moderate number ($d \approx 30$) of spatial locations, and larger problems require advanced inference methods such as composite likelihood (Huser & Davison, 2013; Padoan, Ribatet, & Sisson, 2010) or semiparametric methods (Buhl, Davis, Klüppelberg, & Steinkohl, 2019) similar to what we propose in Section 4. A recent alternative to max-stable models are conditional extremes models (Shooter, Ross, Tawn, & Jonathan, 2019; Wadsworth & Tawn, 2019) which are flexible and have been applied at a larger number of locations, but whose complexity requires the estimation of two functional parameters of distance. In addition to work to develop spatial models whose dependence structure arises from extreme value theory, there is work to characterize how the nature of extreme phenomena vary over a spatial domain (e.g., Cooley, Nychka, & Naveau, 2007; Sharkey & Winter, 2019), work which aims to characterize patterns or trends in extreme behavior (e.g., Cabral, Ferreira, & Friederichs, 2020; Russell, Cooley, Porter, & Heald, 2016), and work which employs alternative spatial models to capture extreme behavior (e.g., Hazra, Reich, & Staicu, 2020; Sang & Gelfand, 2010).

Areal models have only recently become a topic of interest in the spatial extreme value analysis literature. Such models are needed, especially in atmospheric sciences, because much of the spatial data available for analysis are indexed by regular grids. Climate analyses often draw on gridded climate model output, or direct observations that are converted to gridded "data products" to, among other reasons, facilitate comparison to numerical climate model output. Recently, Reich and Shaby (2019) proposed a method for areal extremes that accounts for spatial dependence via latent clustering of neighboring regions, but is not directly analogous to classical areal models.

The aim of this work is to develop a simple spatial extremes model that is computationally feasible for areal data at a large number of locations. Specifically, we propose a new extreme model inspired by the classical SAR model. Our extremal SAR models employ the framework of multivariate regular variation (MVRV) for threshold exceedances.

2 | BACKGROUND

2.1 | Classical SAR model

First introduced by Whittle (1954), the simultaneous autoregressive (SAR) model is a simple model for areal data which can capture spatial dependence via a neighborhood structure. The classical SAR model can be specified by $\mathbf{Y} = S\mathbf{Y} + \boldsymbol{\epsilon}$,

where $\mathbf{Y}$ is a $d\times 1$ vector, S is a $d\times d$ matrix whose diagonal elements are zero, and $\boldsymbol{\epsilon}$ is a $d\times 1$ vector of independent zero-mean errors. Typically, it is further assumed $\boldsymbol{\epsilon} \sim MVN(\mathbf{0}, \sigma^2 I)$, where I is the $d\times d$ identity matrix. If $(I-S)$ is nonsingular, then

$$\mathbf{Y} = (I-S)^{-1}\boldsymbol{\epsilon}. \tag{1}$$

This rewritten form can be thought of as beginning with an independent vector $\boldsymbol{\epsilon}$ and inducing dependence in $\mathbf{Y}$, specifically $\mathbf{Y} \sim N[\mathbf{0},\ \sigma^2(I-S)^{-1}(I-S)^{-\top}]$.

To encode spatial information in S, the typical approach (e.g., Anselin, 1988; Cliff & Ord, 1973) is to let $S=\rho W$, where ρ is a single spatial dependence parameter and W is a $d\times d$ spatial proximity matrix that specifies the neighborhood structure. Many options for spatial proximity measures can be considered. In a simple case, W has entries that are 1 or 0 according to whether or not unit i and unit j are classified as neighbors (with $w_{ii}=0$). Two common neighborhood definitions are the rook's case, in which neighbors share a common edge, and the queen's case, in which neighbors share a common edge or a common vertex. More complex definitions of W which extend nonzero values further can induce longer range dependence, which we explore in Section 4.2.3. Also, it is possible to create higher order SAR models where $S=\rho_1 W_1+\rho_2 W_2+\ \dots\ +\rho_k W_k$ with $\{W_i\}$ specifying neighbors at different distances (Haining, 1990). However, increasing the number of spatial parameters makes estimation more complex, and most SAR applications have had a single spatial parameter ρ.

To ensure invertibility of $(I-\rho W)$, ρ is required to satisfy $\rho\lambda_i \neq 1$ for $i=1,\ \dots\ ,d$, where $\lambda_1<\lambda_2<\ \dots\ <\lambda_d$ are the ordered eigenvalues of W (Kelejian & Robinson, 1995). In practice, ρ is almost always restricted to the interval $(\lambda_1^{-1}\ ,\ \lambda_d^{-1})$, or $(0\ ,\ \lambda_d^{-1})$ when negative spatial autocorrelation is not of practical interest (Haining, 1990; Song & De Oliveira, 2012; Wall, 2004).

2.2 | Extremes and linear operations

2.2.1 | Multivariate regular variation on the nonnegative orthant

We will work within the multivariate regular variation (MVRV) framework for threshold exceedances. MVRV provides a probabilistic characterization of the joint (upper) tail of a random vector, and is defined entirely in terms of the joint tail. The MVRV framework assumes the heavy-tailed case, that is, it implies that the joint tail decays like a power function. We model in the nonnegative orthant to focus attention on the upper tail and in the direction of interest, for example, in our application, we aim to model extremely large precipitation.

There are several equivalent definitions for regular variation (Resnick, 2007, theorem 6.1); the most useful for our purposes is in terms of polar coordinates. Let $\mathbf{Y}$ be a random vector taking values in the nonnegative orthant $\mathbb{R}_+^d=[0,\infty)^d$, and let $M_+(\mathfrak{C})$ denote the space of nonnegative Radon measures on $\mathfrak{C}=[0,\infty]^d\setminus\{\mathbf{0}\}$. Given any norm $\|\cdot\|$, denote the nonnegative unit sphere $\mathbb{S}_+^{d-1}=\{\mathbf{y}\in\mathbb{R}_+^d\ :\ \|\mathbf{y}\|=1\}$. We define "radial" and "angular" components by $R=\|\mathbf{Y}\|$ and $\boldsymbol{\Theta}=\|\mathbf{Y}\|^{-1}\mathbf{Y}$, respectively. Then the random vector $\mathbf{Y}$ is *regularly varying* if there exists a sequence $c_n\to\infty$ and a finite measure H on $\mathbb{S}_+^{d-1}$ such that for any H-continuity Borel set $B\subset\mathbb{S}_+^{d-1}$, and for $r>0$,

$$n\mathbb{P}\left(\frac{R}{c_n}>r, \boldsymbol{\Theta}\in B\right)\xrightarrow{v} r^{-\alpha}H(B) \tag{2}$$

in $M_+(\mathfrak{C})$ as $n\to\infty$, where $\xrightarrow{v}$ denotes vague convergence of measures. The right-hand side of (2) is a product measure, indicating that the radial and angular components become independent in the limit. H is termed the *angular measure* and completely characterizes the limiting tail dependence structure of $\mathbf{Y}$. The tail index α determines the power-law behavior of the tail. We denote a d-dimensional regularly varying random vector $\mathbf{Y}$ with tail index α by $\mathbf{Y}\in RV_+^d(\alpha)$. We denote the limiting measure as ν, that is, for a general set $\mathcal{R}$ in $M_+(\mathfrak{C})$, $n\mathbb{P}\left(\frac{\mathbf{Y}}{c_n}\in\mathcal{R}\right)\xrightarrow{v}\nu(\mathcal{R})=\int_{(r,\boldsymbol{\theta})\in\mathcal{R}}\alpha r^{-\alpha-1}dH(\boldsymbol{\theta})$.

2.2.2 | Transformed-linear operations and MVRV random vectors

The classical SAR construction relies on linear operations. As we model in the nonnegative orthant to focus attention on upper tail, this raises the question of how to define a vector space on the nonnegative orthant. To this end, we employ

"transformed-linear" operations as defined by Cooley and Thibaud (2019). Let t be a bijection from $\mathbb{R}$ onto $(0,\infty)$, with t^{-1} its inverse, and for vectors, apply t and t^{-1} componentwise. Define vector addition in $(0,\infty)^d$ by $\mathbf{x}_1 \oplus \mathbf{x}_2 = t\{t^{-1}(\mathbf{x}_1) + t^{-1}(\mathbf{x}_2)\}$, and for any $a \in \mathbb{R}$, define scalar multiplication of a vector in $(0,\infty)^d$ by $a \circ \mathbf{x} = t\{at^{-1}(\mathbf{x})\}$. Define the additive identity in $(0,\infty)^d$ as $t(\mathbf{0})$, and the additive inverse of any $\mathbf{x} \in (0,\infty)^d$ as $-\mathbf{x} = t\{-t^{-1}(\mathbf{x})\}$. Then, Cooley and Thibaud (2019) show that $(0,\infty)^d$ is a vector space under these transformed-linear operations.

To apply these transformed-linear operations to regularly varying random vectors, we will use the specific transform $t : \mathbb{R} \to (0,\infty)$ defined by $t(v) = \log\{\exp(v)+1\}$, with inverse $t^{-1}(x) = \log\{\exp(x)-1\}$. Extending the definition of t such that $t(-\infty) = 0$, $t^{-1}(0) = -\infty$, and $t(\infty) = t^{-1}(\infty) = \infty$ yields a bijection $t : [-\infty,\infty]^d \to [0,\infty]^d$. This bijection has the important property that $\lim_{v\to\infty} \frac{t(v)}{v} = \lim_{x\to\infty} \frac{t^{-1}(x)}{x} = 1$, that is, the transform and its inverse have a negligible effect on large values. If $\mathbf{Y} \in RV_+^d(\alpha)$, Cooley and Thibaud (2019) show that MVRV is preserved by transformed-linear operations with transform t. Specifically, they show that if $\boldsymbol{Y}_1, \boldsymbol{Y}_2 \in RV_+^d(\alpha)$ and independent, each with normalizing sequence c_n and with respective limiting measures $\nu_{\mathbf{Y}_1}$ and $\nu_{\mathbf{Y}_2}$, then $\mathbf{Y}_1 \oplus \mathbf{Y}_2 \in RV_+^d(\alpha)$, and $n\mathbb{P}\{c_n^{-1}(\mathbf{Y}_1 \oplus \mathbf{Y}_2) \in \cdot\} \overset{v}{\to} \nu_{\mathbf{Y}_1}(\cdot) + \nu_{\mathbf{Y}_2}(\cdot)$. They also show that for $\boldsymbol{Y} \in RV_+^d(\alpha)$ with normalizing sequence c_n and limiting measure $\nu_{\mathbf{Y}}$ and nonnegative scalar a, $n\mathbb{P}\{c_n^{-1}(a \circ Y) \in \cdot\} \overset{v}{\to} a^\alpha \nu_{\mathbf{Y}}(\cdot)$. Cooley and Thibaud (2019) impose a lower tail condition on $\boldsymbol{Y}$ so that scalar multiplication can be extended to negative scalars a while still preserving regular variation; here we restrict our attention to nonnegative scalars.

Matrix multiplication can be defined using transformed-linear operations, and we consider constructing regularly varying random vectors by applying transformed-linear matrix multiplication to a vector of iid regularly varying random variables. Recall in the nonextreme setting, by letting $\mathbf{Y} = A\mathbf{Z}$ where $A \in \mathbb{R}^{d\times q}$ and $\mathbf{Z} \sim MVN(\mathbf{0}, I_{q\times q})$, one can construct a correlated Gaussian random vector with $Cov(\mathbf{Y}) = AA^T$. Let $A = (\mathbf{a}_1, \ldots, \mathbf{a}_q) \in \mathbb{R}^{d\times q}$ with $a_{ij} \geq 0$ and $\max_{i=1,\ldots,d} a_{ij} > 0$ for all $j = 1, \ldots, q$, and let $\mathbf{Z} = (Z_1, \ldots, Z_q)^\top$ be a vector of iid regularly varying random variables with tail index α such that $n\mathbb{P}(c_n^{-1}Z_j > z) \to z^{-\alpha}, j = 1, \ldots, q$. Then $A \circ \mathbf{Z} = \mathbf{a}_1 \circ Z_1 \oplus \ldots \oplus \mathbf{a}_q \circ Z_q = t\{At^{-1}(\mathbf{Z})\} \in RV_+^d(\alpha)$, and when renormalized by $\{c_n\}$ has limiting angular measure

$$H_{A\circ\mathbf{Z}}(\cdot) = \sum_{j=1}^{q} \|\mathbf{a}_j\|^\alpha \delta_{\mathbf{a}_j/\|\mathbf{a}_j\|}(\cdot), \tag{3}$$

where δ is the Dirac mass function. Thus, by applying the matrix A to a vector of independent regularly varying random variables, we obtain a regularly varying random vector with extremal dependence structure described by $H_{A\circ\mathbf{Z}}$.

The angular measure $H_{A\circ\mathbf{Z}}$ is discrete, consisting of point masses with locations corresponding to the normalized columns of A. In fact, it is the same limiting angular measure as for the max-linear construction $A\times_{\max}\mathbf{Z} = (\max_{j=1,\ldots,d} a_{1j}Z_j, \ldots, \max_{j=1,\ldots,d} a_{dj}Z_j)^\top$ (e.g., Fougères, Mercadier, & Nolan, 2013). Although the constructions $A \circ \mathbf{Z}$ and $A\times_{\max}\mathbf{Z}$ are similar, they differ in their realizations, as large realizations of the transformed-linear construction have angular components near but not exactly at locations corresponding to the normalized columns of A. More importantly, transformed linear matrix multiplication has a natural inverse: if A^{-1} denotes the standard matrix inverse, then $A^{-1} \circ (A \circ \boldsymbol{Z}) = A \circ (A^{-1} \circ \boldsymbol{Z}) = \boldsymbol{Z}$.

2.2.3 | Tail pairwise dependence matrix

In standard statistical practice, dependence is often analyzed via the covariance matrix, even when Gaussianity is not assumed. Because the covariance matrix is not designed to capture dependence in the tail, we construct a matrix that summarizes second-order properties of the angular measure. Suppose $\mathbf{Y} = (Y_1, \ldots, Y_d)^\top \in RV_+^d(2)$ such that $n\mathbb{P}(n^{-1/2}\mathbf{Y} \in \cdot) \overset{v}{\to} \nu_{\mathbf{Y}}(\cdot)$, where $\nu_{\mathbf{Y}}(dr \times d\boldsymbol{\theta}) = 2r^{-3}drdH_{\mathbf{Y}}(\boldsymbol{\theta})$ with $H_{\mathbf{Y}}$ a Radon measure on the nonnegative L_2 unit sphere $\mathbb{S}_{+(2)}^{d-1} = \{\boldsymbol{\theta} \in \mathbb{R}_+^d : \|\boldsymbol{\theta}\|_2 = 1\}$. Then Cooley and Thibaud (2019) define the $d \times d$ *tail pairwise dependence matrix* (TPDM) as $\Sigma_{\mathbf{Y}} = (\sigma_{\mathbf{Y}ik})_{i,k=1,\ldots,d}$, where

$$\sigma_{\mathbf{Y}ik} = \int_{\mathbb{S}_{+(2)}^{d-1}} \theta_i\theta_k dH_{\mathbf{Y}}(\boldsymbol{\theta}) \tag{4}$$

is the extremal dependence measure defined in the bivariate case by Larsson and Resnick (2012). The assumption that $\alpha = 2$ and using the L_2 norm result in the TPDM $\Sigma_{\mathbf{Y}}$ having many properties that are analogous to those of a covariance

matrix. $\Sigma_{\mathbf{Y}}$ is positive semidefinite. The diagonal entries of $\Sigma_{\mathbf{Y}}$ provide information about the scale of the components of $\mathbf{Y}$; specifically, $\sigma_{\mathbf{Y}ii}$ is equal to the square of the marginal scale parameter of Y_i. For $i \neq k$, $\sigma_{\mathbf{Y}ik} = 0$ corresponds to asymptotic independence of the components Y_i and Y_k (Cooley & Thibaud, 2019).

Regularly varying random vectors constructed via transformed-linear matrix multiplication have a special form of the TPDM. Let $A \in \mathbb{R}^{d\times q}$ with $a_{ij} \geq 0$ and $\max_{i=1,\dots,d} a_{ij} > 0$ for all $j = 1, \dots, q$, where $q \geq d$, and let $\mathbf{Z} = (Z_1, \dots, Z_q)^\top$ be a vector of iid regularly varying random variables with $\alpha = 2$ such that $n\mathbb{P}(c_n^{-1} Z_j > z) \to z^{-2}, j = 1, \dots, q$. Then $\Sigma_{A \circ \mathbf{Z}} = AA^\top$.

3 | EXTREME SAR MODELS

In this section we propose multivariate extreme analogues of the classical SAR model presented in Section 2.1.

3.1 | A SAR-inspired model for areal extremes

3.1.1 | Formulation

Let $\mathbf{Z} = (Z_1, \dots, Z_d)^\top$ be a vector of iid regularly varying random variables with tail index α such that $n\mathbb{P}(c_n^{-1} Z_j > z) \to z^{-\alpha}$, $j = 1, \dots, d$. Let S be a $d \times d$ matrix whose diagonal elements are zero, and satisfying $(I - S)$ nonsingular. Then, using transformed-linear operations, we can define the equivalent expressions (see Supplement Section S1 for the full expansions):

$$\begin{aligned} &\mathbf{Y} = S \circ \mathbf{Y} \oplus \mathbf{Z}, \\ \Leftrightarrow\ &(I - S) \circ \mathbf{Y} = \mathbf{Z}, \\ \Leftrightarrow\ &\mathbf{Y} = (I - S)^{-1} \circ \mathbf{Z}. \end{aligned} \tag{5}$$

Throughout, we follow the usual SAR single-parameter scheme, but note that the model could be extended to be higher order as with the classical SAR (see Section 2.1). We choose to let $S = \rho W$, where ρ is a single spatial dependence parameter and W is a nonnegative spatial proximity matrix. This results in the model formulation

$$\mathbf{Y} = (I - \rho W)^{-1} \circ \mathbf{Z}. \tag{6}$$

For simplicity of notation, we denote the matrix $(I - \rho W)^{-1}$ by $A = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_d)$. Thus, we can write $\mathbf{Y} = A \circ Z$, matching the construction described in Section 2.2.2.

3.1.2 | Condition on ρ and properties

Recall from Section 2.2.2 that we wish to restrict attention to A with nonnegative entries. We can show that in the case of $A = (I - \rho W)^{-1}$, if ρ takes on a similar restriction as for the classical SAR, then A is both well-defined and nonnegative. The proof (see Supplement Section S2) uses results from nonnegative matrix theory (e.g., Plemmons, 1977).

Proposition 1. *Let W be a $d \times d$ spatial proximity matrix whose diagonal elements are zero and off-diagonal entries are nonnegative. Let $\lambda_1 < \lambda_2 < \dots < \lambda_d$ denote the ordered eigenvalues of W. If*

$$\rho \in (0\ ,\ \min\{|\lambda_1|^{-1}, |\lambda_d|^{-1}\}), \tag{7}$$

then $A = (I - \rho W)^{-1}$ exists and is a nonnegative matrix.

Note that for square lattices with W a binary proximity matrix under the rook neighborhood specification, $\lambda_1 \downarrow -4$ and $\lambda_d \uparrow 4$ as $d \to \infty$ (Haining, 1990), so $\rho \in (0, 0.25)$ is a sufficient restriction to ensure A exists and is entrywise nonnegative.

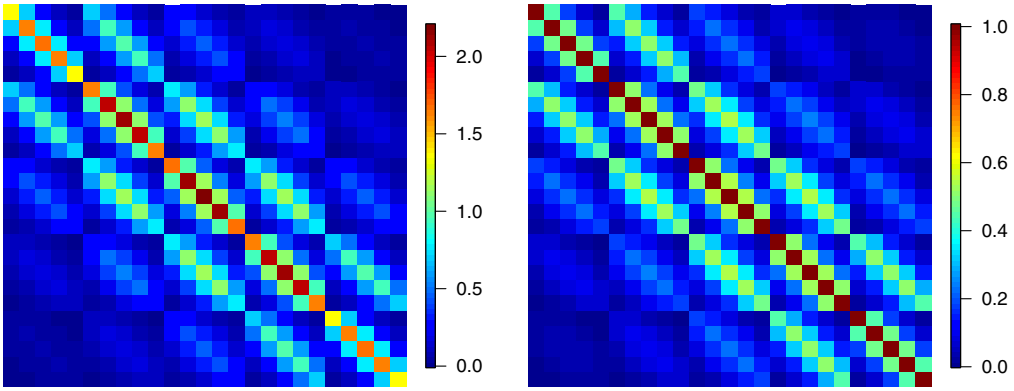

FIGURE 2 Tail pairwise dependence matrices for $\mathbf{Y} = A \circ \mathbf{Z}$ (left) and $\mathbf{Y} = \tilde{A} \circ \mathbf{Z}$ (right), with $d = 25$, $\rho = 0.2$, and W a binary proximity matrix following the rook neighborhood structure. The diagonal elements $\sigma_{\mathbf{Y}ii}$ are equal to the square of the scale of Y_i. The case of $\mathbf{Y} = \tilde{A} \circ \mathbf{Z}$ (right) has a common scale of 1

Henceforth, we assume that ρ meets condition (7). $\boldsymbol{Y}$ as defined in (6) has angular measure given by (3). In the case of $\alpha = 2$, using the L_2 norm and with $c_n = n^{1/2}$, the TPDM of $\boldsymbol{Y}$ simplifies to

$$\Sigma_{\mathbf{Y}}(\rho) = (I - \rho W)^{-1}(I - \rho W)^{-\top}, \tag{8}$$

which is reminiscent of the covariance matrix in the Gaussian SAR model.

3.2 | An extreme SAR model with common scale

There is an issue with the extreme SAR model as defined in Section 3.1 which could prove problematic for implementation. The diagonal elements of the TPDM given in (8) are not identical (see Figure 2, left), and thus the marginal scale parameters of the elements of $\mathbf{Y}$ differ. Similar behavior is found in the classical SAR, which is not marginally stationary. In most classical SAR work, this issue is not addressed, and marginal distributions at the locations are allowed to have different variances. However, like for many spatial extremes models, application of our model likely will involve a marginal transformation. The fact that the marginal scales depend on ρ poses a challenge for the often-used two-stage inference procedure for multivariate extremes, in which one first estimates the marginal effects and then estimates the tail dependence structure after marginal transformation. If ρ was known, it would be possible to transform the marginals accordingly, but here ρ is unknown and must be estimated. To address this issue, we propose an extension, which results in common (unit) marginal scale parameters and is thus more amenable to a two-stage inference procedure.

Let $\mathbf{Z} = (Z_1, \ldots, Z_d)^\top$ be a vector of iid regularly varying random variables with tail index $\alpha = 2$ such that $n\mathbb{P}(n^{-1/2}Z_j > z) \to z^{-2}$, $j = 1, \ldots, d$. Let $A = (I - \rho W)^{-1}$ as before, and let $D = \mathrm{diag}(\|A_{1\cdot}\|_2, \ldots, \|A_{d\cdot}\|_2)$ denote the diagonal matrix of the L_2 norms of the rows of A. Then the common-scale model is defined as follows:

$$\mathbf{Y} = \tilde{A} \circ \mathbf{Z} = D^{-1}A \circ \mathbf{Z}. \tag{9}$$

Note that D, like A, depends on the value of ρ.

Because we assume ρ meets condition (7), A is nonnegative and has positive row norms. Thus $\tilde{A}$ is also nonnegative, and the modified formulation (9) results in the TPDM

$$\Sigma_{\mathbf{Y}}(\rho) = \Sigma_{\tilde{A} \circ \mathbf{Z}} = \tilde{A}\tilde{A}^\top = D^{-1}AA^\top D^{-1}, \tag{10}$$

which has diagonal elements $\sigma_{\mathbf{Y}ii} = \|A_{i\cdot}\|_2^{-1}\|A_{i\cdot}\|_2^2\|A_{i\cdot}\|_2^{-1} = 1$ (Figure 2, right). This means each component of $\mathbf{Y}$ has a marginal scale parameter of 1, that is, $n\mathbb{P}(n^{-1/2}Y_i > y) \to y^{-2}$, $i = 1, \ldots, d$. The limiting angular measure for this common-scale model is

$$H_{\mathbf{Y}}(\cdot) = H_{\tilde{A} \circ \mathbf{Z}}(\cdot) = \sum_{j=1}^{d} \|\tilde{\mathbf{a}}_j\|_2^2 \ \delta_{\tilde{\mathbf{a}}_j/\|\tilde{\mathbf{a}}_j\|_2}(\cdot), \tag{11}$$

where $\tilde{\mathbf{a}}_j$ denotes the jth column of $\tilde{A}$, for $j = 1, \ldots, d$.

4 | ESTIMATION AND INFERENCE

The beauty of classical areal models such as the SAR and CAR is their simplicity: spatial dependence is induced by a single parameter ρ and predetermined spatial proximity matrix W. However, this overly simplistic representation of dependence can lead to difficulties for inference. For example, Besag and Kooperberg (1995) note that a common issue with the CAR is that appreciable correlations between neighboring sites require parameter values extremely close to the boundary of the parameter space.

The discrete nature of the limiting angular measure of the extreme SAR model poses a challenge to estimation. To our knowledge, likelihood-based inference has not been attempted for existing extremes models with discrete angular measures, such as max-linear models (Fougères et al., 2013; Wang & Stoev, 2011). Instead, several minimum distance methods have been proposed. Yuen and Stoev (2014) propose an M-estimator based on the continuous ranked probability score of multivariate cumulative distribution functions. Einmahl, Kiriliouk, Krajina, and Segers (2016) also propose an M-estimator, theirs based on the stable tail dependence function, which is related to the upper tail of the underlying cumulative distribution function. Einmahl, Kiriliouk, and Segers (2018) extend the previous approach with an adaptive weighted least-squares procedure. The max-linear examples examined in the aforementioned studies are of dimension at most $d = 10$, whereas in this section we consider an extreme SAR model of dimension $d = 400$.

Least squares fitting of nonextreme spatial models is common, often by minimizing the difference between empirical and model variograms (Cressie, 1993, section 2.6). Below, we propose a novel least squares method for estimation of ρ based on the TPDM. After providing details of our method, we apply it to three types of data sets: simulated data from the extreme SAR model, simulated data from a max-stable process, and gridded precipitation data.

4.1 | Model fitting via the TPDM

The TPDM $\Sigma_{\mathbf{Y}}(\rho)$ is a useful summary of the extremal dependence induced by the SAR model. The goal of this approach is to find the value of ρ whose corresponding theoretical TPDM most closely matches the observed tail pairwise dependence structure with respect to the Frobenius norm. Implementing this method involves several considerations. First, we must obtain an estimate of the TPDM, $\hat{\Sigma}_{\mathbf{Y}}$, from the data. We use a threshold-based estimator defined using a d-dimensional vector norm. It is well known that in cases of moderate to weak extremal dependence, threshold-based estimators tend to overestimate dependence (Huser, Davison, & Genton, 2016). To mitigate this bias, we propose a simple method to obtain a bias-corrected estimate of the TPDM, $\tilde{\Sigma}_{\mathbf{Y}}$. Finally, we find $\hat{\rho}$ such that we minimize the squared Frobenius norm of $\Sigma_{\mathbf{Y}} - \tilde{\Sigma}_{\mathbf{Y}}$, defined as $\|\Sigma_{\mathbf{Y}} - \tilde{\Sigma}_{\mathbf{Y}}\|_F^2 = \sum_{i=1}^{d}\sum_{j=1}^{d} |\sigma_{\mathbf{Y}ij} - \tilde{\sigma}_{\mathbf{Y}ij}|^2$. We estimate uncertainty via bootstrapping. Below we describe each of these steps in more detail.

4.1.1 | Estimation of the TPDM

The estimator we use for the TPDM was proposed by Cooley and Thibaud (2019), and extends an estimator given by Larsson and Resnick (2012) to dimension $d > 2$. This estimator replaces the true angular measure with an empirical estimate. Let $\mathbf{y}_t = (y_{t1}, y_{t2}, \ldots, y_{td})$, $t = 1, \ldots, n$, denote our observations with the appropriate regularly varying marginals. Note that this may require first performing a suitable marginal transformation. Let the radial component $r_t = \|\mathbf{y}_t\|_2$, and the angular component $\boldsymbol{\theta}_t = \mathbf{y}_t r_t^{-1}$. Set a suitably high threshold r_0 for the radial components, and let $n_{\text{exc}} = \sum_{t=1}^{n} \mathbb{I}(r_t > r_0)$ be the number of threshold exceedances. In practice, it is convenient to define r_0 in terms of an empirical quantile of the data, which we will denote by q_{r_0}. We define

$$\hat{\sigma}_{\mathbf{Y}ij} = \hat{m} \int_{\mathbb{S}_{+(2)}^{d-1}} \theta_i \theta_j d\hat{N}_{\mathbf{Y}}(\boldsymbol{\theta}) = \frac{\hat{m}}{n_{\text{exc}}} \sum_{t=1}^{n} \theta_{ti}\theta_{tj}\mathbb{I}(r_t > r_0), \tag{12}$$

where $N_{\mathbf{Y}}(\cdot) = m^{-1}H_{\mathbf{Y}}(\cdot)$, and $\hat{m}$ is an estimate of $H_{\mathbf{Y}}(\mathbb{S}^{d-1}_{+(2)})$. In application, because we either simulate or preprocess our data to have common unit scale, $m = d$ and does not need to be estimated.

Let $\hat{\Sigma}_{\mathbf{Y}} = (\hat{\sigma}_{\mathbf{Y}ij})_{i,j=1,\dots,d}$. Note that $\hat{\Sigma}_{\mathbf{Y}}$ is symmetric and positive semidefinite, as $\hat{\Sigma}_{\mathbf{Y}} = n_{\text{exc}}^{-1}\hat{m}\hat{\mathbf{\Theta}}^{\top}\hat{\mathbf{\Theta}}$, where $\hat{\mathbf{\Theta}}$ is a matrix whose rows are the vectors $\boldsymbol{\theta}_t$ for which $r_t > r_0$. The estimator (12) is of a form widely used in extreme analyses. Resnick (2004) and Larsson and Resnick (2012) show consistency of such estimators using intermediate asymptotic arguments common to extremes.

4.1.2 | Bias correction of the estimated TPDM

We anticipate that finite-sample estimates of $\Sigma_{\mathbf{Y}}$ will be biased. In particular, threshold-based estimators are known to overestimate dependence when true dependence is moderate to weak (Huser et al., 2016), such as is the case when the distance between grid cell pairs is moderate to large for the extreme SAR. Naively minimizing $\|\Sigma_{\mathbf{Y}}(\rho) - \hat{\Sigma}_{\mathbf{Y}}\|_F^2$ would result in an overestimate of ρ. Instead, we propose an intermediate step to obtain a bias-corrected estimate of the TPDM, denoted by $\tilde{\Sigma}_{\mathbf{Y}}$. Our bias correction approach assumes that the true pairwise tail dependence should be close to zero at large distances. This assumption is based on knowledge of the spatial extent of the modeled phenomenon relative to the size of the study area; the localized behavior of extreme precipitation events in the study region shown in Figure 1 indicate this is a reasonable assumption for our application. We provide more details about the bias correction when we apply to data in Section 4.2.1.

4.1.3 | Estimation of ρ

Recall from Section 3.2 that the theoretical TPDM for the common-scale formulation of the extreme SAR model is

$$\Sigma_{\mathbf{Y}}(\rho) = \tilde{A}\tilde{A}^{\top} = \text{diag}(1/\|A_{1\cdot}\|_2, \dots, 1/\|A_{d\cdot}\|_2)AA^{\top}\text{diag}(1/\|A_{1\cdot}\|_2, \dots, 1/\|A_{d\cdot}\|_2), \tag{13}$$

where $A = (I - \rho W)^{-1}$. After estimating $\hat{\Sigma}_{\mathbf{Y}}$ and performing bias correction to obtain $\tilde{\Sigma}_{\mathbf{Y}}$, the final step is to find

$$\hat{\rho} = \arg\min_{\rho \in (0\ ,\ \min\{|\lambda_1|^{-1}, |\lambda_d|^{-1}\})} \|\Sigma_{\mathbf{Y}}(\rho) - \tilde{\Sigma}_{\mathbf{Y}}\|_F^2, \tag{14}$$

where $\lambda_1 < \dots < \lambda_d$ are the ordered eigenvalues of W, and the restriction on ρ follows from Proposition 1. We find $\hat{\rho}$ using numerical optimization in `R` (R Core Team, 2018).

We estimate uncertainty via a nonparametric bootstrap procedure. Specifically, we first sample with replacement from the original n observations (each of which is d-dimensional), ensuring the preservation of the dependence structure. In cases where marginal transformation is required, doing the resampling after transformation would not include uncertainty in $\hat{\rho}$ resulting from the transformation. For each bootstrap replicate, we perform the inference procedure described above, starting with estimating the TPDM and ending with estimating ρ. We then obtain 95% confidence intervals using the percentile method.

4.2 | Application to simulated and observed data

We investigate our estimation and inference procedure on three different data categories of increasing inferential challenge. Each data set will consist of n realizations or observations of a 20×20 spatial field. Note that this differs from the typical setup in spatial statistics, which includes only one observation of a spatial field, but is standard for the analysis of multivariate and spatial extremes.

4.2.1 | Simulation from the true extreme SAR model

To demonstrate our method, we begin by simulating realizations from the true extreme SAR model (9). Let W be the 400×400 binary proximity matrix following the rook neighborhood specification. We select two values of ρ for

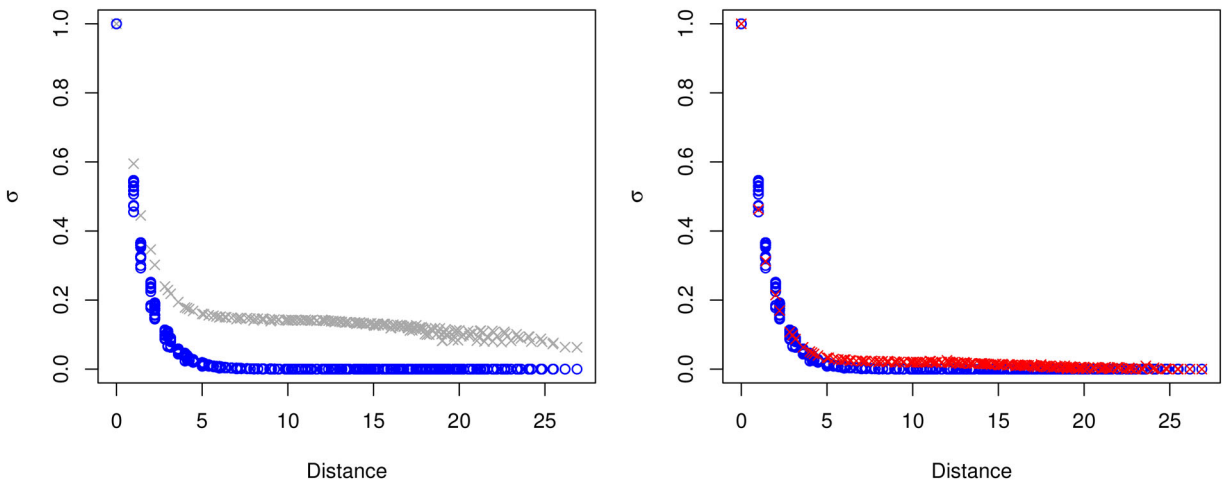

FIGURE 3 Results of estimating $\Sigma_{\mathbf{Y}}$ for data simulated from the true extreme simultaneous autoregressive model with $\rho = 0.2$, $q_{r_0} = 0.99$, and no added noise. True values $\{\sigma_{\mathbf{Y}ij}\}$ (blue circles) are plotted against pairwise Euclidean distances. The left panel compares the true values to means of the estimates $\{\hat{\sigma}_{\mathbf{Y}ij}\}$ at each distance (gray x's). The right panel compares the true values to means of bias-corrected estimates $\{\tilde{\sigma}_{\mathbf{Y}ij}\}$ at each distance (red x's)

simulations: $\rho = 0.1$ (weaker dependence) and $\rho = 0.2$ (stronger dependence). For each value of ρ, we simulate $n = 10,000$ independent realizations of a $d = 400$-dimensional random vector $\mathbf{Z} \overset{iid}{\sim}$ Fréchet($\alpha = 2$, scale $= 1$), that is, $\mathbb{P}(Z_j \leq z) = \exp(-z^{-2})$ for $j = 1, \ldots, 400$. Each realization of $\mathbf{Z}$ represents values found on a 20×20 square lattice. To each realization of $\mathbf{Z}$, we apply transformed-linear multiplication with the $\tilde{A}$ matrix corresponding to W and the given value of ρ, to obtain n independent realizations of $\mathbf{Y}$. These realizations $\{\mathbf{y}_t\}_{t=1}^{n}$ are independent temporal replicates of heavy-tailed spatial fields whose spatial extremal dependence is specified by (9).

Since we do not believe that real-world data will follow the true model exactly, it is important to investigate how our estimation and inference procedures react to deviations from the true model. A first step is to add light-tailed noise. For purposes of illustration, we add $\boldsymbol{\zeta}_t \overset{iid}{\sim} Exp(2)$ to each realization $\mathbf{y}_t$. At moderate thresholds we expect this independent noise to dilute the spatial dependence and lead to underestimation of ρ. Because the noise is light-tailed, we expect its effect to diminish as the threshold increases.

For simulations from the true extreme SAR model, we know both the true ρ and the true $\Sigma_{\mathbf{Y}}(\rho)$. Below, we use the base case of data simulated from the true model with $\rho = 0.2$ with no added noise to illustrate our modeling choices in each of the steps described in Section 4.1. Estimation of $\Sigma_{\mathbf{Y}}$ and hence of ρ depends on the threshold quantile. We choose the 0.99 quantile for our initial illustration. We then explore the effect of varying the threshold quantile, and present results for various combinations of the true ρ, threshold quantile, and presence of light-tailed noise.

Estimation of the TPDM

The left panel Figure 3 shows results of estimating the TPDM for the base case using the estimator defined in Section 4.1.1. As expected, this estimator tends to overestimate the tail pairwise dependence, especially at moderate to large distances when true dependence is weak.

Bias correction of the estimated TPDM

Based on the results of estimating $\Sigma_{\mathbf{Y}}$, we propose a simple bias correction method which reflects our knowledge that the true pairwise extremal dependence should be close to zero when the distance is large enough between grid cell centers. We temporarily view $\hat{\sigma}_{\mathbf{Y}ij}$ as a function of pairwise distance, and fit the curve $\hat{\sigma} = \beta_0 e^{-\beta_1 * \text{distance}} + \beta_2$ to the estimated TPDM values (see Supplement Figure S2, left). Then $\hat{\beta}_2$ is our estimate for the horizontal asymptote. In the base case of $\rho = 0.2$ and $q_{r_0} = 0.99$, we obtain $\hat{\beta}_2 = 0.134$. Note that $\sigma_{\mathbf{Y}ii}$ is unbiased on average, so it is only necessary to bias-correct the off-diagonal entries of $\hat{\Sigma}_{\mathbf{Y}}$. We make a simplifying assumption of constant bias for pairwise distances greater than zero. Our bias-corrected estimate of the TPDM is then $\tilde{\Sigma}_{\mathbf{Y}} = (\tilde{\sigma}_{\mathbf{Y}ij})_{i,j=1,\ldots,d}$, where

$$\tilde{\sigma}_{\mathbf{Y}ij} = \begin{cases} \hat{\sigma}_{\mathbf{Y}ij}, & i = j, \\ (\hat{\sigma}_{\mathbf{Y}ij} - \hat{\beta}_2)_+, & i \neq j. \end{cases} \tag{15}$$

We take the positive part in the latter case because we know that $\Sigma_{\mathbf{Y}}$ is entrywise nonnegative for the true extreme SAR under our model specifications. The right panel of Figure 3 compares the bias-corrected $\tilde{\Sigma}_{\mathbf{Y}}$ to the true $\Sigma_{\mathbf{Y}}$ for the base case.

Estimation of ρ

For the base case, minimizing the squared Frobenius norm $\|\Sigma_{\mathbf{Y}}(\rho) - \tilde{\Sigma}_{\mathbf{Y}}\|_F^2$ results in an estimate of $\hat{\rho} = 0.197$, which is quite close to the true value of $\rho = 0.2$. The corresponding 95% bootstrap confidence interval based on 1,000 bootstrap replicates is $(0.192, 0.202)$, which covers the true value.

Estimation of a risk region occurrence probability

We can also compare the probability of an event occurring in a given risk region between the true and fitted extreme SAR models. Let us define a risk region by

$$\mathcal{R} = \{\mathbf{y} \in \mathbb{R}_+^d : \max_{i \in \mathcal{D}_{\mathcal{R}}} y_i > y_0\}, \tag{16}$$

where $\mathcal{D}_{\mathcal{R}} \subset \mathcal{D} = \{1, \dots, d\}$. Our goal is to estimate $p_{\mathcal{R}} = \mathbb{P}(\mathbf{Y}_t \in \mathcal{R})$. We choose this particular form of the risk region for ease of calculation, but note that occurrence probabilities can be estimated in a similar manner for any risk region of interest.

For the extreme SAR model,

$$\nu(\mathcal{R}) = \int_{\boldsymbol{\theta} \in \mathbb{S}_{+(2)}^{d-1}} \max_{i \in \mathcal{D}_{\mathcal{R}}} \left(\frac{\theta_i}{y_0}\right)^2 dH(\boldsymbol{\theta}) = \frac{1}{y_0^2} \sum_{j=1}^{d} \max_{i \in \mathcal{D}_{\mathcal{R}}} \tilde{A}_{ij}{}^2. \tag{17}$$

Next we obtain our occurrence probability estimate. We assume n is fixed and large enough such that $n\mathbb{P}\left(\frac{\mathbf{Y}}{c_n} \in \cdot\right) \approx \nu(\cdot)$, where the appropriate normalizing constant in this case is $c_n = n^{1/2}$. So we get

$$\begin{aligned} &n\mathbb{P}(n^{-1/2}\mathbf{Y}_t \in n^{-1/2}\mathcal{R}) \approx \nu(n^{-1/2}\mathcal{R}) \\ &\quad \Rightarrow \mathbb{P}(\mathbf{Y}_t \in \mathcal{R}) \approx n^{-1}\nu(n^{-1/2}\mathcal{R}) \approx \nu(\mathcal{R}), \end{aligned} \tag{18}$$

where the last approximation follows from the scaling property of ν. Since the risk region occurrence probability depends on $\tilde{A}$, which in turn depends on ρ, we will use the notation $p_{\mathcal{R}}(\rho)$.

As an example, suppose $\mathcal{D}_{\mathcal{R}} = \{174, 193, 194, 195, 214\}$, corresponding to a plus-shaped region of grid cells within the 20×20 grid, and let $y_0 = 30$. For the true extreme SAR model with $\rho = 0.2$, we can use (17) and (18) to obtain $p_{\mathcal{R}}(\rho) \approx 4.57 \times 10^{-3}$. Our fitted model with $\hat{\rho} = 0.197$ yields the estimate $\hat{p}_{\mathcal{R}} = p_{\mathcal{R}}(\hat{\rho}) \approx 4.63 \times 10^{-3}$, which is very similar. Plugging in our 1,000 bootstrapped estimates for ρ, we obtain a corresponding 95% confidence interval for the probability estimate of $(4.52 \times 10^{-3}, 4.70 \times 10^{-3})$, which covers the true $p_{\mathcal{R}}(\rho)$. For reference, the empirical occurrence probability based on our simulation of 10,000 grids was 5.70×10^{-3}.

Sensitivity to threshold selection, true dependence, and noise

Estimation of $\Sigma_{\mathbf{Y}}$, and therefore of ρ, depends on the choice of threshold. Figure 4 illustrates the effect of threshold selection on estimates of ρ. We let the threshold r_0 be the empirical q_{r_0}-quantile of data simulated from the true model, for $q_{r_0} = 0.95, 0.975, 0.99, 0.995$, and 0.999. Figure 4 shows the corresponding estimates for ρ using our chosen bias-corrected method (in red) as well as the estimates we would have obtained if we did not bias correct $\hat{\Sigma}_{\mathbf{Y}}$ and instead minimized $\|\Sigma_{\mathbf{Y}}(\rho) - \hat{\Sigma}_{\mathbf{Y}}\|_F^2$ (in black). We see that as q_{r_0} increases, point estimates from both methods get closer to the true value of ρ. Also in both cases, the uncertainty increases with q_{r_0} since the sample size (number of exceedances) decreases. As we would expect based on the consistency of $\hat{\Sigma}_{\mathbf{Y}}$, omitting bias correction of $\hat{\Sigma}_{\mathbf{Y}}$ leads to accurate estimates of ρ when q_{r_0} is high enough (e.g., when $q_{r_0} = 0.999$). However, at less extreme threshold quantiles corresponding to a reasonable number of exceedances in practice, our bias-corrected method performs much better and is effective in reducing the bias in estimation of ρ.

In addition to varying the threshold quantile, we also explored different values of the true ρ and the effect of adding light-tailed noise. Table 1 gives point estimates for ρ along with 95% confidence intervals based on 1,000 bootstrap replicates for four cases. The first row starts with the base case of $\rho = 0.2$, $q_{r_0} = 0.990$, and no added noise. Each subsequent row of the table represents a change to either the true ρ, the norm threshold quantile, or the presence of light-tailed noise, from this base case. We see that the estimates for ρ are slightly biased low in all cases, and as expected the bias increases

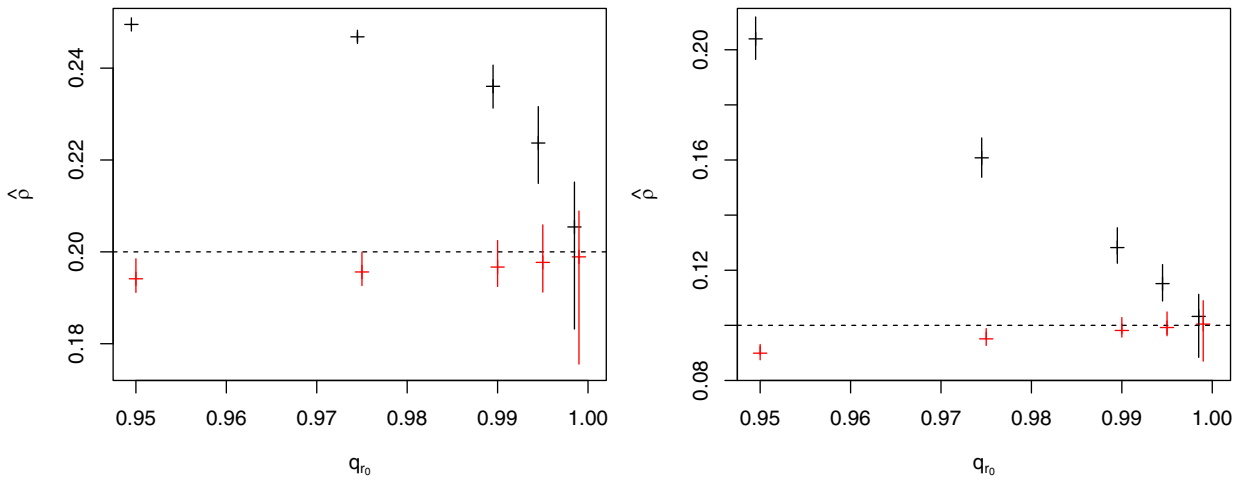

FIGURE 4 Effect of varying threshold quantile q_{r_0} on estimates of ρ using our proposed bias-corrected method (red) and estimates without bias-correction (black) for data simulated from the true extreme simultaneous autoregressive model with $\rho = 0.2$ (left) and $\rho = 0.1$ (right). Dashed lines indicate the true value of ρ. Solid vertical lines represent 95% bootstrap confidence intervals

TABLE 1 Results from fitting via the TPDM for data simulated from the true model

ρ	q_{r_0}	Noise	$\hat{\rho}$	95% CI
0.2	0.990	No	0.197	(0.192, 0.202)
0.1	0.990	No	0.098	(0.097, 0.103)
0.2	0.975	No	0.196	(0.193, 0.200)
0.2	0.990	Yes	0.196	(0.191, 0.202)

Note: Columns represent (from left to right): the true value of ρ, norm threshold quantile, presence of light-tailed noise, point estimate, and 95% bootstrap CI for ρ.
Abbreviations: CI, confidence interval; TPDM, tail pairwise dependence matrix.

when the threshold is lowered or noise is added. However, the 95% bootstrap confidence interval covers the true value of ρ in all cases.

Finally, we assess coverage for the base case by repeating our simulation 100 times. In each repetition, we simulate n d-dimensional fields from the true extreme SAR model, then estimate ρ and the corresponding occurrence probability for the risk region defined previously. We obtain 95% bootstrap intervals based on 1,000 bootstrap replicates, and find an empirical coverage rate of 96% for both ρ and $p_{\mathcal{R}}(\rho)$.

4.2.2 | Simulation from a Brown–Resnick process

Our second category of data comprises simulations from a Brown–Resnick process (Brown & Resnick, 1977; Kabluchko, Schlather, & de Haan, 2009). Brown–Resnick processes form a flexible class of max-stable processes constructed from intrinsically stationary Gaussian processes parameterized by variograms. As a more realistic deviation from the true model (9), we simulate $n = 10,000$ independent realizations from a Brown–Resnick process at the set of $d = 400$ grid cell centers. We use the variogram $2\gamma(\mathbf{s}, \mathbf{s}') = 2(\|\mathbf{s} - \mathbf{s}'\|_2/\lambda)^{\kappa}$, where $\mathbf{s}, s'$ denote locations of grid cell centers, and $\lambda > 0$ and $\kappa \in (0, 2]$ control the range and the smoothness, respectively (Thibaud, Aalto, Cooley, Davison, & Heikkinen, 2016). We simulate our processes to have Fréchet ($\alpha = 2$, scale $= 1$) marginals. Although the Brown–Resnick processes are max-stable, the processes observed at a finite number of locations are also multivariate regularly varying.

Importantly, there is no true ρ value for the Brown–Resnick model. In order to evaluate the model fitting procedure, we choose $\lambda = 0.431$ and $\kappa = 1.148$ to minimize the squared error between TPDM values from the SAR model with $\rho = 0.2$ and pairwise extremal dependence measures for the Brown–Resnick model (derived in Supplement Section S3). We do

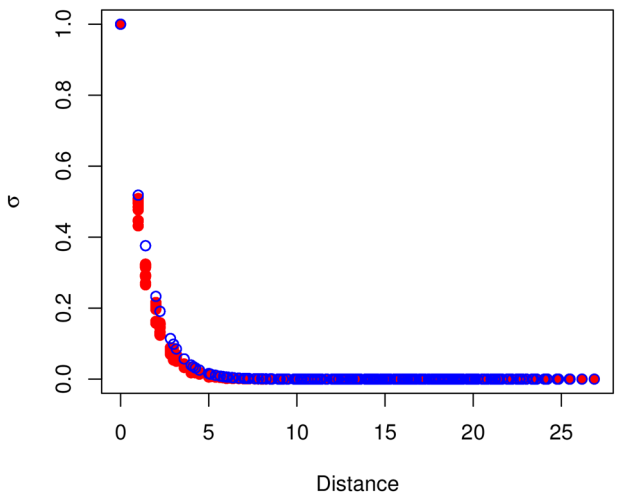

FIGURE 5 Comparison of the tail pairwise dependence matrix $\Sigma_{\mathbf{Y}}(\hat{\rho})$ for the fitted extreme simultaneous autoregressive model (red) and the true Σ^{BR} of the Brown–Resnick process with $\lambda = 0.431$ and $\kappa = 1.148$ (blue)

not compare the SAR model fit to that of a fitted Brown–Resnick model to this data, as fitting a Brown–Resnick model to 400 locations is an arduous task.

Estimation and bias correction of the TPDM

Applying the estimator defined in Section 4.1.1 to the base Brown–Resnick case similarly overestimates pairwise tail dependence at most distances (see Supplement Figure S3, left). Using the same bias correction procedure as before, we obtain a bias-corrected estimate $\tilde{\Sigma}_{\mathbf{Y}}$ with entries given by (15) with $\hat{\beta}_2 = 0.045$ (see Supplement Figure S2, right). As before, this procedure mitigates bias especially at moderate to large distances (see Supplement Figure S3, right).

Estimation of ρ

For the base Brown–Resnick case, minimizing $\|\Sigma_{\mathbf{Y}}(\rho) - \tilde{\Sigma}_{\mathbf{Y}}\|_F^2$ results in an estimate of $\hat{\rho} = 0.192$ with a corresponding 95% bootstrap confidence interval of $(0.188, 0.199)$. Although there is no true ρ in the case of the Brown–Resnick simulations, we can assess our performance based on the known pairwise tail dependence. Figure 5 compares the theoretical TPDM $\Sigma_{\mathbf{Y}}(\hat{\rho})$ based on the fitted extreme SAR model to the true pairwise extremal dependence measures $\{\sigma_{ij}^{BR}\}$ of the Brown-Resnick process that we simulated from. Even though we are fitting the wrong model to the data, we are able to estimate dependence well. Compared with the true pairwise dependence, the fitted extreme SAR model slightly underestimates dependence, especially at shorter distances.

Estimation of a risk region occurrence probability

We can also compare the theoretical occurrence probability for a given risk region to the estimated probability based on the fitted extreme SAR model. Again, we define a risk region $\mathcal{R}$ by (16). For the Brown–Resnick process, we can use the known exponent measure $V(\cdot)$ to calculate $p_{\mathcal{R}}^{BR} = \mathbb{P}(\mathbf{Y}_t \in \mathcal{R})$. Specifically,

$$\begin{aligned}\mathbb{P}(\mathbf{Y}_t \in \mathcal{R}) &= 1 - \mathbb{P}(Y_{ti} \leq y_0 \quad \text{for all } i \in \mathcal{D}_{\mathcal{R}}) \\ &= 1 - \exp\{-V(\mathbf{y_0})\},\end{aligned} \tag{19}$$

where $y_0 = (y_0, \ldots, y_0)$ is a vector of dimension $\text{card}(\mathcal{D}_{\mathcal{R}})$. For the same plus-shaped example of $\mathcal{D}_{\mathcal{R}} = \{174,193,194,195,214\}$ and $y_0 = 30$ as before, we obtain $p_{\mathcal{R}}^{BR} = 4.18 \times 10^{-3}$. In comparison, the probability estimate for the fitted extreme SAR model is $p_{\mathcal{R}}(\hat{\rho}) = 4.71 \times 10^{-3}$, with a corresponding 95% bootstrap confidence interval of $(4.60 \times 10^{-3}, 4.77 \times 10^{-3})$. By fitting the extreme SAR model to simulations from a Brown–Resnick process, we slightly overestimate the occurrence probability. For our particular form of risk region, this corresponds to slightly underestimating the tail dependence, as we already saw from Figure 5. For reference, the empirical occurrence probability for our 10,000 simulated Brown–Resnick fields was 4.30×10^{-3}.

Sensitivity to threshold selection and true dependence

We investigate the effect of threshold selection on our estimation and inference of ρ in the same manner as we did for simulations from the extreme SAR model. We let the threshold r_0 be the empirical q_{r_0}-quantile of data simulated from a given Brown–Resnick process, for $q_{r_0} = 0.95, 0.975, 0.99, 0.995$, and 0.999, and consider estimates with and without bias correction. Similarly as before, as q_{r_0} increases, the point estimates from the two methods get closer to each other, while

TABLE 2 Results from fitting via the TPDM for data simulated from a Brown–Resnick process

λ	κ	q_{r_0}	$\hat{\rho}$	95% CI
0.431	1.148	0.990	0.192	(0.188, 0.199)
0.254	1.341	0.990	0.101	(0.091, 0.112)
0.431	1.148	0.975	0.186	(0.183, 0.191)

Note: Columns represent (from left to right): the range and smoothness parameters of the Brown–Resnick simulations, norm threshold quantile, point estimate, and 95% bootstrap CI for ρ.
Abbreviations: CI, confidence interval; TPDM, tail pairwise dependence matrix.

the uncertainty of these estimates increases (see Supplement Figure S4). This is consistent with our familiar bias-variance tradeoff for threshold selection.

For easy comparison with Table 1, Table 2 gives point estimates for ρ along with 95% confidence intervals based on 1,000 bootstrap replicates for three cases. The first row starts with the base case of $q_{r_0} = 0.99$ and Brown–Resnick parameters $\lambda = 0.431$ and $\kappa = 1.148$, which give rise to the TPDM values closest to the extreme SAR model with $\rho = 0.2$. Each subsequent row of the table represents a change to either the Brown–Resnick parameters (to those which minimize the squared error to TPDM values for the extreme SAR model with $\rho = 0.1$) or the norm threshold quantile from this base case.

4.2.3 | Gridded precipitation observations

For our data application (recall Figure 1), we use the Climate Prediction Center (CPC) unified gauge-based analysis of daily precipitation (National Oceanic and Atmospheric Administration, 2010). This is a gridded product with 0.25° resolution over the contiguous United States and daily observations for the 59 years between 1948 and 2006. We choose a region of 20×20 grid cells containing northeastern Colorado, and restrict our attention to precipitation between May 1 and September 30 of each year for a total of $n = 9,027$ daily observations of this region.

Unlike the simulated data in Sections 4.2.1 and 4.2.2, which already have appropriate regularly varying marginals, the precipitation data requires transforming the marginal distributions to be approximately Fréchet with $\alpha = 2$ and scale $= 1$ before fitting our dependence model. There are many possible ways to implement such a marginal transformation. In Section S5 of the Supplement, we describe our chosen procedure, which uses a parametric form for the upper tail and a nonparametric form below a marginal threshold. Figure 6 shows the same precipitation data as in Figure 1, after applying the marginal transformation.

Following the same procedure as for simulations from the true model and simulations from a Brown–Resnick process, we first estimate the TPDM from the data, using $q_{r_0} = 0.99$ as our threshold quantile. We then bias-correct to obtain $\tilde{\Sigma}_{\mathbf{Y}}$. As a first attempt, we assume W to be a binary proximity matrix following the rook neighborhood structure. Minimizing $\|\Sigma_{\mathbf{Y}}(\rho) - \tilde{\Sigma}_{\mathbf{Y}}\|_F^2$ leads to an estimate of $\hat{\rho} = 0.236$, with corresponding 95% bootstrap confidence interval (0.230, 0.241). The left panel of Figure 7 compares the fitted TPDM $\Sigma_{\mathbf{Y}}(\hat{\rho})$ to the means at each distance of the bias-corrected $\{\tilde{\sigma}_{\mathbf{Y}ij}\}$. The discrepancy shown here indicates that the model with the simple rook structure is not able to fully capture the longer range dependence (i.e., larger storm sizes) apparent in the data.

A natural way to adapt our extreme SAR model for longer range dependence is to modify the neighborhood structure encoded by the proximity matrix W. Instead of the simple rook structure, which puts weights of 1 at the four nearest neighbors, exploration leads us to use a larger template based on a Gaussian kernel that assigns weights to second- and third-order neighbors as shown in Figure 8. Now, letting W be the new spatial weights matrix corresponding to this larger template, (7) gives us an upper bound for ρ of 0.078, and fitting the extreme SAR model results in an estimate of $\hat{\rho} = 0.070$ (0.068, 0.072). The right panel of Figure 7 compares the fitted TPDM $\Sigma_{\mathbf{Y}}(\hat{\rho})$ to the bias-corrected $\tilde{\Sigma}_{\mathbf{Y}}$ estimated from the precipitation data. We can see that using the larger template allows us to better match the observed tail dependence structure, especially in terms of capturing longer-range dependence.

As previously done with the simulated data, we consider a risk region $\mathcal{R}$ defined by (16). Now the set of component indices $\mathcal{D}_{\mathcal{R}} = \{174,193, 194,195, 214\}$ corresponds to the grid cell containing Wray, Colorado, and its four nearest neighbors. As before, we let $y_0 = 30$. Our risk region occurrence probability can be interpreted as the probability of transformed

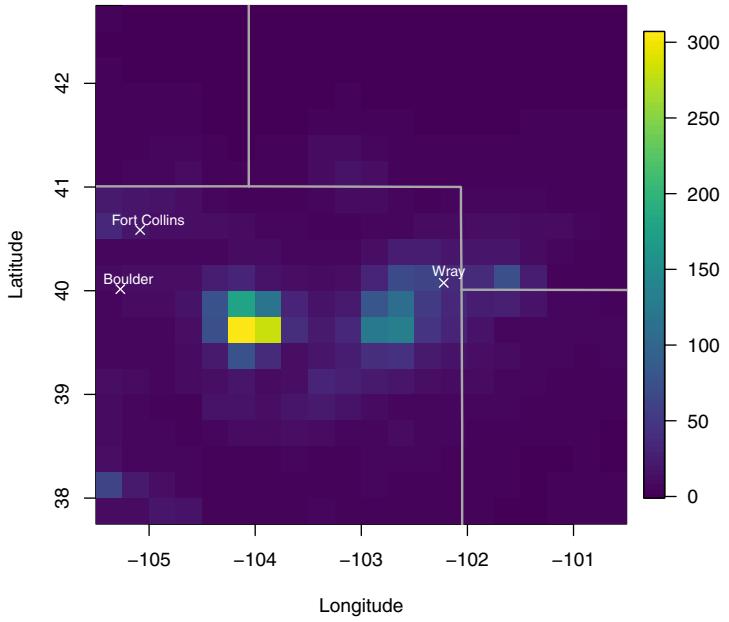

FIGURE 6 Example of marginal transformation: gridded precipitation data over the study region from the Climate Prediction Center (CPC) unified gauge-based analysis for the same example day as Figure 1, after transformation to Fréchet($\alpha = 2$, scale $= 1$) marginals

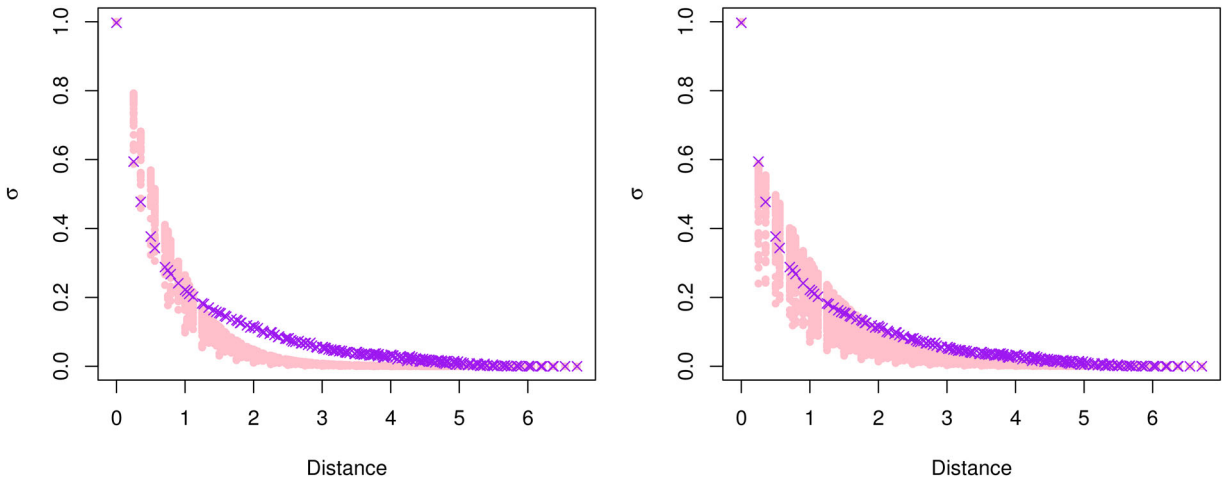

FIGURE 7 Results from fitting an extreme simultaneous autoregressive model to gridded precipitation observations, using two different neighborhood structures: the simple rook neighborhood (left) and a larger template (right). The tail pairwise dependence matrix $\Sigma_{\mathbf{Y}}(\hat{\rho})$ for the fitted model (pink) is compared with means of the bias-corrected estimate $\bar{\Sigma}_{\mathbf{Y}}$ (purple)

daily precipitation exceeding 30 units in at least one of the five grid cells under consideration, which is approximately the 0.999 quantile at each of these five locations. Based on the fitted extreme SAR model, we estimate this probability to be $p_R(\hat{\rho}) = 4.0 \times 10^{-3}$ with 95% bootstrap confidence interval $(3.6 \times 10^{-3}, 4.3 \times 10^{-3})$. Although we do not know the true risk region occurrence probability, we can compare to the empirical probability estimate, which is 3.1×10^{-3} with 95% bootstrap confidence interval $(2.0 \times 10^{-3}, 4.3 \times 10^{-3})$. Based on these results, it seems that our simple spatial extremes model provides reasonable estimates of risk.

5 | DISCUSSION

Inspired by the SAR model from classical spatial statistics, we have proposed a new model for extremes of areal data which utilizes recent results on transformed-linear operations applied to regularly varying random vectors. The model's common-scale extension renders the model amenable to extremes practice where a common marginal distribution is assumed in order to more simply characterize the dependence structure. Because the limiting angular measure is known,

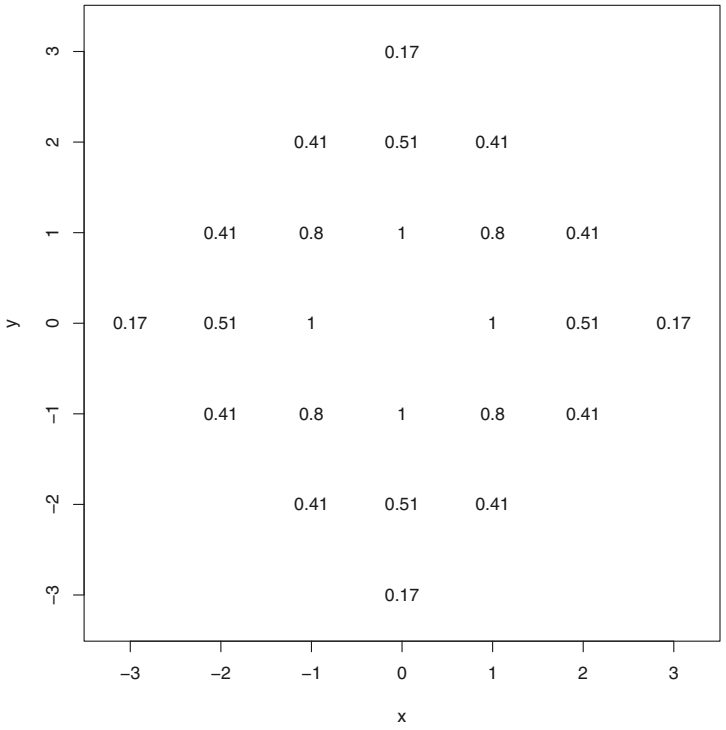

FIGURE 8 Illustration of the larger template used for fitting an extreme SAR model to gridded precipitation observations. The weights shown are assigned to 24 neighbors of a cell at the origin. By contrast, the default rook structure would only assign weights of 1 to each of the four nearest neighbors

the model can be used to assess the probability of landing in any risk region of interest. Our fitting procedure, which optimizes agreement between the sample and model TPDM, is quite fast: for $d = 400$, obtaining $\hat{\rho}$ takes approximately 3 s on a Macbook Pro laptop (2.6-GHz Intel Core i5, 8 Gb memory). We note that our simulated and actual spatial data sets are large by extremes standards, and our methods could be applied to problems of much greater size.

The appeal of this model is its simplicity; given a spatial proximity matrix W, spatial dependence is described by a single parameter ρ. Although realized fields from the extreme SAR model could be viewed as overly simplistic, we illustrate that the model is able to produce reasonable estimates of extremal dependence and corresponding risk region occurrence probabilities even in the case of model misspecification. By using a larger template, the model was able to capture longer range dependence found in the precipitation data. Like max-stable process models, the extreme SAR model only describes dependence in the asymptotically dependent setting. Current models which account for asymptotic independence (Shooter et al., 2019; Wadsworth & Tawn, 2019) are more complex.

We proposed a simple bias correction procedure for the estimated TPDM, but results could alternatively be obtained by letting the bias correction vary by distance. We also explored bias-correction via shrinkage or tapering, but there was not a clear choice of shrinkage target or taper function. These extensions could be avenues of future investigation.

For our application, we performed a marginal transformation of the data prior to fitting the spatial model. This is common practice in spatial and multivariate extremes applications, but it does have consequences. Our bootstrap-based confidence intervals did not include the uncertainty arising from estimation of the marginal distributions, as this would have greatly increased the computational cost of the bootstrap procedure. We did perform a sensitivity analysis where we repeated the application with a rank-based marginal transformation at each location, rather than the parametric and spatially smoothed marginal transformation we presented. The estimate of ρ changed from 0.070 to 0.073, showing that the choice of how the marginal transformation is performed will somewhat affect dependence estimates.

A potential future use of the extreme SAR model could be in a hierarchical setting. In many extremes problems, marginal behavior is of primary interest, such as in estimating high quantiles (return levels) or occurrence probabilities at individual locations. In such cases, it may be effective and computationally efficient to use hierarchical approaches (e.g., Cooley et al., 2007) that spatially model the parameters of univariate extremes. Most such hierarchical models assume extremes at different spatial locations are independent conditional on the latent processes. This conditional independence assumption ignores the fact that multiple locations can be affected by the same event (e.g., storm). Replacing the conditional independence assumption with a spatial extremes model is challenging (Ribatet, Cooley, & Davison, 2012), but the extreme SAR model's simplicity could make it appealing for this purpose.

ACKNOWLEDGEMENTS

Miranda J. Fix was supported in part by the UW NIEHS sponsored Biostatistics, Epidemiologic and Bioinformatic Training in Environmental Health (BEBTEH) Training Grant, Grant #: NIEHS T32ES015459. While at Colorado State University, Miranda J. Fix, Daniel S. Cooley, and Emeric Thibaud were partially supported by NSF grant DMS-1243102. Daniel S. Cooley is also partially supported by DMS-1811657.

CPC US Unified Precipitation data was provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their web site at https://www.esrl.noaa.gov/psd/.

ORCID

Miranda J. Fix https://orcid.org/0000-0002-5282-5456

REFERENCES

Anselin, L. (1988). *Spatial econometrics: Methods and models*. Berlin, Germany: Springer Science & Business Media.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, *36*(2), 192–236.

Besag, J., & Kooperberg, C. (1995). On conditional and intrinsic autoregressions. *Biometrika*, *82*(4), 733–746.

Brown, B. M., & Resnick, S. I. (1977). Extreme values of independent stochastic processes. *Journal of Applied Probability*, *14*(4), 732–739.

Buhl, S., Davis, R. A., Klüppelberg, C., & Steinkohl, C. (2019). Semiparametric estimation for isotropic max-stable space-time processes. *Bernoulli*, *25*(4A), 2508–2537.

Cabral, R., Ferreira, A., & Friederichs, P. (2020). Space–time trends and dependence of precipitation extremes in North-Western Germany. *Environmetrics*, *31*(3), e2605.

Cliff, A., & Ord, J. K. (1973). *Spatial autocorrelation*. London: Pion Limited.

Cooley, D., Nychka, D., & Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association*, *102*(479), 824–840.

Cooley, D., & Thibaud, E. (2019). Decompositions of dependence for high-dimensional extremes. *Biometrika*, *106*, 587–604.

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Cressie, N. (1993). *Statistics for spatial data*. New York, NY: Wiley.

Davison, A., Huser, R., & Thibaud, E. (2019). *Spatial extremes*. In A. Gelfand, M. Fuentes, J. Hoeting, & R. Smith (Eds.), *Handbook of environmental and ecological statistics* (pp. 711–744). Boca Raton, FL: CRC Press.

Einmahl, J. H., Kiriliouk, A., Krajina, A., & Segers, J. (2016). An M-estimator of spatial tail dependence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *78*(1), 275–298.

Einmahl, J. H., Kiriliouk, A., & Segers, J. (2018). A continuous updating weighted least squares estimator of tail dependence in high dimensions. *Extremes*, *21*(2), 205–233.

Fougères, A.-L., Mercadier, C., & Nolan, J. (2013). Dense classes of multivariate extreme value distributions. *Journal of Multivariate Analysis*, *116*, 109–129.

Haining, R. (1990). *Spatial data analysis in the social and environmental sciences*. Cambridge , MA: Cambridge University Press.

Hazra, A., Reich, B. J., & Staicu, A.-M. (2020). A multivariate spatial skew-t process for joint modeling of extreme precipitation indexes. *Environmetrics*, *31*(3), e2602.

Huser, R., & Davison, A. C. (2013). Composite likelihood estimation for the Brown-Resnick process. *Biometrika*, *100*(2), 511–518.

Huser, R., Davison, A. C., & Genton, M. G. (2016). Likelihood estimators for multivariate extremes. *Extremes*, *19*(1), 79–103.

Kabluchko, Z., Schlather, M., & de Haan, L. (2009). Stationary max-stable fields associated to negative definite functions. *The Annals of Probability*, *37*(5), 2042–2065.

Kelejian, H. H., & Robinson, D. P. (1995). *Spatial correlation: A suggested alternative to the autoregressive model*. In *New directions in spatial econometrics* (pp. 75–95). New York, NY: Springer.

Larsson, M., & Resnick, S. I. (2012). Extremal dependence measure and extremogram: The regularly varying case. *Extremes*, *15*(2), 231–256.

National Oceanic and Atmospheric Administration (2010). *CPC unified gauge-based analysis of daily precipitation over CONUS* Retrieved from https://www.esrl.noaa.gov/psd/

Padoan, S. A., Ribatet, M., & Sisson, S. A. (2010). Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, *105*(489), 263–277.

Plemmons, R. J. (1977). M-matrix characterizations. I—nonsingular M-matrices. *Linear Algebra and Its Applications*, *18*(2), 175–188.

Reich, B. J., & Shaby, B. A. (2019). A spatial Markov model for climate extremes. *Journal of Computational and Graphical Statistics*, *28*(1), 117–126.

Resnick, S. (2004). The extremal dependence measure and asymptotic independence. *Stochastic Models*, *20*(2), 205–227.

Resnick, S. I. (2007). *Heavy-tail phenomena: Probabilistic and statistical modeling Springer series in operations research and financial engineering*. New York, NY: Springer.

Ribatet, M., Cooley, D., & Davison, A. C. (2012). Bayesian inference for composite likelihood models and an application to spatial extremes. *Statistica Sinica*, *22*, 813–846.

Russell, B. T., Cooley, D. S., Porter, W. C., & Heald, C. L. (2016). Modeling the spatial behavior of the meteorological drivers' effects on extreme ozone. *Environmetrics*, *27*(6), 334–344.

Sang, H., & Gelfand, A. E. (2010). Continuous spatial process models for spatial extreme values. *Journal of Agricultural, Biological, and Environmental Statistics*, *15*(1), 49–65.

Sharkey, P., & Winter, H. C. (2019). A Bayesian spatial hierarchical model for extreme precipitation in Great Britain. *Environmetrics*, *30*(1), e2529.

Shooter, R., Ross, E., Tawn, J., & Jonathan, P. (2019). On spatial conditional extremes for ocean storm severity. *Environmetrics*, *30*(6), e2562.

Song, J. J., & De Oliveira, V. (2012). Bayesian model selection in spatial lattice models. *Statistical Methodology*, *9*(1), 228–238.

Thibaud, E., Aalto, J., Cooley, D. S., Davison, A. C., & Heikkinen, J. (2016). Bayesian inference for the Brown–Resnick process, with an application to extreme low temperatures. *The Annals of Applied Statistics*, *10*(4), 2303–2324.

Wadsworth, J. L. & Tawn, J. (2019). *Higher-dimensional spatial extremes via single-site conditioning*. arXiv preprint arXiv:1912.06560.

Wall, M. M. (2004). A close look at the spatial structure implied by the CAR and SAR models. *Journal of Statistical Planning and Inference*, *121*(2), 311–324.

Wang, Y., & Stoev, S. A. (2011). Conditional sampling for spectrally discrete max-stable random fields. *Advances in Applied Probability*, *43*(2), 461–483.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, *41*(3/4), 434–449.

Yuen, R., & Stoev, S. (2014). CRPS M-estimation for max-stable models. *Extremes*, *17*(3), 387–410.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Fix MJ, Cooley DS, Thibaud E. Simultaneous autoregressive models for spatial extremes. *Environmetrics*. 2020;e2656. https://doi.org/10.1002/env.2656