

ABSTRACT

MORRIS, SAMUEL ALAN. Spatial Methods for Modeling Extreme and Rare Events. (Under the direction of Brian J. Reich.)

In recent years, extreme values analysis (EVA) have started to benefit from methods that allow for dependence between observations. In a spatial setting, this is important for many phenomenon (e.g. ozone, rainfall, wildfires) because spatial methods borrow information across space for informing inference and making predictions at unobserved locations. When working with dependent extremes, max-stable processes are the analog to Gaussian processes. Max-stable processes have generalized extreme value (GEV) marginal distributions. These processes are exceptionally flexible, but they can be computationally challenging to work with for high dimensional data.

Some current methods address the dimensionality issues for max-stable processes using pairwise likelihoods, and others circumvent the computational challenges with low-rank approximations to max-stable processes. While these provide feasible solutions, they can still be very slow to fit, and can be sensitive to distributional assumptions. Despite these limitations, it is important to use models that are designed to handle dependence in the tail of the distribution. Our contributions to the literature on spatial extremes are threefold.

First, we develop a space-time skew- t method for threshold exceedances. In this method, we use a censored, local skew- t process to model asymptotic dependence. We use censoring so as not to allow the bulk of the data to influence the fit of the model in the extremes. We introduce a partitioning to alleviate long-range asymptotic dependence. We compare our method to Gaussian and max-stable processes with a simulation study. We also conduct a data analysis of ozone measurements throughout the U.S. in July 2005. We find that in most cases the method performs comparably or better than Gaussian processes and max-stable processes.

Second, we extend the GEV link for modeling binary data with a max-stable spatial dependence structure. Traditionally, spatial methods for binary data use a latent Gaussian process to model dependence, but this may not be appropriate for rare data due to the lack of asymptotic dependence for Gaussian processes. We compare our model to spatial probit and logistic methods through a simulation study. We also conduct a data analysis of *Tamarix ramosissima* and *Hedysarum scoparium*. We find some evidence to suggest that for very rare data, under certain sampling strategies, the max-stable extension provides an improvement in area under the receiver operating characteristic curve (AUROC).

Lastly, we present a method to construct empirical basis functions (EBF) as a low-rank approximation for max-stable processes. Similar to principal component analysis (PCA), these EBFs provide an exploratory method to examine important spatial trends as well as for Bayesian inference and predictions as unobserved locations. One notable distinction from PCA is that these EBFs are not

orthogonal due to the fact that they are restricted to be positive. We demonstrate our method for a data analysis on wildfire data in the state of Georgia as well as precipitation data in the eastern U.S. The results show that in the presence of spatial dependence, the EBF method demonstrates an improvement in quantile scores over a more traditional approach of using spatial knots with Gaussian weights.

© Copyright 2016 by Samuel Alan Morris

All Rights Reserved

Spatial Methods for Modeling Extreme and Rare Events

by
Samuel Alan Morris

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Statistics

Raleigh, North Carolina

2016

APPROVED BY:

Alyson G. Wilson

Arnab Maity

Adam J. Terando

Brian J. Reich
Chair of Advisory Committee

DEDICATION

To Zoe, Catherine, Stella, Benjamin, and Alice.
May you always have the courage and the support to achieve your dreams.

BIOGRAPHY

Sam graduated from Jacksonville High School, Jacksonville, North Carolina in 1999. He graduated magna cum laude with his Bachelor of Music in Clarinet Performance, Piano, and French from the University of North Carolina at Greensboro in 2005. In 2007, he received his Master of Education in Student Affairs Administration from North Carolina State University. He then worked for three years as the Systems Administrator and Assessment Coordinator for the Study Abroad Office at North Carolina State University at which point he decided to pursue graduate studies in statistics. In 2012, he received his Master of Statistics from North Carolina State University. During his graduate work in statistics, he has served as a Lecturer on the faculty at the University of Wisconsin–La Crosse.

ACKNOWLEDGEMENTS

First and foremost, I would like to acknowledge my partner, Adam Van Liere. Your constant support and encouragement to persevere through my studies have meant so much to me.

I am deeply grateful to my advisor Dr. Brian Reich, for his guidance, teaching, and mentoring. His advice, assistance, and encouragement in the preparation of this dissertation were invaluable. I would like to thank my committee, Dr. Alyson Wilson, Dr. Arnab Maity, and Dr. Adam Terando. Your comments and feedback were invaluable. I am also grateful to Dr. Emeric Thibaud and Dr. Daniel Cooley from Colorado State University for their insight and feedback on Chapters 2 and 4.

I am truly grateful to my professors and the staff of the Statistics Department at North Carolina State University for providing me with a strong basis for my statistical education and helping me make sure I met all the requirements to graduate. In particular, I would like to recognize Alison McCoy for the emotional support she provided, and Dr. Roger Woodard and Dr. Herle McGowan for the mentoring they provided me. I would like acknowledge the Statistics Department, the EPA, and STATMOS for the funding they have provided during my studies.

I could not have made it through my graduate studies without some amazing friends. Many thanks go out to Vickie Weber, Susheela Singh, Neal Grantham, Matt Austin, Andrew Wilcox, Luke Smith, Ryan Parker, Robert Allen, and Doug Baumann, to name a few.

Lastly, I also could not have finished this without the love and support from my family. To them, I am forever grateful.

TABLE OF CONTENTS

LIST OF TABLES	vii
LIST OF FIGURES	viii
Chapter 1 Introduction	1
1.1 A space-time skew- <i>t</i> model	2
1.2 A spatial model for rare binary events	2
1.3 Empirical basis functions	3
Chapter 2 A Space-time Skew-<i>t</i> Model for Threshold Exceedances	4
2.1 Introduction	4
2.2 Spatial skew processes	7
2.2.1 Skew- <i>t</i> process	7
2.2.2 Extremal dependence	8
2.3 Extending the model	8
2.3.1 Censoring to focus on the tail	8
2.3.2 Partitioning to remove long-range asymptotic dependence	9
2.3.3 Extension to space-time data	10
2.4 Hierarchical model	12
2.4.1 Computation	13
2.5 Simulation study	13
2.5.1 Design	13
2.5.2 Cross validation	15
2.5.3 Results	15
2.6 Data analysis	16
2.6.1 Model comparisons	16
2.6.2 Results	19
2.7 Discussion	20
Chapter 3 A Spatial Model for Rare Binary Events	23
3.1 Introduction	23
3.2 Spatial dependence for binary regression	24
3.3 Joint distribution	26
3.4 Quantifying spatial dependence	26
3.5 Computation	27
3.6 Simulation study	28
3.6.1 Latent processes	28
3.6.2 Sampling methods	29
3.6.3 Methods	29
3.6.4 Priors	30
3.6.5 Model comparisons	30
3.6.6 Results	31
3.7 Data analysis	33

3.7.1	Methods	33
3.7.2	Results	34
3.8	Discussion and future research	34
Chapter 4	Empirical Basis Functions for Max-Stable Spatial Dependence	37
4.1	Introduction	37
4.2	Model	38
4.3	Estimating the spatial dependence function	40
4.4	Bayesian implementation details	42
4.5	Data analysis	43
4.5.1	Gaussian kernel basis functions	44
4.5.2	Analysis of extreme Georgia fires	44
4.5.3	Results for fire analysis	47
4.5.4	Analysis of annual precipitation	49
4.5.5	Results for precipitation analysis	49
4.6	Discussion	53
BIBLIOGRAPHY	56
APPENDICES	60
Appendix A	A Space-time Skew- <i>t</i> Model for Threshold Exceedances	61
A.1	MCMC details	61
A.2	Posterior distributions	63
A.3	Proof that $\lim_{h \rightarrow \infty} \pi(h) = 0$	66
A.4	Skew- <i>t</i> distribution	67
A.5	Comparisons with other parameterizations	68
A.6	Temporal dependence	69
A.7	Brier scores for ozone prediction	70
A.8	Simulation study results	70
Appendix B	Spatial Model for Rare Binary Events	74
B.1	Binary regression using the GEV link	74
B.2	Derivation of the likelihood	74
Appendix C	Empirical Basis Functions for Max-stable Spatial Dependence	77
C.1	Extreme value distributions	77
C.2	Grid approximation to PS density	78
C.3	Standardized Gaussian kernel functions	78
C.4	Principal components	79

LIST OF TABLES

Table 2.1	Top two performing models for predicting ozone exceedance of level L with Relative Brier score	20
Table 3.1	Brier scores ($\times 100$) [SE] and AUROC [SE] for GEV, Probit, and Logistic methods from the simulation study.	31
Table 3.2	Brier scores ($\times 100$) [SE] and AUROC [SE] for GEV, Probit, and Logistic methods for <i>Tamarix ramosissima</i>	34
Table 4.1	Average Brier scores ($\times 100$), average quantile scores for $q(0.95)$ and $q(0.99)$, and time (in minutes) for 1,000 iterations for fire analysis.	48
Table 4.2	Average Brier scores ($\times 100$), average quantile scores for $q(0.95)$ and $q(0.99)$, and time (in minutes) for 1,000 iterations for precipitation analysis.	54
Table A.1	Setting 1 – Gaussian marginal, $K = 1$ knot	72
Table A.2	Setting 2 – Skew- t marginal, $K = 1$ knot	72
Table A.3	Setting 3 – Skew- t marginal, $K = 5$ knots	72
Table A.4	Setting 4 – Max-stable, Asymmetric logistic	72
Table A.5	Setting 5 – Max-stable, Brown-Resnick	73

LIST OF FIGURES

Figure 2.1	Ozone values (ppb) on July 10, 2005	5
Figure 2.2	Extremal dependence measure $\chi(h)$, as a function of distance, h , for $K = 1, 3, 5$, and 10 knots.	10
Figure 2.3	Brier scores relative to the Gaussian method for simulation study results. A ratio lower than 1 indicates that the method outperforms the Gaussian method.	17
Figure 2.4	Gaussian Q-Q plot (left) and skew- t with $a = 10$ and $\lambda = 1$ Q-Q plot (right) of the residuals.	18
Figure 2.5	Daily quantiles for two monitoring locations near Columbus, OH (left) and daily quantiles for a monitoring location in Los Angeles, CA and Columbus, OH (right)	18
Figure 2.6	Relative Brier scores for time-series models (left) and non-time-series models (right). Relative brier score for the max-stable model is between 1.13 and 1.18	20
Figure 2.7	Panels (a) – (d) give the posterior predictive $\hat{q}(0.99)$ for the month of July under four different models, panel (e) gives the difference between $\hat{q}(0.99)$ in panels (d) and (a), panel (f) gives the difference between $\hat{q}(0.99)$ in panels (d) and (b).	21
Figure 3.1	One simulated dataset from spatial GEV (left), spatial logistic (center), and hotspot (right) designs.	29
Figure 3.2	Smooth of BS (left) and AUROC (right) by rareness for GEV link and cluster sampling with $n_s = 250$	32
Figure 3.3	Smooth of BS (left) and AUROC (right) by rareness for GEV link and simple random sampling with $n_s = 250$	32
Figure 3.4	Smooth of BS (left) and AUROC (right) by rareness for logistic link and cluster sampling with $n_s = 250$	33
Figure 3.5	True occupancy of <i>Tamarix ramosissima</i> (left) and <i>Hedysarum scoparium</i> (right) from a 1-km ² study region of PR China.	34
Figure 3.6	Vertically averaged ROC curves for <i>Tamarix ramosissima</i>	35
Figure 4.1	Time series of log acres burned for 25 randomly selected counties with colors coding the county's quadrant (left), and spatially smoothed threshold values, T_i for each county (right).	44
Figure 4.2	Mean residual plot for the data pooled across counties after standardizing using the county's median and interquartile range. The two panels show different ranges on the x-axis and include a vertical line at the sample 95th percentile.	45
Figure 4.3	First six EBFs for the Georgia fire data and the cumulative sum of contributions v_1, \dots, v_{25}	46

Figure 4.4	Posterior mean of $\beta_{1,\text{time}}$ (top left), posterior mean of $\beta_{2,\text{time}}$ (middle left), estimate of $\Delta Q90$ (bottom left), $P[\beta_{1,\text{time}} > 0]$ (top right), $P[\beta_{2,\text{time}} > 0]$ (middle right), and $P[\Delta Q90 > 0]$ for fire data using EBF. In three counties (labeled), $\Delta Q90 > 2500$: County 1 - Ware (11,109), County 2 - Clinch (7,128), and County 3 - Charlton (6,545)	50
Figure 4.5	Time series of yearly max precipitation for current (1969 – 2000) (left). Time series of yearly max precipitation for future (2039 – 2070) (right).	51
Figure 4.6	First six EBFs for the combined precipitation data and the cumulative sum of contributions v_1, \dots, v_{25}	52
Figure 4.7	Posterior mean of $\Delta\mu$ (top left), posterior mean of $\Delta \log(\sigma)$ (middle left), estimate of $\Delta Q90$ (bottom left), $P[\Delta\mu > 0]$ (top right), $P[\Delta \log(\sigma) > 0]$ (middle right), and $P[\Delta Q90 > 0]$ between 2000 and 2070 for precipitation data using EBF	55
Figure A.1	Illustration of the partition of A	66
Figure A.2	Simulated lag- m χ for varying levels of φ	70
Figure A.3	Map of Brier scores for Gaussian (top) vs Symmetric- t , $K = 10$ knots, $T = 75$, time series (bottom).	71
Figure C.1	First six principal components and the cumulative sum of the first 25 eigenvalues for the Georgia fire data.	80
Figure C.2	First six principal components and the cumulative sum of the first 25 eigenvalues for the precipitation data.	81

CHAPTER

1

INTRODUCTION

In extreme values analysis (EVA), extremes are separated from the bulk of the distribution by either analyzing only points above a threshold or block maximums [Col01], e.g., the annual maximum of the daily precipitation. A natural spatial model for block maximum at several spatial locations is the max-stable process, which, under certain conditions, arises as the limit of the location-wise maximum of infinitely-many spatial processes [HF06]. Max-stable processes provide a very flexible framework, but they can be computationally challenging to implement. This is because the full likelihood exists for only a few trivial situations.

In addition to computational limitations, the use of max-stable processes also necessitates the use of dependence measures other than things like covariances and correlations. In extremes analysis, using measures like covariance to describe dependence may introduce other problems because they focus on deviations around the mean of the distribution. This is problematic because the finite dimensional marginal distributions of a max-stable process follow a generalized extreme value (GEV) distribution which is not guaranteed to have a finite first moment.

In this chapter, we provide a basic overview of our contributions to solving some of the current issues in EVA. The first contribution explores the use of the skew- t distribution for modeling spatial extremes. This distribution has nice properties for modeling extremes, but it requires some modification to address long-range asymptotic dependence and possible misspecification of distribution.

The second contribution is the development of a max-stable model for spatial dependence in rare binary data. Spatial methods for binary data traditionally use a Gaussian process, but in the case of rare data, the Gaussian process may not be appropriate due to the lack of asymptotic dependence for the Gaussian process. The third contribution is a new method to construct empirical basis functions (EBF) as an exploratory technique to examine spatial dependence in the extremes. This method is motivated by Principal Components Analysis (PCA), but differs in that the functions are positive and thus not orthogonal to one another. Together these three contributions provide new solutions for existing challenges in spatial methods for EVA.

1.1 A space-time skew- t model

To assess the compliance of air quality regulations, the Environmental Protection Agency (EPA) must know if a site exceeds a pre-specified level. In the case of ozone, compliance is fixed such that the year's 99th quantile should not exceed 75 parts per billion (ppb), which is high, but not extreme at all locations. One standard approach in EVA is to consider the yearly maximum at each spatial location. However, this approach is not appropriate because compliance is not based on the maximum at each site. The other standard approach is to model threshold exceedances, but in this setting, the threshold is usually selected to be extreme in the data as opposed to being fixed in advance by a regulatory agency. Furthermore, the computing required to fit a max-stable process to a large spatiotemporal dataset can be onerous.

In Chapter 2, we propose a new method that shares some asymptotic dependence characteristics of the max-stable process while retaining some of the nice computing properties of a Gaussian process. Our method is space-time model for threshold exceedances based on the skew- t process. Without some modification, the skew- t process has some undesirable characteristics such as long-range asymptotic dependence. To address this concern, we incorporate a random partition to permit long-distance asymptotic independence while allowing for sites that are near one another to be asymptotically dependent, and we incorporate thresholding to allow the tails of the data to speak for themselves. Finally, many extremal phenomena (e.g. high temperatures, extreme precipitation) exhibit some type of temporal dependence. Therefore we introduce a transformed AR(1) time-series to allow for temporal dependence.

1.2 A spatial model for rare binary events

Spatial methods for binary data commonly implement a Gaussian process to account for spatial dependence. However, for rare data, the latent variable is typically extreme, so Gaussian processes

may not adequately capture the dependence between observations. As a solution to this problem, in Chapter 3 we extend the GEV link [WD10] function to allow for a max-stable dependence structure based on the low-rank representation of a max-stable process given by Reich & Shaby [RS12]. For very rare data, the max-stable dependence structure provides some gain over probit and logistic methods with latent Gaussian processes.

1.3 Empirical basis functions

Large datasets provide a rich source of data, but sometimes they can lead to onerous computing. This is especially true for max-stable methods which can be computationally burdensome for as few as 20 simultaneously observed variables. Low-rank methods provide an attractive solution to dealing with this problem. One such low-rank max-stable model is given by Reich & Shaby [RS12]. In this model, a set of spatial knots are placed throughout the domain of interest, and a positive stable (PS) random effect is then associated with each knot. Chapter 4 presents the development of empirical basis functions to be used in place of the spatial knots. As mentioned previously, these basis functions play a similar role to principal components, but they are positive and thus not orthogonal. The basis functions can be plotted to reveal important spatial trends, and they can also be used for Bayesian inference on the marginal parameters, modeling spatial dependence, and testing for covariate effects.

CHAPTER

2

A SPACE-TIME SKEW- T MODEL FOR THRESHOLD EXCEEDANCES

2.1 Introduction

Epidemiological studies have linked air quality to public health concerns regarding morbidity and mortality [Sam00]. As a result, the Environmental Protection Agency (EPA) has developed a set of standards to help reduce air pollution thereby improving air quality. Our study is motivated by an air pollution application where the focus is not on the average behavior, but on the behavior over a high level determined by government regulation. More specifically, we consider the case of compliance for ozone. A site is said to be in compliance if the fourth highest daily maximum 8-hour concentration averaged over three years does not exceed 75 parts per billion (ppb). Figure 2.1 shows the ozone levels from July 10, 2005, at 1089 stations across the United States. We see a large area above the compliance level in the midwest covering Ohio, Indiana, Illinois, and parts of the surrounding states. We analyze these data with the goals of spatial prediction for unmonitored locations and to map the probability of extreme events.

A spatial model for threshold exceedances warrants special consideration and standard spatial methods are likely to perform poorly. First, because we are interested only in high values, we want

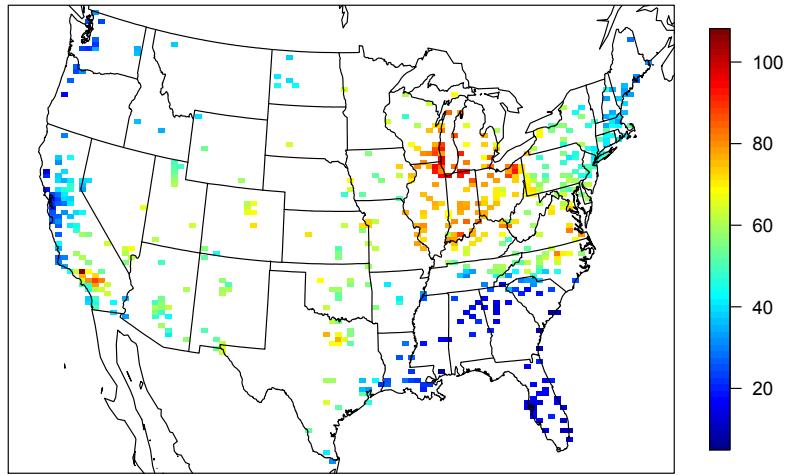


Figure 2.1 Ozone values (ppb) on July 10, 2005

to “let the tail speak for itself”. That is, if we fit a model to the entire data set, low-to-moderate values would influence the fit of the overall model. As there are more of these values, they can unduly influence the distribution at the higher levels about which we are interested. Our inference method will only use data which exceed a pre-selected threshold and will censor data below the threshold, thereby tailoring the fit to the levels of interest. Second, likelihood-based spatial modeling typically assumes a Gaussian process, which is appropriate when mean behavior is of interest. However, the Gaussian distribution is light-tailed and symmetric, and therefore may be inappropriate for modeling data which does not share this tail behavior. Third, we aim to capture the dependence structure when ozone is at high levels, and dependence at these levels may not be well-represented by covariances which focus again on mean behavior. Asymptotic dependence/independence (see Section 2.2.2) are notions which describe how two random variables’ probability of simultaneous exceedance of an extremely high level. The Gaussian distribution always exhibits asymptotic independence, except in the case of perfect dependence, thus is an inappropriate model for data which exhibits asymptotic dependence. To allow for more flexibility in the marginal tail and to allow for asymptotic dependence, the skew-*t* distribution forms the basis for our model.

Our approach differs from threshold modeling approaches based on extreme value distributions.

There has been extensive work on threshold modeling in the field of extreme value statistics where extreme events are naturally defined in terms of exceedances over a high threshold. Davison & Smith [DS90] considered modeling threshold exceedances of univariate time series by the generalized Pareto distribution. Threshold based inference for multivariate extreme value distributions was considered by Ledford & Tawn [LT96] who introduced a censored approach that provides a way to deal with different types of exceedances of a threshold. These models were extended to spatial models for threshold exceedances by Wadsworth & Tawn [WT12] and Thibaud et al. [Thi13] who fit various models to spatial extremes using a censored pairwise likelihood [Pad10] based on the approach of Ledford & Tawn [LT96]. Huser & Davison [HD14] further extended this to space-time modeling. Wadsworth & Tawn [WT14], Engelke et al. [Eng15], and Thibaud & Opitz [TO15] introduced more efficient inference for threshold exceedances of extremal spatial processes with full likelihood methods. The previous approaches to threshold modeling are motivated by extreme value theory and assume the threshold is high enough that extremal models are valid for the data and for extrapolation beyond the range of observed values. Moreover, these approaches are computationally intensive and limited to rather small datasets. Our application with ozone data does not fit into this framework because we do not focus on exceedances of a very high level, and we have observations at 1,089 ozone monitoring locations.

We propose a new spatiotemporal threshold exceedance model based on the skew- t process [Pad11]. We use a skew- t distribution because of its flexibility to model asymmetry and heavy-tailed data with the aim of modeling exceedances of a high fixed level at an unobserved location. Our model allows for inference and predictions using the full likelihood with computing on the order of Gaussian models. This allows us to use Bayesian methods, which we use to fit the model, handle censored data below the threshold, and make predictions at unobserved locations. The multivariate skew normal distribution was introduced by Azzalini & Dalla Valle [AD96], and this was extended to the multivariate skew- t by Branco & Dey [BD01]. These skew-elliptical distributions have been used in the spatial setting [Gen04; KM04]. Zhang & El-Shaarawi [ZES10] propose the skew-Gaussian process as a class of stationary processes that have skewed marginal distributions. Padoan [Pad11] examined the usage of skew-Gaussian and skew- t distributions for multivariate extremes. In a spatial setting, the multivariate skew- t distribution demonstrates asymptotic dependence between observations at all sites regardless of the distance between the sites. In order to address this concern, we introduce a random spatial partition similar to the method used by Kim et al. [Kim05] for non-stationary Gaussian data.

The paper is organized as follows. Section 2.2 is a brief review of the spatial skew- t process. In Section 2.3, we build upon the traditional skew- t process by incorporating censoring, partitioning, and extending the model to space-time data. The computing is described in Section 2.4. In Section 2.5,

we present a simulation study that examines the predictive capabilities of this model compared to Gaussian and max-stable methods. We compare our method to Gaussian and max-stable methods with a data analysis of ozone measurements throughout the US in Section 2.6.

2.2 Spatial skew processes

The skew-elliptical family of distributions provides models that are mathematically tractable while introducing a slant parameter to account for asymmetric data. A brief review of the additive process [AC14, p. 129] by which a skew- t process is created is given here.

2.2.1 Skew- t process

Let $Y(\mathbf{s})$ be a spatial process defined for spatial location \mathbf{s} in a spatial domain of interest $\mathcal{D} \in \mathbb{R}^2$. The spatial skew- t process can be written as

$$Y(\mathbf{s}) = \mathbf{X}(\mathbf{s})^\top \boldsymbol{\beta} + \lambda \sigma |z| + \sigma \nu(\mathbf{s}) \quad (2.1)$$

where $\mathbf{X}(\mathbf{s})$ is the observed covariate vector at site \mathbf{s} , $\boldsymbol{\beta}$ is the p -vector of regression parameters, $\lambda \in \mathbb{R}$ is a parameter controlling skewness, $z \sim N(0, 1)$, $\sigma^2 \sim \text{IG}(a/2, b/2)$ is random scale parameter, IG is the distribution function of an inverse gamma random variable, a is the degrees of freedom, b controls the precision of the process, and $\nu(\mathbf{s})$ is a standard Gaussian process with positive definite correlation function $\text{Cor}[Y(\mathbf{s}_1), Y(\mathbf{s}_2)] = \rho(\mathbf{s}_1, \mathbf{s}_2)$. Although any positive definite correlation function could be used, we choose to use the stationary isotropic Matérn correlation with

$$\rho(h) = (1 - \gamma)I(h=0) + \gamma \frac{1}{\Gamma(\nu)2^{\nu-1}} \left(\sqrt{2\nu} \frac{h}{\varphi} \right)^\nu K_\nu \left(\sqrt{2\nu} \frac{h}{\varphi} \right) \quad (2.2)$$

where $I(\cdot)$ is an indicator function, $\varphi > 0$ is the spatial range, $\nu > 0$ is the smoothness, $\gamma \in [0, 1]$ is the proportion of variance accounted for by the spatial variation, K_ν is a modified Bessel function of the second kind, and $h = \|\mathbf{s}_1 - \mathbf{s}_2\|$ is the Euclidean distance between sites \mathbf{s}_1 and \mathbf{s}_2 .

For a finite collection of locations $\mathbf{s}_1, \dots, \mathbf{s}_n$, we denote by $\mathbf{Y} = [Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n)]^\top$ the vector of observations, and the covariate matrix $\mathbf{X}_{n \times p} = [\mathbf{X}(\mathbf{s}_1), \dots, \mathbf{X}(\mathbf{s}_n)]^\top$. After marginalizing over both z and σ , using the notation from Azzalini & Capitanio [AC14, p. 176],

$$\mathbf{Y} \sim \text{ST}_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, a), \quad (2.3)$$

that is, \mathbf{Y} follows an n -dimensional skew- t distribution with location $\mathbf{X}\boldsymbol{\beta} \in \mathbb{R}^n$; covariance matrix

$\Omega_{n \times n} = \boldsymbol{\omega} \left[\frac{1}{1+\lambda^2} (\Sigma + \lambda^2 \mathbf{1}\mathbf{1}^\top) \right] \boldsymbol{\omega}$, $\Sigma_{n \times n}$ is the positive definite correlation matrix which is obtained from $\rho(h)$, and $\boldsymbol{\omega}_{n \times n} = \text{diag} \left(\sqrt{\frac{b}{a(1+\lambda^2)}}, \dots, \sqrt{\frac{b}{a(1+\lambda^2)}} \right)$; slant parameters $\boldsymbol{\alpha} \in \mathbb{R}^n = \lambda(1 + \lambda^2)^{1/2}(1 + \lambda^2 \mathbf{1}^\top \Sigma^{-1} \mathbf{1})^{-1/2} \Sigma^{-1} \mathbf{1}$, and degrees of freedom a . Furthermore, the marginal distributions at each location also follow a univariate skew- t distribution [AC14]. This process is desirable because it is heavy tailed with tail index α , and the shape of the distribution is controlled by the skewness parameter. For a comparison with other parameterizations, see Appendix A.5.

2.2.2 Extremal dependence

Our interest lies in spatial dependence in the tail of the skew- t process. One measure of extremal dependence is the χ statistic [Col99]. For a stationary and isotropic spatial process, the χ statistic for two locations separated by distance h is

$$\chi(h) = \lim_{c \rightarrow c^*} \Pr[Y(\mathbf{s} + h) > c | Y(\mathbf{s}) > c] \quad (2.4)$$

where c^* is the upper limit of the support of Y ; for the skew- t distribution $c^* = \infty$. If $\chi(h) = 0$, then observations are asymptotically independent at distance h . For Gaussian processes, $\chi(h) = 0$ regardless of the distance h , so they are not suitable for modeling asymptotically dependent extremes. Unlike the Gaussian process, the skew- t process is asymptotically dependent (the explicit expression for $\chi(h)$ is given in Appendix A.4). However, one problem with the spatial skew- t process is that $\lim_{h \rightarrow \infty} \chi(h) > 0$. This occurs because all observations, both near and far, share the same z and σ terms. Therefore, this long-range dependence feature of the skew- t process is not desirable for spatial analysis of large geographic regions where we expect only local spatial dependence. We propose a solution to this in Section 2.3.2.

2.3 Extending the model

In this section, we propose extensions to the skew- t process to model spatial extremes over a large geographic region by introducing censoring to focus on tail behavior and a random partition to remove long-range asymptotic dependence. For notational convenience, we introduce the model for a single replication, and then extend this model to the spatiotemporal setting in Section 2.3.3.

2.3.1 Censoring to focus on the tail

We do not want the low-to-moderate values to influence the fit of the model. We propose the use of a censored approach to fit threshold exceedances only. More specifically, we assume our

skew- t model $Y(\mathbf{s})$ is valid at each location \mathbf{s} above a threshold T , and censor the values below T for which we don't assume the model to be valid. We define our partially censored observations as $\tilde{\mathbf{Y}} = [\tilde{Y}(\mathbf{s}_1), \dots, \tilde{Y}(\mathbf{s}_n)]^\top$ where $\tilde{Y}(\mathbf{s}) = \max\{Y(\mathbf{s}), T\}$, and fit the skew- t process to these $\tilde{\mathbf{Y}}$. In our Bayesian framework, inference can be easily performed by imputing censored observations below T (see Section 2.4.1).

As our goal is to model exceedances above a high level L , we should select a value for $T \leq L$. For example, in predicting ozone exceedances, we might set $T = 50$ ppb in order to predict exceedances of $L = 75$ ppb. Selecting T too small may lead to bias in estimating the tail parameters; selecting T too large increases variance. We impute the censored values as a step in the algorithm used to fit the model described in Section 2.4.1, and use cross-validation to select T .

2.3.2 Partitioning to remove long-range asymptotic dependence

For a large spatial domain, it may not be reasonable to assume sites that are far apart demonstrate asymptotic dependence. As discussed in Section 2.2, the source of long-range dependence is the shared z and σ . Therefore, to alleviate this dependence, we allow z and σ to vary by site using a partitioning approach. The model becomes

$$Y(\mathbf{s}) = \mathbf{X}(\mathbf{s})^\top \boldsymbol{\beta} + \lambda \sigma(\mathbf{s})|z(\mathbf{s})| + \sigma(\mathbf{s})\nu(\mathbf{s}). \quad (2.5)$$

To model spatial variation, consider a set of spatial knots $\mathbf{w}_1, \dots, \mathbf{w}_K$ from a homogeneous Poisson process over spatial domain $\mathcal{D} \in \mathbb{R}^2$. The knots define a random partition of \mathcal{D} by subregions P_1, \dots, P_K defined as

$$P_k = \{\mathbf{s} : k = \arg \min_\ell \|\mathbf{s} - \mathbf{w}_\ell\|\}. \quad (2.6)$$

In other words, P_k is composed of all sites for which the closest knot is \mathbf{w}_k . For all $\mathbf{s} \in P_k$, with $k = 1, 2, \dots, K$, the functions $z(\mathbf{s})$ and $\sigma(\mathbf{s})$ are equal to the constants z_k and σ_k respectively, and the z_k and σ_k^2 are distributed as $z_k \stackrel{\text{iid}}{\sim} N(0, 1)$ and $\sigma_k^2 \stackrel{\text{iid}}{\sim} \text{IG}(a/2, b/2)$. So, within each partition, $Y(\mathbf{s})$ follows the spatial skew- t process defined in Section 2.2. Across partitions, the $Y(\mathbf{s})$ remain dependent via the correlation function for $\nu(\mathbf{s})$ because $\nu(\mathbf{s})$ spans all partitions. However, the bivariate distribution for sites in different partitions is neither Gaussian nor skew- t and does not have asymptotic dependence.

The partitioning model removes long-range dependence. Conditional on knots $\mathbf{w}_1, \dots, \mathbf{w}_K$, the

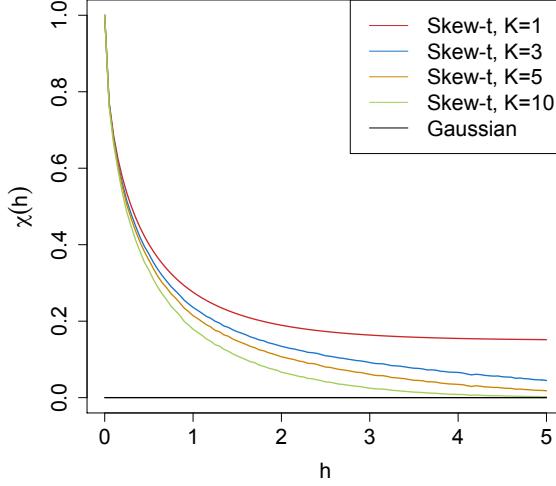


Figure 2.2 Extremal dependence measure $\chi(h)$, as a function of distance, h , for $K = 1, 3, 5$, and 10 knots.

χ statistic for two sites \mathbf{s}_1 and \mathbf{s}_2 in partitions k_1 and k_2 respectively is

$$\chi(h) = I(k_1 = k_2)\chi_{\text{skew-}t}(h) \quad (2.7)$$

where $\chi_{\text{skew-}t}(h)$ is the χ statistic for a skew- t process given in equation (A.8) of Appendix A.4, and $h = \|\mathbf{s}_1 - \mathbf{s}_2\|$. Marginally, over the knots, $\chi(h) = \pi(h)\chi_{\text{skew-}t}(h)$, where $\pi(h) = \Pr(k_1 = k_2)$ is the probability that two sites separated by distance h are in the same partition. In Appendix A.3, we show that assuming the knots follow a homogeneous Poisson process, $\lim_{h \rightarrow \infty} \pi(h) = 0$, and thus long-range dependence is removed. In practice we fix K at a finite value and use a uniform distribution for the knots $\mathbf{w}_1, \dots, \mathbf{w}_K$. In Figure 2.2, we estimate $\chi(h)$ for $K = 1, 3, 5, 10$ partitions for a skew- t distribution with $\alpha = 10$, and 3 degrees of freedom.

2.3.3 Extension to space-time data

When using daily measurements, the assumption of temporal independence is often inappropriate. In this section, we extend (2.5) to the spatiotemporal setting. There are several places where temporal dependence could be incorporated in the model, including the Gaussian process $v_t(\mathbf{s})$. However, we choose to allow for temporal dependence in the \mathbf{w} , z , and σ terms because these terms dictate the tail behavior of the process which is our primary focus (see Appendix A.6 for a discussion of the

induced temporal asymptotic dependence). Let

$$Y_t(\mathbf{s}) = \mathbf{X}_t(\mathbf{s})^\top \boldsymbol{\beta} + \lambda \sigma_t(\mathbf{s}) |z_t(\mathbf{s})| + \sigma_t(\mathbf{s}) v_t(\mathbf{s}), \quad (2.8)$$

where $t \in \{1, \dots, n_t\}$ denotes the day of each observation. Let $\mathbf{w}_{tk} = (w_{tk1}, w_{tk2})$ be a spatial knot on day t , and let $\mathbf{w}_{t1}, \dots, \mathbf{w}_{tK}$ be the collection of spatial knots on day t . As in Section 2.3.2, these knots define a daily partition P_{t1}, \dots, P_{tK} , and for $\mathbf{s} \in P_{tk}$,

$$z_t(\mathbf{s}) = z_{tk} \quad \text{and} \quad \sigma_t(\mathbf{s}) = \sigma_{tk}. \quad (2.9)$$

We allow the partition structure to vary from day to day in order to account for sharp spikes in a response that may not be present every day (e.g. the impact of a forest fire on ozone levels).

We use an AR(1) time series model for \mathbf{w}_{tk} , z_{tk} , and σ_{tk} . The time series model must be specified after a transformation to preserve the skew- t process at each time point. For each time-varying parameter, we transform the parameter to obtain a standard normal marginal distribution, place a Gaussian prior with autocorrelation on the transformed parameter, and then transform back to the appropriate marginal distribution for the skew- t process. We first transform the spatial knots from \mathcal{D} to \mathbb{R}^2 as follows. Let

$$w_{tki}^* = \Phi^{-1} \left[\frac{w_{tki} - \min(\mathbf{s}_i)}{\max(\mathbf{s}_i) - \min(\mathbf{s}_i)} \right], \quad i = 1, 2 \quad (2.10)$$

where Φ is a univariate standard normal density function and $\mathbf{s}_i = [s_{1i}, \dots, s_{ni}]$. Then the transformed knots $\mathbf{w}_{tk}^* \in \mathbb{R}^2$. We use a transformation of a Gaussian random variable on $z_t(\mathbf{s})$ to ensure that the marginal distributions of $z_t(\mathbf{s})$ are half-normal. Let

$$z_t^*(\mathbf{s}) = \Phi^{-1} \{ \text{HN}[z_t(\mathbf{s})] \} \quad (2.11)$$

where HN is the distribution function of a half-normal random variable. We also use a transformation of a Gaussian random variable on $\sigma_t^2(\mathbf{s})$ to ensure that the marginal distributions of $\sigma_t^2(\mathbf{s})$ are inverse gamma. Let

$$\sigma_t^{2*}(\mathbf{s}) = \Phi^{-1} \{ \text{IG}[\sigma_t^2(\mathbf{s})] \} \quad (2.12)$$

where IG is defined as before. The AR(1) process for each tail parameter is $\mathbf{w}_{1k}^* \sim N_2(\mathbf{0}, \mathbf{I}_2)$ where

$\mathbf{I}_2 = \text{diag}(1, 1)$, $z_{1k}^* \sim N(0, 1)$, $\sigma_{1k}^{2*} \sim N(0, 1)$, and for $t > 1$ the time series is modeled as

$$\mathbf{w}_{tk}^* | \mathbf{w}_{t-1,k}^* \sim N_2[\phi_w \mathbf{w}_{t-1,k}^*, (1 - \phi_w^2) \mathbf{I}_2] \quad (2.13)$$

$$z_{tk}^* | z_{t-1,k}^* \sim N[\phi_z z_{t-1,k}^*, (1 - \phi_z^2)] \quad (2.14)$$

$$\sigma_{tk}^{2*} | \sigma_{t-1,k}^{2*} \sim N[\phi_\sigma \sigma_{t-1,k}^{2*}, (1 - \phi_\sigma^2)] \quad (2.15)$$

where $|\phi_w|, |\phi_z|, |\phi_\sigma| < 1$. These are stationary time series models with marginal distributions $\mathbf{w}_k^* \sim N_2(\mathbf{0}, \mathbf{I}_2)$, $z_k^* \sim N(0, 1)$, and $\sigma_k^{2*} \sim N(0, 1)$. After transformation back to the original space, $\mathbf{w}_{tk} \sim \text{Unif}(\mathcal{D})$, $z_{tk} \sim \text{HN}(0, 1)$, and $\sigma_{tk}^2 \sim \text{IG}(a/2, b/2)$. We then create the partition for day t using $\mathbf{w}_{t1}, \dots, \mathbf{w}_{tK}$. For each day, the model is identical to the spatial-only model in (2.5) by construction.

2.4 Hierarchical model

We define a Bayesian hierarchical model based on the skew- t process and use a Markov chain Monte Carlo (MCMC) algorithm to fit the data. We model our data as partially censored observations (see Section 2.3.1) from the skew- t model defined in Section 2.3. In the first step of the MCMC algorithm, we impute censored values of \tilde{Y} such that the estimation of model parameters can be based on the completed \mathbf{Y} . Conditioned on $z_{tk}(\mathbf{s})$, $\sigma_{tk}^2(\mathbf{s})$, and P_{tk} , joint distribution of \mathbf{Y} is multivariate Gaussian. We do not fix the partitions; instead, the locations of the K knots are treated as unknown and random. One approach would be to allow K to be unknown and follow a Poisson process prior, but this would lead to onerous computing. Therefore, we elect to treat K as a tuning parameter for the MCMC by fixing it at different values and assessing its impact on prediction as described in Section 2.5.2. Then the hierarchical model is given as

$$\begin{aligned} Y_t(\mathbf{s}) | z_t(\mathbf{s}), \sigma_t^2(\mathbf{s}), P_{tk}, \Theta &= \mathbf{X}_t(\mathbf{s})^\top \beta + \lambda \sigma_t(\mathbf{s}) | z_t(\mathbf{s}) | + \sigma_t(\mathbf{s}) v_t(\mathbf{s}) \\ z_t(\mathbf{s}) &= z_{tk} \text{ if } \mathbf{s} \in P_{tk} \\ \sigma_t^2(\mathbf{s}) &= \sigma_{tk}^2 \text{ if } \mathbf{s} \in P_{tk} \\ \lambda &\sim N(0, \sigma_\lambda^2) \\ v_t(\mathbf{s}) | \Theta &\sim \text{Matérn}(\mathbf{0}, \Sigma) \end{aligned} \quad (2.16)$$

where $\Theta = \{\varphi, \nu, \gamma, \lambda, \boldsymbol{\beta}\}$, Σ is a Matérn covariance matrix as described in Section 2.2.1, and the priors on \mathbf{w}_{tk}^* , z_{tk}^* , and σ_{tk}^{2*} are given in (2.13) – (2.15).

2.4.1 Computation

We use MCMC methods to estimate the posterior distribution of the model parameters. At each MCMC iteration, we first impute values below the threshold conditional on observations above the threshold. After conditioning on λ , $z_t(\mathbf{s})$ and non-censored observations, $Y_t(\mathbf{s})$ has truncated normal full conditionals $Y_t(\mathbf{s}) \sim N_{(-\infty, T)}(\mathbf{X}_t^\top(\mathbf{s})\beta + \lambda|z_t(\mathbf{s})|, \Sigma)$.

We update model parameters Θ using Gibbs sampling with Metropolis-Hastings steps when needed. In our case, we also wish to be able to make predictions at sites where we do not have data. We can easily implement Bayesian Kriging as a part of the algorithm to generate a predictive distribution for $Y_t(\mathbf{s}^*)$ at prediction location \mathbf{s}^* . This step is similar to the imputation for censored observations except that the full conditionals are no longer truncated at T . See Appendices A.1 and A.2 for details regarding the MCMC algorithm.

2.5 Simulation study

In this section, we present the results from a simulation study to investigate how the number of partitions and the amount of thresholding impact the accuracy of predictions made by the model and to compare with Gaussian and max-stable methods.

2.5.1 Design

For all simulation designs, we generated data from model (2.5) in Section 2.3.2 using $n_s = 144$ sites and $n_t = 50$ independent days. The sites were generated uniformly on the square $[0, 10] \times [0, 10]$. We generated data from five different simulation designs:

1. Gaussian, $K = 1$ knot
2. Skew- t , $K = 1$ knots
3. Skew- t , $K = 5$ knots
4. Reich & Shaby [RS12] max-stable process
5. Brown-Resnick max-stable process [Kab09]

In the first three designs, the realizations from $v_t(\mathbf{s})$ were generated using a Matérn covariance with smoothness parameter $\nu = 0.5$, spatial range $\varphi = 1$ and $\gamma = 0.9$. In the first design, $\sigma^2 = 2$ was used for all days which results in a Gaussian distribution. For designs 2 and 3, $\sigma_{tk}^2 \stackrel{\text{iid}}{\sim} \text{IG}(3, 8)$ to give a t distribution with 6 degrees of freedom. For design 1, we set $\lambda = 0$. For designs 2 and 3, $\lambda = 3$ was

used as to simulate moderate skewness, and the z_t were generated as described in Section 2.3.2. In designs 1 – 3, the mean $\mathbf{X}^\top \boldsymbol{\beta} = 10$ was assumed to be constant across space. In the fourth design, we generated from the max-stable model of Reich & Shaby [RS12]. The marginal distributions follow a generalized extreme value distribution with location parameter 1, scale parameter 1, and shape parameter 0.2. Spatial dependence in the form an asymmetric logistic dependence function is induced by random effects for kernel basis functions associated with 144 spatial knots defined on a square grid on $[1, 9] \times [1, 9]$. We set the dependence parameter (α in Reich and Shaby, 2012) to 0.5 which represents moderate spatial dependence. For the final design, we generated data from a Brown-Resnick max-stable process using `rmaxtab` in the `SpatialExtremes` package of R [Rib15]. For this design we fixed unit Fréchet margins, and we used a range of 1 and smoothness 0.5.

$M = 50$ data sets were generated for each design. For each data set we fit the data using six models

1. Gaussian marginal, $K = 1$ knots
2. Skew- t marginal, $K = 1$ knots, $T = -\infty$
3. Symmetric- t marginal, $K = 1$ knots, $T = q(0.80)$
4. Skew- t marginal, $K = 5$ knots, $T = -\infty$
5. Symmetric- t marginal, $K = 5$ knots, $T = q(0.80)$
6. Reich & Shaby [RS12] max-table model thresholded at $T = q(0.80)$

where $q(0.80)$ is the 80th sample quantile of the data. All methods were fit using a fully-Bayesian approach that simultaneously estimates marginal and spatial dependence parameters. The design matrix \mathbf{X} includes an intercept with a first-order spatial trend with priors of $\beta_{\text{int}}, \beta_{\text{lat}}, \beta_{\text{long}} \stackrel{\text{iid}}{\sim} N(0, 10)$ although only the intercept is used in the data generation. The spatial covariance parameters have priors $\log(\nu) \sim N(-1.2, 1)$, $\gamma \sim \text{Unif}(0, 1)$, $\varphi \sim \text{Unif}(0, 15)$. The skewness parameter has prior $\lambda \sim N(0, 20)$. The residual variance terms have priors $\sigma_t^2(\mathbf{s}) \sim \text{IG}(a/2, b/2)$, where a has a discrete uniform prior on a mesh from 0.2 to 20 with spacing of 0.1 and b has a $\text{Gamma}(0.1, 0.1)$ prior. As described in Section 2.2.1, these priors are meant to be fairly uninformative to allow the data to dictate both the degrees of freedom a and precision b of the process. The knots have priors $\mathbf{w} \sim \text{Unif}(\mathcal{D})$. We tried also fitting the skew- t marginals for the thresholded models, but it is very challenging for the MCMC to properly identify the skewness parameter with a censored left tail. Each chain of the MCMC ran for 20,000 iterations with a burn-in period of 10,000 iterations. Although the goal of this simulation study is not to assess parameter estimation, for design (2) and method (2)

where data are generated and fit with a skew- t distribution, the samples converges for λ and a , and the empirical coverage of posterior 95% intervals is near the nominal level (96% coverage for λ and 90% for a). It should be noted that in the models with multiple partitions (i.e. models 4 and 5) it is hard to assess the convergence of \mathbf{w} , $z(\mathbf{s})$, and $\sigma^2(\mathbf{s})$ because of partition label switching throughout the MCMC; however, we are not interested in these parameters but rather spatial predictions and tail probabilities which converge well. Finally, we did not fit a Brown-Resnick model because we cannot use a Bayesian approach with so many sites.

2.5.2 Cross validation

Models were compared using cross validation, with 100 sites used as training sites to fit the models, and 44 sites withheld for testing the predictions. Because one of the primary goals of this model is to predict exceedances over a high level, we use Brier scores to compare the models [GR07]. The Brier score for predicting exceedance of a level L is given by $[e(L) - P(L)]^2$ where $e(L) = I[y > L]$ is an indicator function indicating that a test set value, y , has exceeded the level, L , and $P(L)$ is the predicted probability of exceeding L . We average the Brier scores over all test sites and days. For the Brier score, a lower score indicates a better fit.

2.5.3 Results

We compared the Brier scores for exceeding four different high levels for each dataset. The levels used for the Brier scores are extreme quantiles from the simulated data for $L = q(0.90)$, $q(0.95)$, $q(0.98)$, $q(0.99)$. Figure 2.3 gives the Brier score relative to the Brier score for the Gaussian method calculated as

$$\text{BS}_{\text{rel}} = \frac{\text{BS}_{\text{method}}}{\text{BS}_{\text{Gaussian}}}. \quad (2.17)$$

We analyzed the results for the simulation study using a Friedman [Hol14] test at $\alpha = 0.05$ to see if at least one method had a significantly different Brier score. For Friedman tests that came back with a significant p-value, we conducted a Wilcoxon-Nemenyi-McDonald-Thompson [Hol14] test to see which of the methods had different results. The full results for the Wilcoxon-Nemenyi-McDonald-Thompson tests are given in Appendix A.8.

In general, we find that when the method to fit the data matches the data generation scheme, there is some improvement over other methods. The results show that when the data are generated from a Gaussian process, our method performs comparably to a Gaussian approach. In general, when the underlying process is not Gaussian, our method results in an improvement over both

the max-stable and Gaussian methods. We also see that the non-thresholded methods tend to outperform the thresholded methods, but this is not surprising given that in most cases, the data are generated directly from the model. Finally, in the case where the data are generated from a Brown-Resnick process, we find that our method is competitive with using a max-stable model. In summary, our method provides great flexibility for data that demonstrate some level of asymmetry and heavy tails, while still performing comparably to Gaussian methods when the data are symmetric and have light tails.

2.6 Data analysis

We consider daily observations of maximum 8-hour ozone measurements for the 31 days of July 2005 at 1,089 Air Quality System (AQS) monitoring sites in the United States as the response (see Figure 2.1). For each site, we also have covariate information containing the estimated ozone from the Community Multi-scale Air Quality (CMAQ) modeling system. Initially, we fit a linear regression with $\mathbf{X}_t(\mathbf{s}) = [1, \text{CMAQ}_t(\mathbf{s})]^\top$. Figure 2.4 shows a Q-Q plot of the residuals compared to a skew- t distribution with $a = 10$ and $\lambda = 1$, suggesting the data are heavy tailed.

Standard exploratory data analysis techniques for extremal dependence are very challenging with only 31 days worth of data because it is difficult to estimate extreme quantiles at each site to obtain empirical estimates of χ . Despite the fact that there is only one month of data, we can get some sense of extremal dependence between sites by looking at joint occurrences of high sample quantiles. For example, Figure 2.5 suggests there is more agreement between sites that are close to one another than sites that are far from one another.

2.6.1 Model comparisons

We fit the model using Gaussian and skew- t marginal distributions with $K = 1, 5, 6, 7, 8, 9, 10, 15$ partitions. We censored $Y(\mathbf{s})$ at $T = 0$, $T = 50$ (0.42 sample quantile), and $T = 75$ (0.92 sample quantile) ppb in order to compare results from no, moderate, and high censoring. The upper threshold of 75 ppb was used because the current air quality standard is based on exceedance of 75 ppb. As with the simulation study, for models with a threshold of $T = 75$, we used a symmetric- t marginal distribution. We also compared models with no time series to models that included the time series. Finally, as a comparison to max-stable methods, we fit the model using the hierarchical max-stable model of Reich & Shaby [RS12] with the data thresholded at $T = 75$. All methods assumed $\mathbf{X}_t(\mathbf{s}) = [1, \text{CMAQ}_t(\mathbf{s})]^\top$. To ensure that the max-stable method ran in a reasonable amount of time, we used a stratified sub-sample of 800 sites. We conducted two-fold cross validation using 400

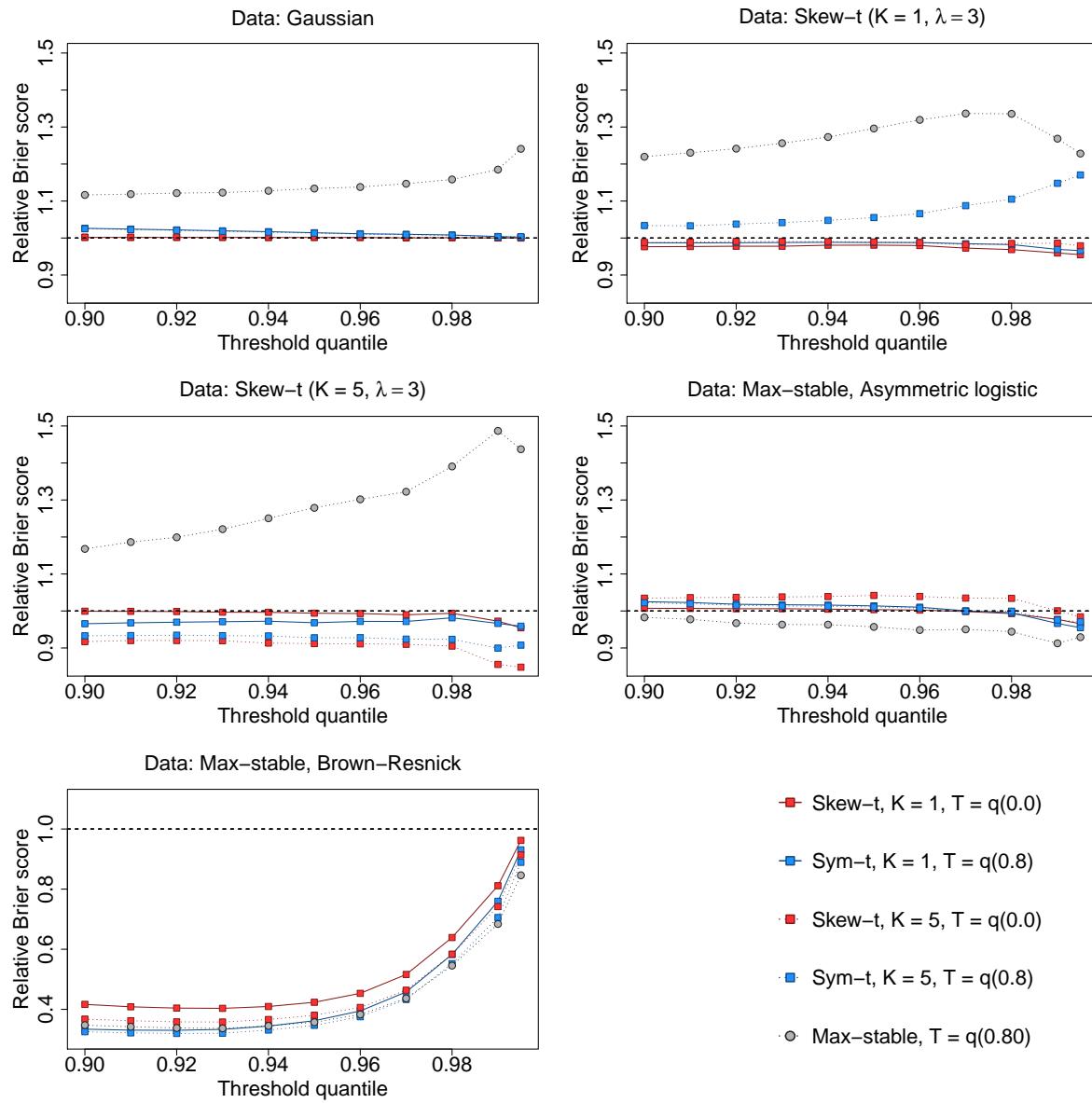


Figure 2.3 Brier scores relative to the Gaussian method for simulation study results. A ratio lower than 1 indicates that the method outperforms the Gaussian method.

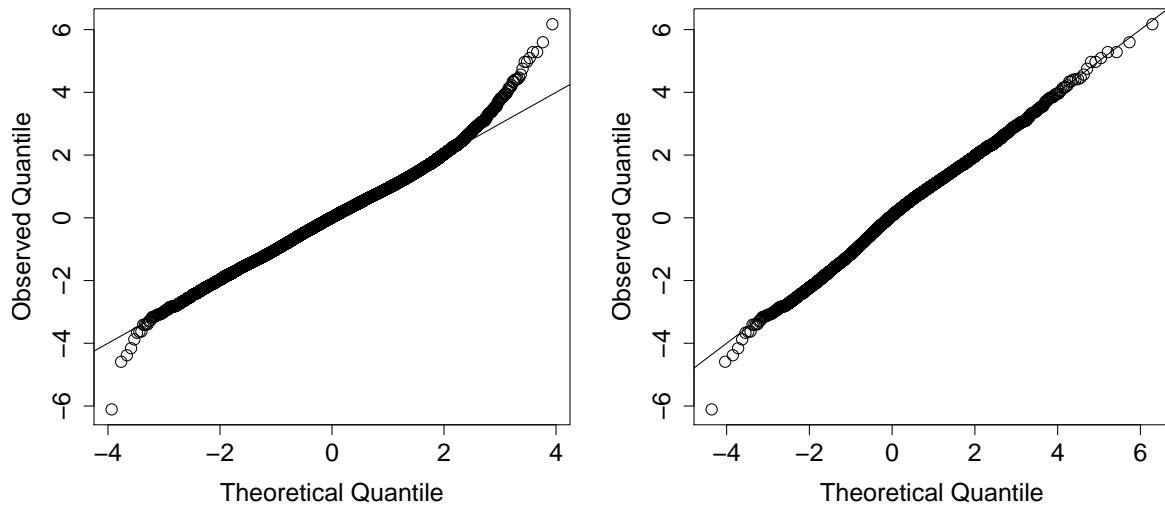


Figure 2.4 Gaussian Q-Q plot (left) and skew-*t* with $\alpha = 10$ and $\lambda = 1$ Q-Q plot (right) of the residuals.

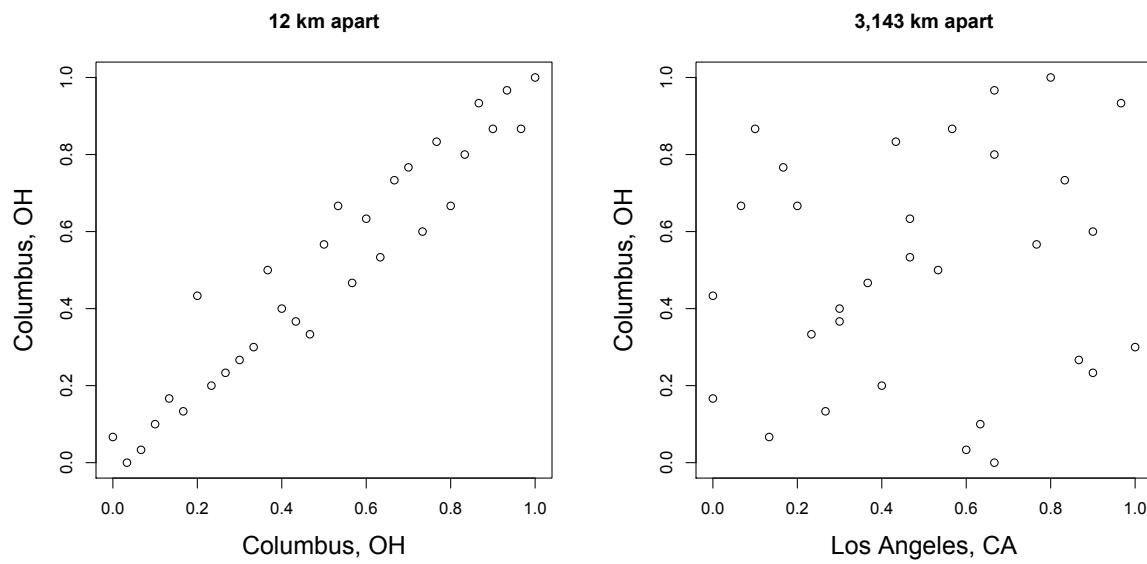


Figure 2.5 Daily quantiles for two monitoring locations near Columbus, OH (left) and daily quantiles for a monitoring location in Los Angeles, CA and Columbus, OH (right)

training sites and 400 validation sites as described in Section 2.5.2

Each chain of the MCMC ran for 30,000 iterations with a burn-in period of 25,000 iterations. We used the same priors for the spatial covariance parameters, skewness parameter, and knots as in the simulation study. The prior for the residual variance terms was $\sigma_t^2(\mathbf{s}) \sim \text{IG}(a/2, b/2)$ where a was the same as the simulation study, but b had a $\text{Gamma}(1, 1)$ prior. Parameters appeared to converge properly; however, as before, for models with multiple partitions it was hard to assess the convergence of \mathbf{w} , $z(\mathbf{s})$, and $\sigma^2(\mathbf{s})$ because of partition label switching throughout the MCMC. For each model, we averaged Brier scores over all sites and days to obtain a single Brier score for each dataset. At a particular level, the model that fit the best was the one with the lowest score. We then computed the relative (to Gaussian) Brier scores (see Section 2.5.3) to compare each model.

2.6.2 Results

The results suggest that the skew- t , thresholded, partitioned, and time series models all give an improvement in predictions over the Gaussian model, whereas the max-stable method results in relative Brier scores between 1.13 and 1.18 indicating poorer performance than the Gaussian model. The plots in Figure 2.6 show the relative Brier scores for time-series and non-time-series models, using $K = 1, 7$, and 15 knots at thresholds $T = 0, 50$, and 75 ppb. Most of the models perform similarly across all the Brier scores; however, for single-partition models without thresholding, performance tends to diminish in the extreme quantiles. The results also suggest that thresholding improves performance for estimates in the extreme quantiles. Both plots have similar features suggesting that most settings do reasonably well. In particular, for all extreme quantiles, selecting a moderate number of knots (e.g. $K = 5, \dots, 10$) tends to give the best results. Table 2.1 shows the best two models for selected extreme quantiles.

We illustrate the predictive capability of our model in Figure 2.7 by plotting the 99th quantile for South Carolina and Georgia, a subset of the spatial domain, in order to study local features. The four methods used are

1. Gaussian
2. Skew- t , $K = 1$ knot, $T = 0$, no time series
3. Skew- t , $K = 5$ knots, $T = 50$, no time series
4. Symmetric- t , $K = 10$ knots, $T = 75$, time series.

In the bottom two plots, we plot the differences between method 4 and methods 1 and 2. The most noticeable differences between the reference methods and the comparison methods is that the

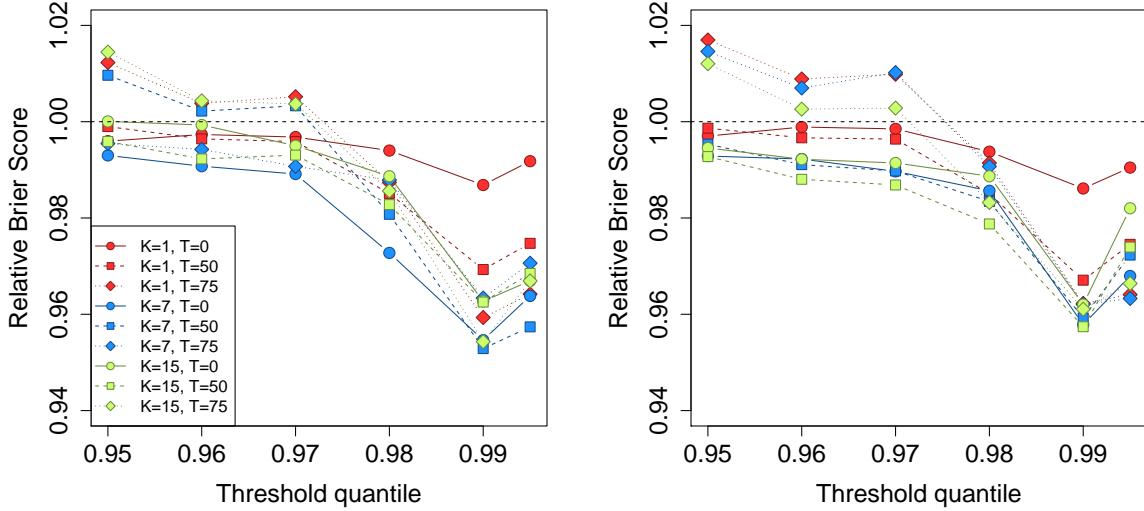


Figure 2.6 Relative Brier scores for time-series models (left) and non-time-series models (right). Relative brier score for the max-stable model is between 1.13 and 1.18

Table 2.1 Top two performing models for predicting ozone exceedance of level L with Relative Brier score

L	1st				2nd				
		No time series	$K = 6$	$T = 50$	BS: 0.992	No time series	$K = 1$	$T = 0$	BS: 0.992
$q(0.90)$		No time series	$K = 6$	$T = 50$	BS: 0.992	No time series	$K = 1$	$T = 0$	BS: 0.992
$q(0.95)$		Time series	$K = 5$	$T = 50$	BS: 0.988	No time series	$K = 6$	$T = 50$	BS: 0.989
$q(0.98)$		Time series	$K = 7$	$T = 50$	BS: 0.973	Time series	$K = 5$	$T = 50$	BS: 0.975
$q(0.99)$		No time series	$K = 8$	$T = 0$	BS: 0.946	Time series	$K = 9$	$T = 75$	BS: 0.947
$q(0.995)$		No time series	$K = 8$	$T = 0$	BS: 0.951	Time series	$K = 9$	$T = 75$	BS: 0.956

comparison methods tend to give higher estimates of the 99th quantile along the I-85 corridor between Charlotte and Atlanta. Among these methods, the fourth method demonstrates the best performance. For a map of Brier scores for the 99th quantile between Gaussian and the fourth method, see Appendix A.7.

2.7 Discussion

In this paper we propose a new threshold exceedance approach for spatiotemporal modeling based on the skew- t process. The proposed model gives flexible tail behavior, demonstrates asymptotic dependence for observations at sites that are near to one another, and has computation on the order of Gaussian models for large space-time datasets. In the simulation study, we demonstrate that this

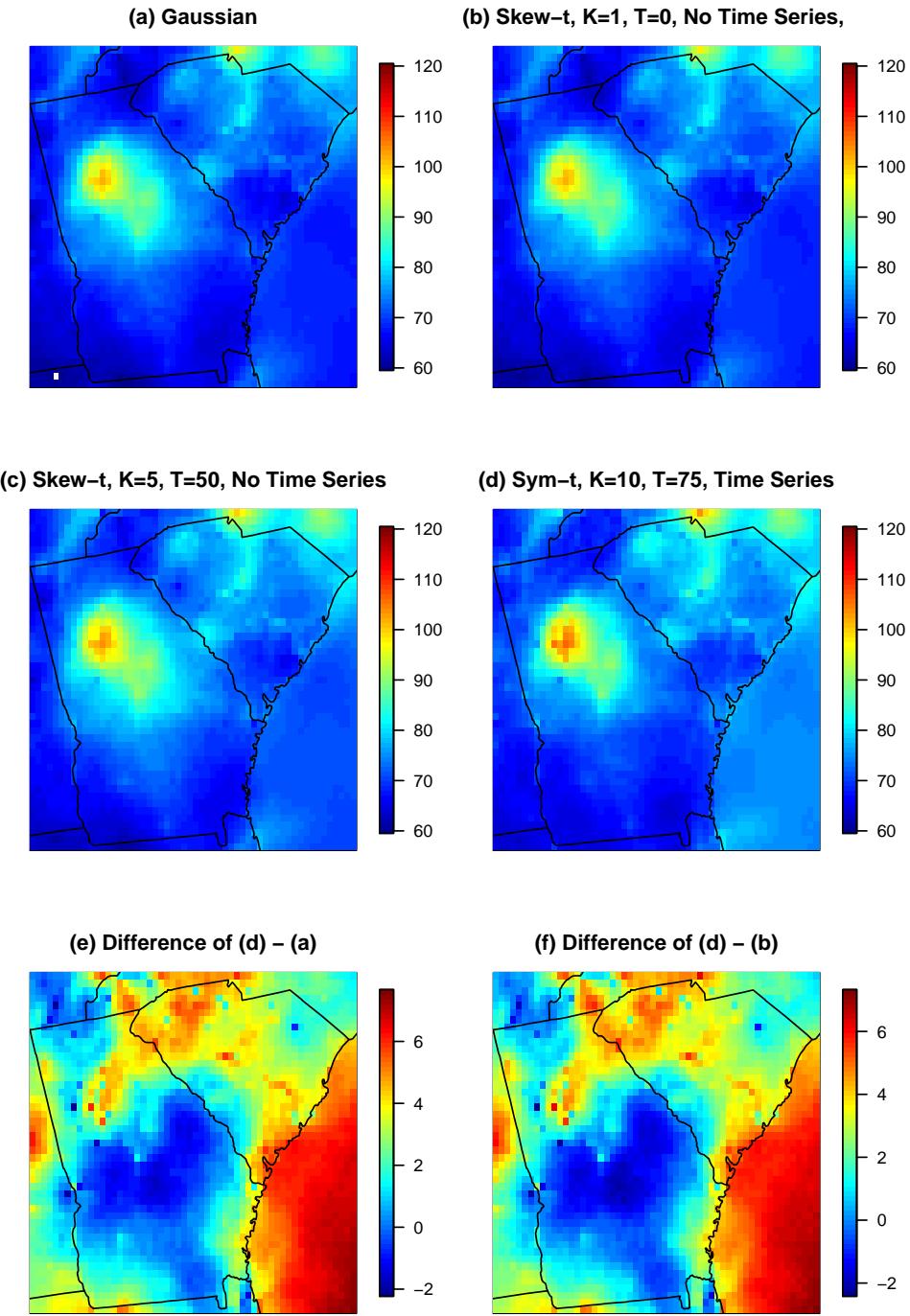


Figure 2.7 Panels (a) – (d) give the posterior predictive $\hat{q}(0.99)$ for the month of July under four different models, panel (e) gives the difference between $\hat{q}(0.99)$ in panels (d) and (a), panel (f) gives the difference between $\hat{q}(0.99)$ in panels (d) and (b).

model shows statistically significant improvements over a naïve Gaussian approach and in most cases, a max-stable approach. In both the simulation study, and the application to ozone data, we find that incorporating a partition in the model can improve extreme predictions. Furthermore the results from the data analysis suggest that thresholding can improve performance when predicting in the extreme tails of the data.

This model presents new avenues for future research. One possibility is the implementation of a different partition structure. We choose to define the random effects for a site by using an indicator function based on closeness to a knot. However, this indicator function could be replaced by kernel function that would allow for multiple knots to impact each site, with the weight of each knot to be determined by some characteristic such as distance. Another area that should be explored is the temporal dependence in the model. Instead of implementing a time series on the random effects, a three-dimensional covariance structure on the residuals could be implemented to address temporal dependence. Finally, we acknowledge that by specifying the number of knots, we may be underestimating the uncertainty in the model. This could be incorporated by treating the number of knots as a model parameter instead of fixing it to be a specific value.

CHAPTER

3

A SPATIAL MODEL FOR RARE BINARY EVENTS

3.1 Introduction

The goals of spatial binary data analysis are often to estimate covariate effects while accounting for spatial dependence and to make predictions at locations without samples. A common approach to incorporate spatial dependence in the model for binary data is relating a continuous spatial process $Z(\mathbf{s}) \in \mathbb{R}$ to the binary response $Y(\mathbf{s})$ by thresholding, $Y(\mathbf{s}) = I[Z(\mathbf{s}) > c]$, where $I[\cdot]$ is an indicator function. In many spatial analyses of binary data, a Gaussian process is used to model $Z(\mathbf{s})$. This is true for both spatial probit and spatial logistic regression. In this model, spatial dependence is determined by the joint probability that two sites simultaneously exceed the threshold c . However, when c is large, and thus $Y(\mathbf{s}) = 1$ is rare, then the asymptotic theory suggests that the Gaussian process will model dependence poorly. In fact, even under strong spatial correlation for Z , it gives asymptotic independence [Sib60], suggesting that for rare binary data, the Gaussian model will not perform very well.

We propose using a latent max-stable process [Haa84] for $Z(\mathbf{s})$ because it allows for asymptotic dependence. The max-stable process arises as the limit of the location-wise maximum of infinitely

many spatial processes, and any finite-dimensional representation of a max-stable process has generalized extreme value distribution (GEV) marginal distributions. Max-stable processes are extremely flexible, but are often challenging to work with in high dimensions [WT14; TO15]. To address this challenge, methods have been proposed that implement composite likelihood techniques for max-stable processes [Pad10; Gen11; HD14]. Composite likelihoods have been used to model binary spatial data [HL98], but not using max-stable processes. As an alternative to these composite approaches, Reich & Shaby [RS12] present a hierarchical model that implements a low-rank representation for a max-stable process. We chose to use this low-rank representation for our rare binary spatial regression model. Our model builds on related work by Wang & Dey [WD10] who use a GEV link for non-spatial binary data. The proposed model generalizes this to have spatial dependence.

The paper proceeds as follows. In Section 3.2 we present the proposed latent max-stable process for rare binary data analysis. In Section 3.3 we give the bivariate distribution for our model. In Section 3.4 we show a link between a commonly used measure of dependence between binary variables and another metric for extremal dependence. The computing for our model is outlined in Section 3.5. Finally, we present a simulation study in Section 3.6, and a data analysis on *Tamarix ramosissima* in Section 3.7. Lastly, in Section 3.8 we provide some discussion and possibilities for future research.

3.2 Spatial dependence for binary regression

Let $Y(\mathbf{s})$ be the binary response at spatial location \mathbf{s} in a spatial domain of interest $\mathcal{D} \in \mathbb{R}^2$. We assume $Y(\mathbf{s}) = I[Z(\mathbf{s}) > 0]$ where $Z(\mathbf{s})$ is a latent continuous max-stable process. The marginal distribution of $Z(\mathbf{s})$ at site \mathbf{s} is GEV with location $\mathbf{X}(\mathbf{s})^\top \boldsymbol{\beta}$, scale $\sigma > 0$, and shape ξ , where $\mathbf{X}(\mathbf{s})$ is a p -vector of spatial covariates at site \mathbf{s} and $\boldsymbol{\beta}$ is a p -vector of regression coefficients. We set $\sigma = 1$ for identifiability because only the sign and not the scale of Z affects Y . If $\mathbf{X}(\mathbf{s})^\top \boldsymbol{\beta} = \mu$ for all \mathbf{s} , then $P(Y = 1)$ is the same for all observations, and the two parameters μ and ξ are not individually identifiable, so when there are no covariates, we fix $\xi = 0$. Although $\boldsymbol{\beta}$ and ξ could be permitted to vary across space, we assume that they are constant across \mathcal{D} . At spatial location \mathbf{s} , the marginal distribution (over $Z(\mathbf{s})$) is $P[Y(\mathbf{s}) = 1] = 1 - \exp\left[-\frac{1}{z(\mathbf{s})}\right]$ where $z(\mathbf{s}) = [1 - \xi \mathbf{X}(\mathbf{s})^\top \boldsymbol{\beta}]^{1/\xi}$. This is the same as the marginal distribution given by Wang & Dey [WD10].

For a finite collection of locations $\mathbf{s}_1, \dots, \mathbf{s}_n$, we denote by $\mathbf{Y} = [Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n)]^T$ the vector of observations. The spatial dependence of \mathbf{Y} is determined by the joint distribution of the latent variable $\mathbf{Z} = [Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n)]^T$. To incorporate spatial dependence, we consider the hierarchical representation of the max-stable process proposed in Reich & Shaby [RS12]. Consider a set of positive

stable random effect $A_1, \dots, A_L \stackrel{\text{iid}}{\sim} \text{PS}(\alpha)$ associated with spatial knots $\mathbf{v}_1, \dots, \mathbf{v}_L \in \mathbb{R}^2$. The hierarchical model is given by

$$\mathbf{Z}(\mathbf{s}_i) | A_1, \dots, A_L \stackrel{\text{ind}}{\sim} \text{GEV}[\mathbf{X}(\mathbf{s}_i)^\top \boldsymbol{\beta} + \theta(\mathbf{s}_i), \alpha \theta(\mathbf{s}_i), \xi \alpha] \quad \text{and} \quad \theta(\mathbf{s}_i) = \left[\sum_{l=1}^L A_l w_l(\mathbf{s}_i)^{1/\alpha} \right]^\alpha \quad (3.1)$$

where $w_l(\mathbf{s}_i) > 0$ are a set of L weight functions that vary smoothly across space and satisfy $\sum_{l=1}^L w_l(\mathbf{s}) = 1$ for all \mathbf{s} , and $\alpha \in (0, 1)$ determines the strength of dependence, with α near zero giving strong dependence and $\alpha = 1$ giving joint independence.

Because the latent $\mathbf{Z}(\mathbf{s})$ are independent given the random effects $\theta(\mathbf{s})$, the binary responses are also conditionally independent. This leads to the tractable likelihood

$$Y(\mathbf{s}_i) | A_1, \dots, A_L \stackrel{\text{ind}}{\sim} \text{Bern}[\pi(\mathbf{s}_i)] \quad (3.2)$$

where

$$\pi(\mathbf{s}_i) = 1 - \exp \left\{ - \sum_{l=1}^L A_l \left(\frac{w_l(\mathbf{s}_i)}{z(\mathbf{s}_i)} \right)^{1/\alpha} \right\} \quad (3.3)$$

and $z(\mathbf{s}) = [1 + \xi \mathbf{X}(\mathbf{s})^\top \boldsymbol{\beta}]^{1/\xi}$. Marginally over the A_l , this gives

$$Z(\mathbf{s}) \sim \text{GEV}[\mathbf{X}(\mathbf{s})^\top \boldsymbol{\beta}, 1, \xi], \quad (3.4)$$

and thus $P[Y(\mathbf{s}) = 1] = 1 - \exp \left\{ - \frac{1}{z(\mathbf{s})} \right\}$.

Many weight functions are possible, but the weights must be constrained so that $\sum_{l=1}^L w_l(\mathbf{s}_i) = 1$ for $i = 1, \dots, n$ to preserve the marginal GEV distribution. For example, [RS12] take the weights to be scaled Gaussian kernels with knots \mathbf{v}_l ,

$$w_l(\mathbf{s}_i) = \frac{\exp[-0.5((\mathbf{s}_i - \mathbf{v}_l)/\rho)^2]}{\sum_{j=1}^L \exp[-0.5((\mathbf{s}_i - \mathbf{v}_j)/\rho)^2]} \quad (3.5)$$

where $\|\mathbf{s}_i - \mathbf{v}_l\|$ is the distance between site \mathbf{s}_i and knot \mathbf{v}_l , and the kernel bandwidth $\rho > 0$ determines the spatial range of the dependence, with large ρ giving long-range dependence and vice versa.

After marginalizing out the positive stable random effects, the joint distribution of \mathbf{Z} is

$$G(\mathbf{z}) = P[Z(\mathbf{s}_1) < z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n) < z(\mathbf{s}_n)] = \exp \left\{ - \sum_{l=1}^L \left[\sum_{i=1}^n \left(\frac{w_l(\mathbf{s}_i)}{z(\mathbf{s}_i)} \right)^{1/\alpha} \right]^\alpha \right\}, \quad (3.6)$$

where $G(\cdot)$ is the CDF of a multivariate GEV distribution. This is a special case of the multivariate GEV distribution with asymmetric Laplace dependence function [Taw90].

3.3 Joint distribution

We give an exact expression in the case where there are only two spatial locations which is useful for constructing a pairwise composite likelihood [Pad10] and studying spatial dependence. When $n = 2$, the probability mass function is given by

$$P[Y(\mathbf{s}_i) = y_i, Y(\mathbf{s}_j) = y_j] = \begin{cases} \varphi(\mathbf{z}), & y_i = 0, y_j = 0 \\ \exp \left\{ -\frac{1}{z(\mathbf{s}_i)} \right\} - \varphi(\mathbf{z}), & y_i = 1, y_j = 0 \\ \exp \left\{ -\frac{1}{z(\mathbf{s}_j)} \right\} - \varphi(\mathbf{z}), & y_i = 0, y_j = 1 \\ 1 - \exp \left\{ -\frac{1}{z(\mathbf{s}_i)} \right\} - \exp \left\{ -\frac{1}{z(\mathbf{s}_j)} \right\} + \varphi(\mathbf{z}), & y_i = 1, y_j = 1 \end{cases} \quad (3.7)$$

where $\varphi(\mathbf{z}) = \exp \left\{ - \sum_{l=1}^L \left[\left(\frac{w_l(\mathbf{s}_i)}{z(\mathbf{s}_i)} \right)^{1/\alpha} + \left(\frac{w_l(\mathbf{s}_j)}{z(\mathbf{s}_j)} \right)^{1/\alpha} \right]^\alpha \right\}$. For more than two locations, we are also able to compute the exact likelihood when the n is large but the number of events $K = \sum_{i=1}^n Y(\mathbf{s}_i)$ is small, as might be expected for very rare events, see Appendix B.2.

3.4 Quantifying spatial dependence

Assume that Z_1 and Z_2 are both $GEV(\beta, 1, 1)$ so that $P(Y_i = 1)$ decreases to zero as β increases. A common measure of dependence between binary variables is Cohen's Kappa [Coh60],

$$\kappa(\beta) = \frac{P_A - P_E}{1 - P_E} \quad (3.8)$$

where P_A is the joint probability of agreement $P(Y_1 = Y_2)$ and P_E is the joint probability of agreement under an assumption of independence $P(Y_i = 1)^2 + P(Y_i = 0)^2$. For the spatial model,

$$\begin{aligned} P_A(\beta) &= 1 - 2 \exp\left\{-\frac{1}{\beta}\right\} + 2 \exp\left\{-\frac{\vartheta(\mathbf{s}_1, \mathbf{s}_2)}{\beta}\right\} \\ P_E(\beta) &= 1 - 2 \exp\left\{-\frac{1}{\beta}\right\} + 2 \exp\left\{-\frac{2}{\beta}\right\}, \end{aligned}$$

and

$$\kappa(\beta) = \frac{P_A(\beta) - P_E(\beta)}{1 - P_E(\beta)} = \frac{\exp\left\{-\frac{\vartheta(\mathbf{s}_1, \mathbf{s}_2) - 1}{\beta}\right\} - \exp\left\{-\frac{1}{\beta}\right\}}{1 - \exp\left\{-\frac{1}{\beta}\right\}} \quad (3.9)$$

where $\vartheta(\mathbf{s}_i, \mathbf{s}_j) = \sum_{l=1}^L [w_l(\mathbf{s}_i)^{1/\alpha} + w_l(\mathbf{s}_j)^{1/\alpha}]^\alpha$ is the pairwise extremal coefficient given by Reich & Shaby [RS12]. To measure extremal dependence, let $\beta \rightarrow \infty$ so that events are increasingly rare. Then,

$$\kappa = \lim_{\beta \rightarrow \infty} \kappa(\beta) = 2 - \vartheta(\mathbf{s}_1, \mathbf{s}_2) \quad (3.10)$$

which is the same as the χ statistic of Coles [Col01], a commonly used measure of extremal dependence.

3.5 Computation

For small K , we can evaluate the likelihood directly. When K is large, we use Markov chain Monte Carlo (MCMC) methods with the random effects model to explore the posterior distribution. To overcome challenges with evaluating the positive stable density, we follow Reich & Shaby [RS12] and introduce a set of auxiliary variables B_1, \dots, B_L following the auxiliary variable technique of Stephenson [Ste09] (for more details, see Appendix A.3 of Reich & Shaby [RS12]). So, the hierarchical model is given by

$$\begin{aligned} Y(\mathbf{s}_i) | \pi(\mathbf{s}_i) &\stackrel{\text{ind}}{\sim} \text{Bern}[\pi(\mathbf{s}_i)] \\ \pi(\mathbf{s}_i) &= 1 - \exp\left\{-\sum_{l=1}^L A_l \left(\frac{w_l(\mathbf{s}_i)}{z(\mathbf{s}_i)}\right)^{1/\alpha}\right\} \\ A_l &\sim \text{PS}(\alpha) \end{aligned} \quad (3.11)$$

with priors $\boldsymbol{\beta} \sim N(\mathbf{0}, \sigma_\beta^2 \mathbf{I}_p)$, $\xi \sim N(0, \sigma_\xi^2)$, $\rho \sim \text{Unif}(\rho_l, \rho_u)$, and $\alpha \sim \text{Beta}(a_\alpha, b_\alpha)$. The model parameters are updated using Metropolis Hastings (MH) update steps, and the random effects A_1, \dots, A_L , and auxiliary variables B_1, \dots, B_L are updated using Hamiltonian Monte Carlo (HMC) update steps. The code for this is available online through <https://github.com/sammorris81/rare-binary>.

3.6 Simulation study

For our simulation study, we generate $n_m = 50$ datasets under 12 different simulation settings to explore the impact of sample size, sampling technique, and misspecification of link function. We generate data assuming three possible types of underlying process. For each of the underlying processes, we generate complete datasets on a 100×100 rectangular grid of $n = 10,000$ locations. If a dataset is generated with $K < 50$, it is discarded and a new dataset is generated. This is done to guarantee that datasets have no less than 0.5% rarity. For model fitting, we select a subsample and use the remaining sites to evaluate predictive performance. For all models, we run the MCMC sampler for 25,000 iterations with a burn-in period of 20,000 iterations. Convergence is assessed through visual inspection of traceplots.

3.6.1 Latent processes

The first process is a latent max-stable process that uses the GEV link described in (3.1) with knots on a 50×50 regularly spaced grid on $[0, 1] \times [0, 1]$. For this process, we set $\alpha = 0.35$, $\rho = 0.1$, and $\beta_0 \approx 2.97$ which gives $K = 500$ (5% rarity), on average. Because there are no covariates, we set $\xi = 0$. We then set $Y(\mathbf{s}) = I[Z(\mathbf{s}) > 0]$.

For the second process, we generate a latent variable from a spatial Gaussian process with a mean of $\text{logit}(0.05) \approx -2.94$ and an exponential covariance given by

$$\text{Cov}(\mathbf{s}_1, \mathbf{s}_2) = \tau_{\text{Gau}}^2 \exp\left\{-\frac{\|\mathbf{s}_1 - \mathbf{s}_2\|}{\rho_{\text{Gau}}}\right\} \quad (3.12)$$

where $\tau_{\text{Gau}} = 10$ and $\rho_{\text{Gau}} = 0.1$. Finally, we generate $Y(\mathbf{s}_i) \stackrel{\text{ind}}{\sim} \text{Bern}[\pi(\mathbf{s}_i)]$ where $\pi(\mathbf{s}_i) = \frac{\exp\{z(\mathbf{s})\}}{1 + \exp\{z(\mathbf{s})\}}$.

For the third process, we generate data using a hotspot method. For this process, we first generate hotspots throughout the space. Let n_{hs} be the number of hotspots in the space. Then $n_{\text{hs}} - 1 \sim \text{Poisson}(2)$. This generation scheme ensures that every dataset has at least one hotspot. We generate the hotspot locations $\mathbf{h}_1, \dots, \mathbf{h}_{n_{\text{hs}}} \sim \text{Unif}(0, 1)^2$. Let B_h be a circle of radius of radius r_h around hotspot $h = 1, \dots, n_{\text{hs}}$. The r_h differ for each hotspot and are generated i.i.d. from a $\text{Unif}(0.03, 0.08)$ distribution. We set $P[Y(\mathbf{s}_i) = 1] = 0.85$ for all \mathbf{s}_i in B_h , and $P[Y(\mathbf{s}_i)] = 0.0005$ for all \mathbf{s}_i outside of B_h .

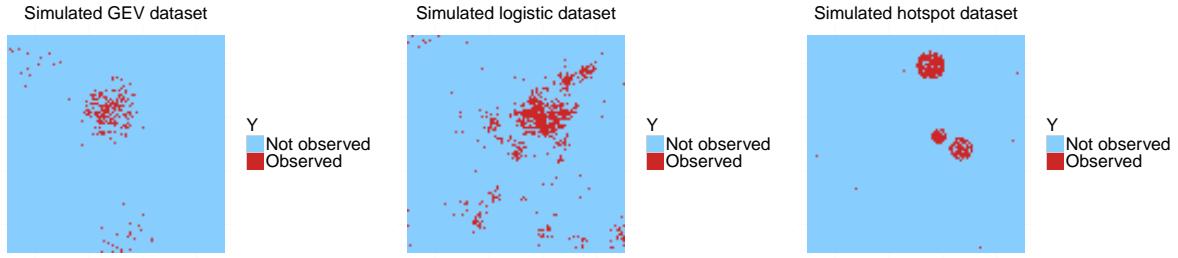


Figure 3.1 One simulated dataset from spatial GEV (left), spatial logistic (center), and hotspot (right) designs.

These settings are selected to give an average of approximately $K = 500$ for the datasets. Figure 3.1 gives an example dataset from each of the data settings.

3.6.2 Sampling methods

We subsample the generated data using $n_s = 100, 250$ initial locations for two different sampling designs. The first is a two-stage spatially-adaptive cluster technique (CLU) taken from Pacifici et al. [Pac16]. In this design, if an initial location is occupied, we also include the four rook neighbor (north, east, south, and west) sites in the sample. For the second design, we use a simple random sample (SRS) with the same number of sites included in the cluster sample. For the GEV setting, when $n_s = 100$, there are on average 117 sites and at most 142 sites in a sample, and when $n_s = 250$, there are on average 286 sites and at most 332 sites in a sample. For the logistic setting, when $n_s = 100$, there are on average 118 sites and at most 147 sites in a sample, and when $n_s = 250$, there are on average 290 sites and at most 330 sites in a sample. For the hotspot setting, when $n_s = 100$, there are on average 110 sites and at most 128 sites in a sample, and when $n_s = 250$, there are on average 275 sites and at most 306 sites in a sample.

3.6.3 Methods

For each dataset, we fit the model using three different models: the proposed spatial GEV model, a spatial probit model, and a spatial logistic model. Logistic and probit methods assume the underlying process is Gaussian. In this case, we assume that $Z(\mathbf{s})$ follows a Gaussian process with mean $\mathbf{X}(s)^T \boldsymbol{\beta}$ and unit variance. For the simulation study, we use an intercept only model. The marginal

distributions are given by

$$P[Y(\mathbf{s}) = 1] = \begin{cases} \frac{\exp[\mathbf{X}^\top(\mathbf{s})\boldsymbol{\beta} + \mathbf{W}(\mathbf{s})\boldsymbol{\epsilon}]}{1 + \exp[\mathbf{X}^\top(\mathbf{s})\boldsymbol{\beta} + \mathbf{W}(\mathbf{s})\boldsymbol{\epsilon}]}, & \text{logistic} \\ \Phi[\mathbf{X}^\top\boldsymbol{\beta}(\mathbf{s}) + \mathbf{W}(\mathbf{s})\boldsymbol{\epsilon}], & \text{probit} \end{cases} \quad (3.13)$$

where $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \tau_L^2 \mathbf{I}_L)$ are Gaussian random effects at the knot locations, and $\mathbf{W}(\mathbf{s})$ are a set of L basis functions given to recreate the Gaussian process at all sites. We use our own code for the spatial probit model, but we use the `spGLM` function in the `spBayes` package [Fin15] to fit the spatial logistic model. For the probit model, we use

$$\mathbf{W}_l(\mathbf{s}_i) = \frac{\exp[-(||\mathbf{s}_i - \mathbf{v}_l||/\rho)^2]}{\sqrt{\sum_{j=1}^L \exp[-(||\mathbf{s}_i - \mathbf{v}_j||/\rho)^2]^2}}. \quad (3.14)$$

For the logistic model, the $\mathbf{W}_l(\mathbf{s}_i)$ are the default implementation from the `spGLM`.

3.6.4 Priors

For all models, we only include an intercept term β_0 in the model, and the prior for the intercept is $\beta_0 \sim N(0, 10)$. Additionally, for all models, the prior for the bandwidth is $\rho \sim \text{Unif}(0.001, 1)$. In all methods, we place knots at all data points. For the GEV method, the prior for the spatial dependence parameter is $\alpha \sim \text{Beta}(2, 5)$. We select this prior because it gives greater weight to $\alpha < 0.5$, which is the point at which spatial dependence becomes fairly weak, but also avoids values below 0.1 which can lead to numerical problems. We fix $\xi = 0$ because we do not include any covariates. For both the spatial probit and logistic models, the prior on the variance term for the random effects is $\text{IG}(0.1, 0.1)$ where $\text{IG}(\cdot)$ is an Inverse Gamma distribution.

3.6.5 Model comparisons

For each dataset, we fit the model using the n_s observations as a training set, and validate the model's predictive power at the remaining grid points. Let \mathbf{s}_j^* be the j th site in the validation set. From the posterior distributions of the parameters we can calculate $P[Y(\mathbf{s}_j^*) = 1]$. To obtain $\hat{P}[Y(\mathbf{s}_j^*) = 1]$, we take the average of the posterior distribution for each j . We consider a few different metrics for comparing model performance. One score is the Brier scores [GR07, BS]. The Brier score for predicting an occurrence at site \mathbf{s} is given by $\{I[Y(\mathbf{s}) = 1] - \hat{P}[Y(\mathbf{s}) = 1]\}^2$ where $I[Y(\mathbf{s}) = 1]$ is an

Table 3.1 Brier scores ($\times 100$) [SE] and AUROC [SE] for GEV, Probit, and Logistic methods from the simulation study.

Setting	<i>n</i>	Sample Type	BS			AUROC		
			GEV	Probit	Logistic	GEV	Probit	Logistic
GEV	100	CLU	3.10 [0.27]	2.45 [0.19]	2.79 [0.25]	0.926 [0.009]	0.942 [0.009]	0.900 [0.020]
		SRS	2.92 [0.20]	2.54 [0.18]	2.92 [0.25]	0.938 [0.007]	0.951 [0.007]	0.879 [0.021]
	250	CLU	2.18 [0.15]	1.87 [0.13]	2.05 [0.14]	0.951 [0.008]	0.948 [0.011]	0.922 [0.017]
		SRS	2.29 [0.15]	2.06 [0.13]	2.26 [0.15]	0.949 [0.009]	0.949 [0.010]	0.908 [0.020]
Logistic	100	CLU	5.29 [0.25]	4.94 [0.23]	5.10 [0.25]	0.659 [0.012]	0.676 [0.014]	0.643 [0.013]
		SRS	5.32 [0.23]	5.09 [0.24]	5.34 [0.26]	0.690 [0.012]	0.693 [0.012]	0.613 [0.012]
	250	CLU	4.81 [0.21]	4.55 [0.21]	4.66 [0.22]	0.731 [0.010]	0.749 [0.010]	0.714 [0.014]
		SRS	4.86 [0.22]	4.63 [0.20]	5.01 [0.23]	0.742 [0.010]	0.760 [0.010]	0.698 [0.015]
Hotspot	100	CLU	2.29 [0.17]	2.01 [0.15]	1.81 [0.12]	0.841 [0.016]	0.833 [0.019]	0.824 [0.020]
		SRS	2.09 [0.13]	1.87 [0.12]	2.13 [0.15]	0.885 [0.015]	0.906 [0.013]	0.844 [0.015]
	250	CLU	1.65 [0.11]	1.25 [0.08]	1.40 [0.09]	0.934 [0.009]	0.949 [0.008]	0.939 [0.011]
		SRS	1.53 [0.10]	1.31 [0.08]	1.63 [0.11]	0.947 [0.007]	0.960 [0.005]	0.918 [0.015]

indicator function indicating that an event occurred at site s . We average the Brier scores over all test sites, and a lower score indicates a better fit. The Brier score equally penalizes false negatives and false positives, but in the case of rare data, this may not be the best metric due to the unbalanced nature of the data. Therefore, we also consider the receiver operating characteristic (ROC) curve, and the area under the ROC curve (AUROC) for the different methods and settings. The ROC curve and AUROC are obtained via the `ROCR` [Sin05] package in R [R C16]. We then average AUROC across all datasets for each method and setting to obtain a single AUROC for each combination of method and setting.

3.6.6 Results

Overall, we find that the spatial probit model actually performs quite well in all cases. Table 3.1 gives the Brier scores and AUROC for each of the methods. Looking at Brier scores, we see that our model is outperformed by the probit model in all cases, by the logistic models in many settings. For AUROC, in a few of the settings, we do demonstrate a small improvement over the probit and logistic models. Because these results are somewhat surprising, we also considered the performance metrics by rareness of the data. We plot the Brier scores and AUROC for a few different sampling methods using the GEV and logistic links in Figure 3.2 – Figure 3.4 using a Loess smoother. These plots give evidence to suggest that as rareness increases, the spatial GEV method has potential to outperform the spatial probit and logistic models based on AUROC.

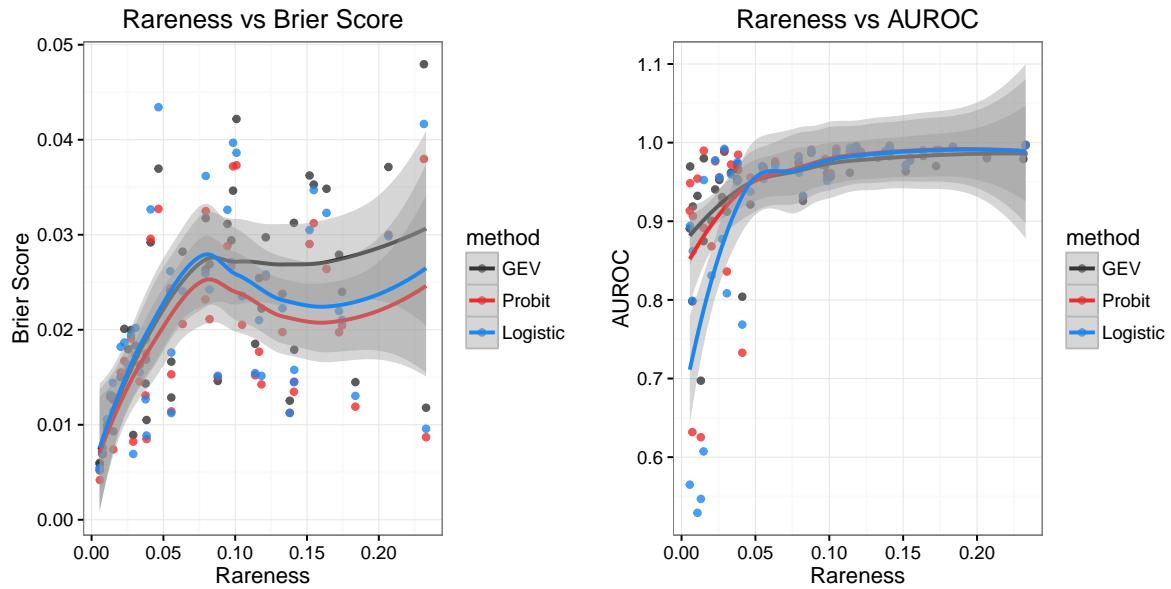


Figure 3.2 Smooth of BS (left) and AUROC (right) by rareness for GEV link and cluster sampling with $n_s = 250$.

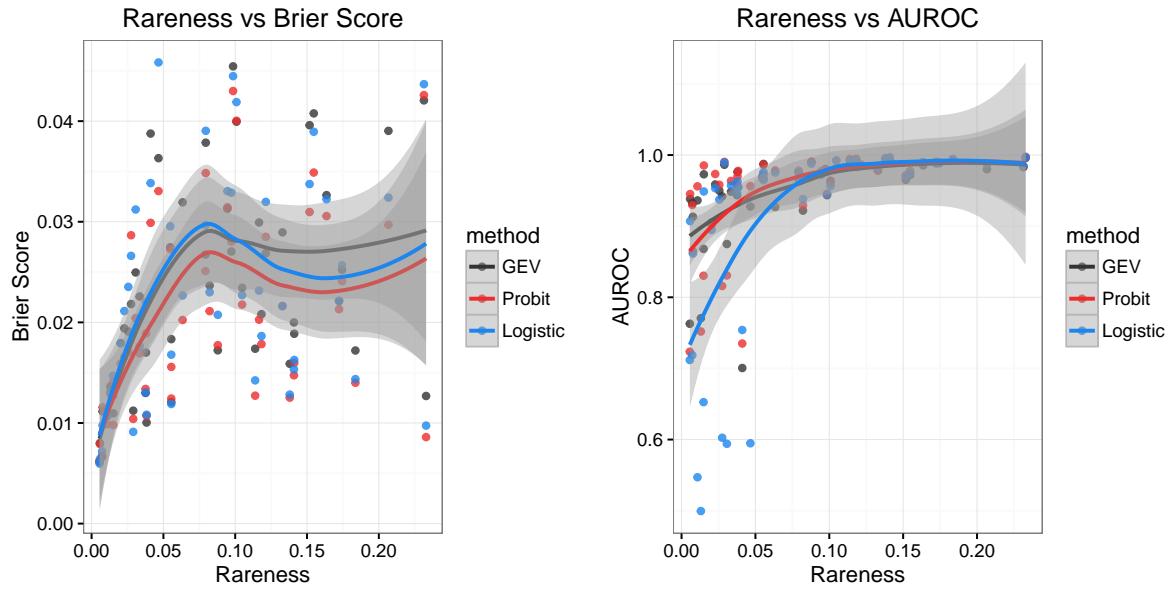


Figure 3.3 Smooth of BS (left) and AUROC (right) by rareness for GEV link and simple random sampling with $n_s = 250$.

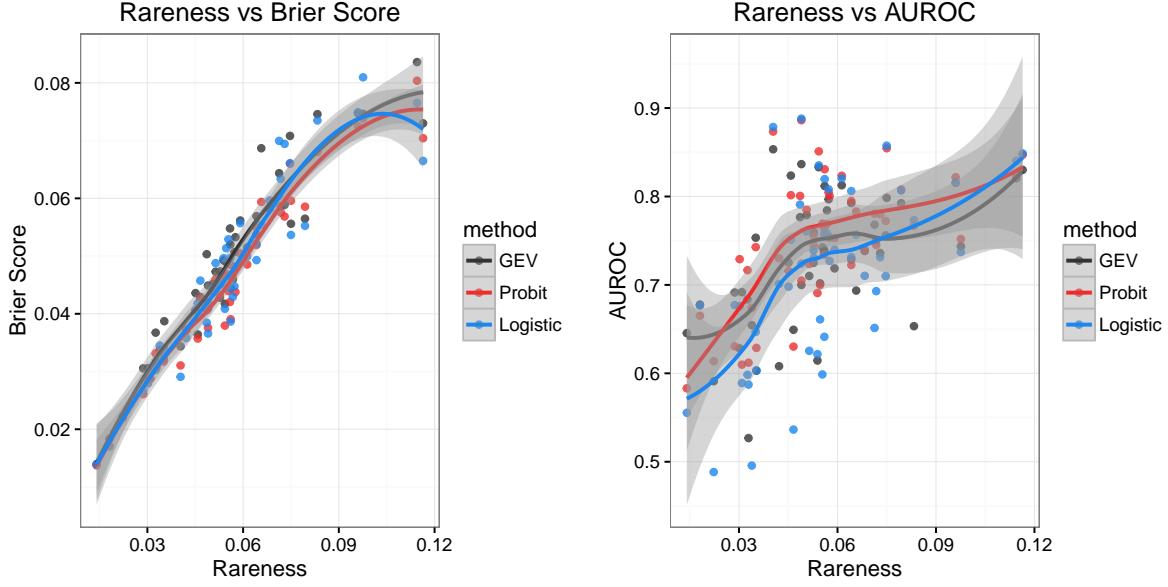


Figure 3.4 Smooth of BS (left) and AUROC (right) by rareness for logistic link and cluster sampling with $n_s = 250$.

3.7 Data analysis

We compare our method to the spatial probit and logit for mapping the probability of the occurrence of *Tamarix ramosissima* (TR) and *Hedysarum scoparium* (HS), two plant species, for a 1-km² study region of PR China [Smi12]. The Chinese Academy of Forestry conducted a full census of the area, and the true occupancy of the species are plotted in Figure 3.5. The region is split into 10-m × 10-m grid cells. *Tamarix ramosissima* can be found in approximately 6% of the grid cells, and *Hedysarum scoparium* can be found in approximately 0.54% of the grid cells.

3.7.1 Methods

For the data analysis, we generate 100 subsamples using the CLU and SRS sampling methods with $n_s = 100, 250$ initial locations. For each subsample, we fit the spatial GEV, spatial probit, and spatial logistic models. Knot placement, prior distributions, and MCMC details for the data analysis are the same as the simulation study. To compare models, we use similar metrics as in the simulation study, but we average the metrics over subsamples.

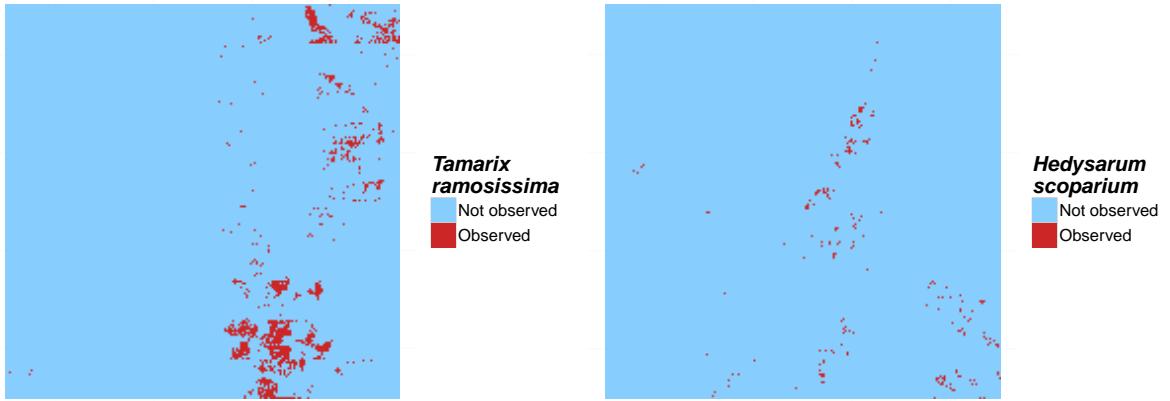


Figure 3.5 True occupancy of *Tamarix ramosissima* (left) and *Hedysarum scoparium* (right) from a 1-km² study region of PR China.

Table 3.2 Brier scores ($\times 100$) [SE] and AUROC [SE] for GEV, Probit, and Logistic methods for *Tamarix ramosissima*.

n	Sample Type	BS			AUROC		
		GEV	Probit	Logistic	GEV	Probit	Logistic
100	CLU	5.15 [0.05]	5.08 [0.04]	5.35 [0.02]	0.747 [0.009]	0.742 [0.010]	0.701 [0.009]
	SRS	5.03 [0.04]	4.97 [0.04]	5.51 [0.02]	0.796 [0.006]	0.802 [0.006]	0.631 [0.008]
250	CLU	4.78 [0.03]	4.68 [0.03]	4.98 [0.04]	0.772 [0.009]	0.783 [0.009]	0.790 [0.008]
	SRS	4.83 [0.04]	4.75 [0.03]	5.17 [0.05]	0.826 [0.007]	0.848 [0.005]	0.712 [0.013]

3.7.2 Results

As with the simulation study, we find that in most cases the spatial probit model gives the best performance. Table 3.2 give summary Brier scores ($\times 100$) and AUROC for the *Tamarix ramosissima*. Figure 3.6 gives the vertically averaged ROC curves for each method and sampling setting.

3.8 Discussion and future research

In this paper, we present a max-stable spatial method for rare binary data. The principal finding in this paper is that the spatial probit model tends to outperform the proposed model. This finding is surprising given that the max-stable process is the theoretically justified spatial process for extreme value distributions, and it leads to possible research questions in the future.

It is unusual that the spatial probit model should outperform the proposed model, particularly when the data are generated directly from the proposed model. One possible explanation is that for the simulated data, there is a wide range of rarity in the data (GEV: 0.5% – 35.9%, Logistic: 1.4% –

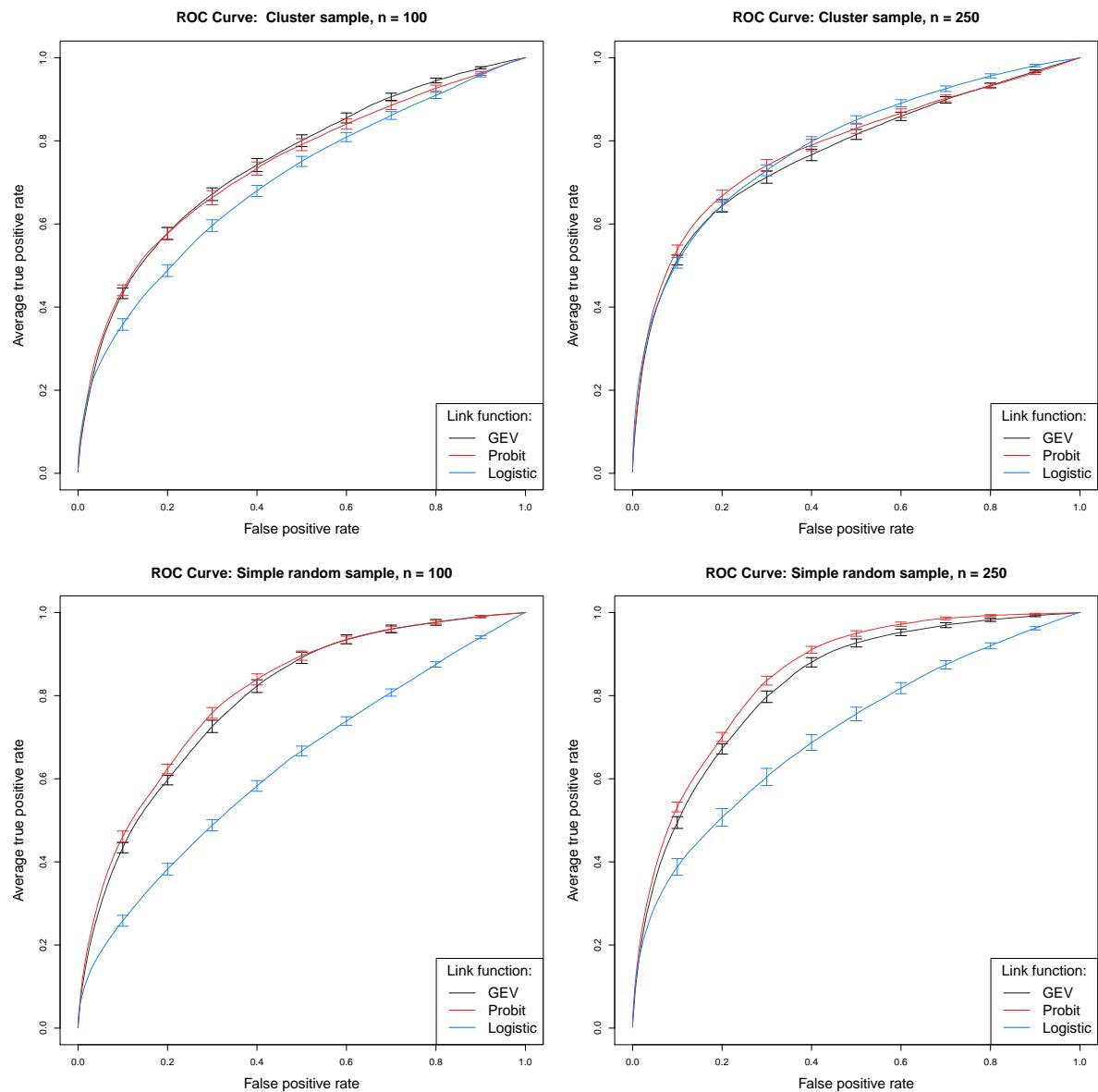


Figure 3.6 Vertically averaged ROC curves for *Tamarix ramosissima*.

14.4%, and Hotspot: 0.5% – 6.8%). Given that for both the GEV and logistic data settings, we have a number of datasets with a relatively high rate of occurrence, it is possible that probit is a competitive method. In particular, it may be useful to have a slightly more restrictive data generation strategy (i.e. restrict datasets to $K < 500$).

CHAPTER

4

EMPIRICAL BASIS FUNCTIONS FOR MAX-STABLE SPATIAL DEPENDENCE

4.1 Introduction

The spatial Extreme Value Analysis (EVA) literature is expanding rapidly [Dav12] to meet the demands of researchers to improve estimates of rare-event probabilities by borrowing information across space and to estimate the probability of extreme events occurring simultaneously at multiple locations. Environmental datasets commonly include observations from hundreds or thousands of locations, and advanced tools are required to explore and analyze these data. For Gaussian data, Principle Components Analysis [EH08, PCA], also known as Empirically Orthogonal Functions [Han07, EOF], has proven to be a powerful tool to study correlation between spatial locations; understand the most important large-scale spatial features; and reduce the dimension of the problem to allow for simple computation even for massive datasets. Computation and exploration are arguably more difficult for EVA than Gaussian data, yet to our knowledge no tool analogous to spatial PCA has been developed for EVA.

In EVA, extremes are separated from the bulk of the distribution by either analyzing only points above a threshold or block maximums [Col01], e.g., the annual maximum of the daily precipitation.

A natural spatial model for block maximum at several spatial locations is the max-stable process, which, under certain conditions, arises as the limit of the location-wise maximum of infinitely-many spatial processes [HF06]. Max-stable processes were also used to model spatial exceedances over a high threshold [Thi13; HD14]. Haan [Haa84] showed that any max-stable process can be represented in terms of a countable number of spatial processes (e.g., stationary log Gaussian processes), and a finite truncation of this representation has been used for conditional simulation [WS11]. Fully-Bayesian analysis using max-stable processes is cumbersome for large data sets [WT14; TO15]. One option is to use non-max-stable models that retain extremal dependence such as the skew-t process in [Morde]. Alternatively, Reich & Shaby [RS12] propose a low-rank method based on spatial kernel functions, and others have used pairwise [Pad10; HD14] and trivariate [Gen11] likelihood methods for parameter estimation.

In this paper we propose an empirical basis function (EBF) approach that builds on a finite truncation of the spectral representation, and develops a method-of-moments estimator for the underlying spatial processes. Unlike PCA/EOFs, but similar to dictionary learning [Mai14] and non-negative matrix factorizations [LS99], the EBFs are not orthogonal, nonetheless these spatial functions can be plotted for exploratory analysis to reveal important spatial trends. In addition to exploratory analysis, we show that the EBFs can be used for Bayesian inference on the marginal parameters at each location, modeling spatial dependence, and to test for covariate effects. By basing the spatial dependence on EBFs, the resulting spatial analysis does not require dubious assumptions such as stationarity. In addition, a Bayesian analysis for either block-maximum or point above a threshold is computationally feasible for large datasets because the entire spatial process is represented by a small number of basis functions.

The paper proceeds as follows. In Section 4.2 we present the low-rank model. Section 4.3 describes the algorithm used to estimate the spatial basis functions, and Section 4.4 describes the model fit using Markov chain Monte Carlo (MCMC) methods. In Section 4.5 we demonstrate the use of the EBFs for an analysis of wildfire data in Georgia and precipitation data in the eastern U.S. Lastly in Section 4.6 we give some summary conclusions and a brief discussion of the findings.

4.2 Model

Let $Y_t(\mathbf{s})$ be the observation at spatial location \mathbf{s} and time t . We temporarily drop the subscript t and describe the model for the process $Y(\mathbf{s})$ for a single time point, but return to the spatiotemporal notation in Section 4.3. To focus attention on the extreme values, we emphasize the statistical model for exceedances above a location-specific threshold $T(\mathbf{s})$. We begin by specifying a spatial model for the complete data $Y(\mathbf{s})$ and then use the censored likelihood defined by $T(\mathbf{s})$ for inference as

described in Section 4.4. Although the model presented implements a censored likelihood, the model also can fit uncensored data (such as block-maxima) by setting $T(\mathbf{s}) = -\infty$.

Spatial dependence is captured by modeling $Y(\mathbf{s})$ as a max-stable process [HF06]. Max-stable processes have generalized extreme value (GEV; see Appendix C.1) marginal distribution. The GEV has three parameters: location $\mu(\mathbf{s})$; scale $\sigma(\mathbf{s})$; and shape $\xi(\mathbf{s})$. Spatial dependence is present both in the GEV parameters but also the standardized residual process

$$Z(\mathbf{s}) = \left\{ 1 + \frac{\xi(\mathbf{s})}{\sigma(\mathbf{s})} [Y(\mathbf{s}) - \mu(\mathbf{s})] \right\}^{1/\xi(\mathbf{s})}, \quad (4.1)$$

which has unit Fréchet (i.e., GEV with location, scale, and shape all equal one) marginal distribution for all \mathbf{s} .

Our objective is to identify a low-rank model for the spatial dependence of $Z(\mathbf{s})$. Haan [Haa84, Chapter 9] show that any max-stable process can be written as

$$Z(\mathbf{s}) = \bigvee_{l=1}^{\infty} B(\mathbf{s}, \mathbf{k}_l) A_l \quad (4.2)$$

where the functions $B(\mathbf{s}, \mathbf{k}_l)$ satisfy $B(\mathbf{s}, \mathbf{k}_l) > 0$ for all \mathbf{s} and $\int B(\mathbf{s}, \mathbf{k}_l) d\mathbf{k}_l = 1$ for all \mathbf{s} , and (\mathbf{k}_l, A_l) for $l = 1, \dots, \infty$ are a Poisson process with intensity measure $dA d\mathbf{k}/A^2$. In many representations of max-stable process, such as Smith [Smi90] and Reich & Shaby [RS12], the \mathbf{k}_l are spatial locations that represent the center of process l ; however, in our proposed method the basis functions are not associated with one particular location and so to simplify notation we let $B_l(\mathbf{s}) = B(\mathbf{s}; \mathbf{k}_l)$

To arrive at a low-rank model, we assume there are a finite and known number of spatial basis functions $B_1(\mathbf{s}), \dots, B_L(\mathbf{s})$ that explain the important spatial variation in the process. As in de Haan's expansion, the basis functions are restricted so that $B_l(\mathbf{s}) > 0$ and $\sum_{l=1}^L B_l(\mathbf{s}) = 1$ for all \mathbf{s} . Because it is unrealistic to assume that realizations of Z are exactly functions of L basis functions, we include independent error variables $\varepsilon(\mathbf{s})$ to capture variation not explained by the $B_l(\mathbf{s})$. We follow Reich & Shaby [RS12] and decompose $Z(\mathbf{s})$ as $Z(\mathbf{s}) = \theta(\mathbf{s})\varepsilon(\mathbf{s})$ where $\theta(\mathbf{s})$ is a spatial process and $\varepsilon(\mathbf{s}) \stackrel{iid}{\sim} \text{GEV}(1, \alpha, \alpha)$ is independent error. The spatial component is

$$\theta(\mathbf{s}) = \left(\sum_{l=1}^L B_l(\mathbf{s})^{1/\alpha} A_l \right)^{\alpha}. \quad (4.3)$$

If $B_l(\mathbf{s}) > 0$, $\sum_{l=1}^L B_l(\mathbf{s}) = 1$ for all \mathbf{s} , and the A_l have positive stable (PS; Appendix C.2) distribution

$A_l \stackrel{\text{iid}}{\sim} \text{PS}(\alpha)$, then $Z(\mathbf{s})$ is max-stable and has unit Fréchet marginal distributions.

Extremal spatial dependence for max-stable processes can be summarized by the extremal coefficient [ST03, EC] $\vartheta(\mathbf{s}_1, \mathbf{s}_2) \in [1, 2]$, where

$$\text{Prob}[Z(\mathbf{s}_1) < c, Z(\mathbf{s}_2) < c] = \text{Prob}[Z(\mathbf{s}_1) < c]^{\vartheta(\mathbf{s}_1, \mathbf{s}_2)}. \quad (4.4)$$

For the PS random effects model the EC has the form

$$\vartheta(\mathbf{s}_1, \mathbf{s}_2) = \sum_{l=1}^L [B_l(\mathbf{s}_1)^{1/\alpha} + B_l(\mathbf{s}_2)^{1/\alpha}]^\alpha. \quad (4.5)$$

In particular, $\vartheta(\mathbf{s}, \mathbf{s}) = 2^\alpha$ for all \mathbf{s} .

4.3 Estimating the spatial dependence function

To estimate the extremal coefficient function, we consider the process at n_s spatial locations $\mathbf{s}_1, \dots, \mathbf{s}_{n_s}$ and n_t times $t = 1, \dots, n_t$. The basis functions are fixed over time, but the random effects and errors are independent over time. That is $Z_t(\mathbf{s}) = \theta_t(\mathbf{s})\epsilon_t(\mathbf{s})$ where $\theta_t(\mathbf{s}) = \left(\sum_{l=1}^L B_l(\mathbf{s})^{1/\alpha} A_{lt} \right)^\alpha$,

$A_{lt} \stackrel{\text{iid}}{\sim} \text{PS}(\alpha)$, and $\epsilon_t(\mathbf{s}) \stackrel{\text{iid}}{\sim} \text{GEV}(1, \alpha, \alpha)$. Denote $Y_t(\mathbf{s}_i) = Y_{it}$, $B_l(\mathbf{s}_i) = B_{il}$, $T(\mathbf{s}_i) = T_i$, and $\vartheta(\mathbf{s}_i, \mathbf{s}_j) = \vartheta_{ij}$.

In this section we develop an algorithm to estimate the spatial dependence parameter α and the $n_s \times L$ matrix $\mathbf{B} = \{B_{il}\}$. Our algorithm has the following steps:

- (1) Obtain an initial estimate of the extremal coefficient for each pair of locations, $\hat{\vartheta}_{ij}$.
- (2) Spatially smooth these initial estimates $\hat{\vartheta}_{ij}$ using kernel smoothing to obtain $\tilde{\vartheta}_{ij}$.
- (3) Estimate the spatial dependence parameters by minimizing the difference between model-based coefficients, ϑ_{ij} , and smoothed coefficients, $\tilde{\vartheta}_{ij}$.

The first-stage estimates are obtained using an empirical estimate as follows. To estimate the spatial dependence we first remove variation in the marginal distribution. Let $U_{it} = \sum_{k=1}^{n_t} I[Y_{ik} < Y_{it}]/n_t$, so that the U_{it} are approximately uniform at each location. Then for some extreme probability $q \in (0, 1)$, solving (4.4) suggests the estimate

$$\hat{\vartheta}_{ij}(q) = \frac{\log[Q_{ij}(q)]}{\log(q)}, \quad (4.6)$$

where $Q_{ij}(q) = \sum_{t=1}^{n_t} I[U_{it} < q, U_{jt} < q]/n_t$ is the sample proportion of the time points at which both sites are less than q . Since all large q give valid estimates, we average over a grid of q with $q_1 < \dots < q_{n_q}$

$$\hat{\vartheta}_{ij} = \frac{1}{n_q} \sum_{j=1}^{n_q} \hat{\vartheta}_{ij}(q_j). \quad (4.7)$$

Assuming the true EC is smooth over space, the initial estimates $\hat{\vartheta}_{ij}$ can be improved by smoothing. Let

$$\tilde{\vartheta}_{ij} = \frac{\sum_{u=1}^{n_s} \sum_{v=1}^{n_s} w_{iu} w_{jv} \hat{\vartheta}_{uv}}{\sum_{u=1}^{n_s} \sum_{v=1}^{n_s} w_{iu} w_{jv}}, \quad (4.8)$$

where $w_{iu} = \exp[-(\|\mathbf{s}_i - \mathbf{s}'_u\|/\phi)^2]$ is the Gaussian kernel function with bandwidth ϕ . The elements $\hat{\vartheta}_{ii}$ do not contribute any information as $\hat{\vartheta}_{ii} = 1$ for all i by construction. To eliminate the influence of these estimates we set $w_{ii} = 0$. However, this approach does give imputed values $\tilde{\vartheta}_{ii}$, which provide information about small-scale spatial variability.

The dependence parameters B_{lt} and α are estimated by comparing estimates $\tilde{\vartheta}_{ij}$ with the model-based values ϑ_{ij} . For all i , $\vartheta_{ii} = 2^\alpha$, and therefore we set α to $\hat{\alpha} = \log_2 \left(\sum_{i=1}^{n_s} \tilde{\vartheta}_{ii} / n_s \right)$. Given $\alpha = \hat{\alpha}$, it remains to estimate \mathbf{B} . Similarly to Smith [Smi90] for a stationary max-stable process, we use squared-error loss, so the estimate $\hat{\mathbf{B}}$ is the minimizer of

$$\sum_{i < j} (\tilde{\vartheta}_{ij} - \vartheta_{ij})^2 = \sum_{i < j} \left(\tilde{\vartheta}_{ji} - \sum_{l=1}^L [B_{il}^{1/\hat{\alpha}} + B_{jl}^{1/\hat{\alpha}}]^{\hat{\alpha}} \right)^2 \quad (4.9)$$

under the restrictions that $B_{il} \geq 0$ for all i and l and $\sum_{l=1}^L B_{il} = 1$ for all i . Since the minimizer of (4.9) does not have a closed form, we use block coordinate descent to obtain $\hat{\mathbf{B}}$. We cycle through spatial locations and update the vectors $(\hat{B}_{i1}, \dots, \hat{B}_{iL})$ conditioned on the values for the other location and repeat until convergence. At each step, we use the restricted optimization routine in the R function `optim`. This algorithm gives estimates of the B_{il} at the n_s data locations, but is easily extended to all \mathbf{s} for spatial prediction. The kernel smoothing step ensures that the estimates for \hat{B}_{il} are spatially smooth, and thus interpolation of the \hat{B}_{il} gives spatial functions $\hat{B}_l(\mathbf{s})$.

These functions provide useful exploratory data analysis techniques. Maps of $\hat{B}_l(\mathbf{s})$ show important spatial features in the extremal dependence. Furthermore, they allow for a non-stationary spatial dependence structure. The relative contribution of each term can be measured by

$$\nu_l = \frac{1}{n_s} \sum_{i=1}^{n_s} \hat{B}_{il}. \quad (4.10)$$

Since $\sum_{l=1}^L \hat{B}_{il} = 1$ for all i , we have $\sum_{l=1}^L \nu_l = 1$. Therefore, terms with large ν_l are the most important. The order of the terms is arbitrary, and so we reorder the terms so that $\nu_1 \geq \dots \geq \nu_L$.

4.4 Bayesian implementation details

For our data analysis in Section 4.5 we allow the GEV location and scale parameters, denoted $\mu_t(\mathbf{s})$ and scale $\sigma_t(\mathbf{s})$ respectively, to vary with space and time. The model we choose is as follows

$$\mu_t(\mathbf{s}) = \beta_{1,int}(\mathbf{s}) + \beta_{1,time}(\mathbf{s})t \quad (4.11)$$

$$\log[\sigma_t(\mathbf{s})] = \beta_{2,int}(\mathbf{s}) + \beta_{2,time}(\mathbf{s})t \quad (4.12)$$

where

$$\begin{aligned} \beta_{1,int}(\mathbf{s}) &\sim N(\mu_{1,int} \mathbf{1}, \sigma_{1,int}^2 \Sigma) & \beta_{1,time}(\mathbf{s}) &\sim N(\mu_{1,time} \mathbf{1}, \sigma_{1,time}^2 \Sigma) \\ \beta_{2,int}(\mathbf{s}) &\sim N(\mu_{2,int} \mathbf{1}, \sigma_{2,int}^2 \Sigma) & \beta_{2,time}(\mathbf{s}) &\sim N(\mu_{2,time} \mathbf{1}, \sigma_{2,time}^2 \Sigma) \end{aligned} \quad (4.13)$$

are Gaussian process priors and Σ is an exponential spatial correlation matrix obtained from $\rho(h) = \exp\left\{-\frac{h}{\phi}\right\}$ where $h = \|\mathbf{s}_1 - \mathbf{s}_2\|$ is the Euclidean distance between sites \mathbf{s}_1 and \mathbf{s}_2 . The GEV shape parameter ξ is held constant over space and time because this parameter is challenging to estimate. Collectively, let the marginal GEV parameters at location i and time t be $\Theta_{it} = \{\mu_{it}, \sigma_{it}, \xi\}$ where $\mu_{it} = \mu_t(\mathbf{s}_i)$ and $\sigma_{it} = \sigma_t(\mathbf{s}_i)$.

As shown in Reich & Shaby [RS12], the uncensored responses $Y_t(\mathbf{s})$ are conditionally independent given the spatial random effects, with conditional distribution

$$Y_{it} | \theta_{it}, \Theta_{it} \stackrel{\text{ind}}{\sim} \text{GEV}(\mu_{it}^*, \sigma_{it}^*, \xi^*), \quad (4.14)$$

where $\mu_{it}^* = \mu_{it} + \frac{\sigma_{it}}{\xi}(\theta_{it}^\xi - 1)$, $\sigma_{it}^* = \alpha \sigma_{it} \theta_{it}^\xi$, and $\xi^* = \alpha \xi$. Therefore, the conditional likelihood conveniently factors across observations; marginalizing over the random effect θ_{it} induces extremal

spatial dependence. To focus on the extreme values above the local threshold T_i , we use the censored likelihood

$$d(y; \theta_{it}, \Theta_{it}, T_i) = \begin{cases} F(y; \mu_{it}^*, \sigma_{it}^*, \xi^*) & y \leq T_i \\ f(y; \mu_{it}^*, \sigma_{it}^*, \xi^*) & y > T_i, \end{cases} \quad (4.15)$$

where F and f are the GEV distribution and density functions, respectively, defined in Appendix C.1.

In summary, given the estimates of α and \mathbf{B} , the hierarchical model is

$$\begin{aligned} Y_{it} | \theta_{ij} &\stackrel{\text{indep}}{\sim} d(y; \theta_{it}, \Theta_{it}, T_i) \\ \theta_{it} &= \left(\sum_{l=1}^L \hat{B}_{il}^{1/\hat{\alpha}} A_{lt} \right)^{\hat{\alpha}} \quad \text{where} \quad A_{lt} \stackrel{\text{iid}}{\sim} PS(\hat{\alpha}) \\ \mu_{it} &= \beta_{1,\text{int}}(\mathbf{s}_i) + \beta_{1,\text{time}}(\mathbf{s}_i)t \\ \log(\sigma_{it}) &= \beta_{2,\text{int}}(\mathbf{s}) + \beta_{2,\text{time}}(\mathbf{s})t. \end{aligned} \quad (4.16)$$

We estimate parameters $\Theta = \{A_{lt}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \xi\}$ using Markov chain Monte Carlo methods. We use a Metropolis-Hastings algorithm to update the model parameters with random walk candidate distributions for all parameters. The PS density is challenging to evaluate as it does not have a closed form. One technique to avoid this complication is to incorporate auxiliary random variables [Ste09], but we opt for a numerical approximation to the integral as described in Appendix C.2. The hyperparameters $\mu_{1,\text{int}}, \mu_{1,\text{time}}, \mu_{2,\text{int}}, \mu_{2,\text{time}}$ and $\sigma_{1,\text{int}}^2, \sigma_{1,\text{time}}^2, \sigma_{2,\text{int}}^2, \sigma_{2,\text{time}}^2$ are updated using Gibbs sampling since their prior distributions are conjugate.

The first-stage estimate of the extremal coefficients has three tuning parameters: the quantile thresholds q_1, \dots, q_{n_q} , the kernel bandwidth ϕ , and the number of terms L . In Section 4.5 we explore a few possibilities for L and discuss sensitivity to this choice. The second-stage Bayesian analysis requires selecting thresholds T_i, \dots, T_{n_s} . For this we use spatially smoothed sample quantiles. That is, we set T_i to the 0.95 quantile of the Y_{it} and its five nearest neighbors.

4.5 Data analysis

In this section, we illustrate our method with two data analyses. In Section 4.5.2, we present a points above a threshold analysis using annual acreage burned due to wildfires in Georgia from 1965 – 2014. This is followed in Section 4.5.4 by an analysis of block maxima precipitation data in the eastern U.S. We compare our method with another method that uses standardized Gaussian kernels for the

spatial basis functions.

4.5.1 Gaussian kernel basis functions

To provide a comparison of our model with another approach, we also fit a model that uses standardized Gaussian kernels for the spatial basis functions [RS12]. In this method, Reich & Shaby introduce a set of $\mathbf{k}_1, \dots, \mathbf{k}_L$ spatial knots and use standardized Gaussian kernel functions (GSK; see Appendix C.3) instead of using EBFs for the $\hat{B}_l(\mathbf{s})$. For the comparison between EBF and GSK methods, we use the same number of basis functions. We obtain estimates of the kernel bandwidth $\hat{\rho}$ and spatial dependence $\hat{\alpha}$, using the same least squares minimization as with the EBF method, and treat these as fixed in the MCMC.

4.5.2 Analysis of extreme Georgia fires

The dataset used for our application is composed of yearly acreage burned due to wildfires for each county in Georgia from 1965 – 2014 (<http://weather.gfc.state.ga.us/FireData/>). Figure 4.1 shows the time series of log(acres burned) for 25 randomly selected counties. Based on this plot and other exploratory analysis, we see no evidence of non-linear trends and proceed with linear time trends for the GEV location and scale parameters.

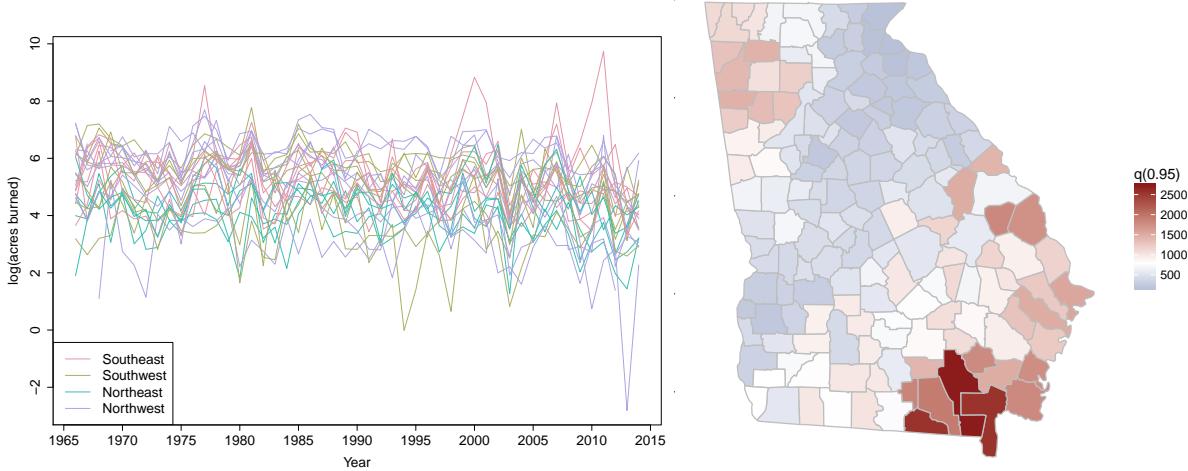


Figure 4.1 Time series of log acres burned for 25 randomly selected counties with colors coding the county's quadrant (left), and spatially smoothed threshold values, T_i for each county (right).

We estimate the extremal coefficient function $\hat{\theta}_{ij}$ by setting $q_1 = 0.90$ and using $n_q = 100$.

With more data, it would be possible to increase q_1 , but we set $q_1 = 0.90$ to increase the stability when estimating $\hat{\eta}_{ij}$. Because these data are not block-maxima, we select a site-specific threshold T_i to use in the analysis with the following algorithm. Without some adjustment to the data, it is challenging to borrow information across sites to inform the threshold selection. We first standardize the data, separately by county, by subtracting the site's median and dividing by the site's interquartile range. Denote the standardized data by \tilde{Y}_i . Then we combine all sites together and plot a mean residual plot for \tilde{Y}_{it} , $i = 1, \dots, n_s$ and $t = 1, \dots, n_t$. The mean residual plot is given in Figure 4.2. Based upon the mean residual plot, we select the 95th percentile for the threshold. To calculate T_i for each county, we use the 95th percentile for the combined data for county i and its five closest counties (see Figure 4.1).

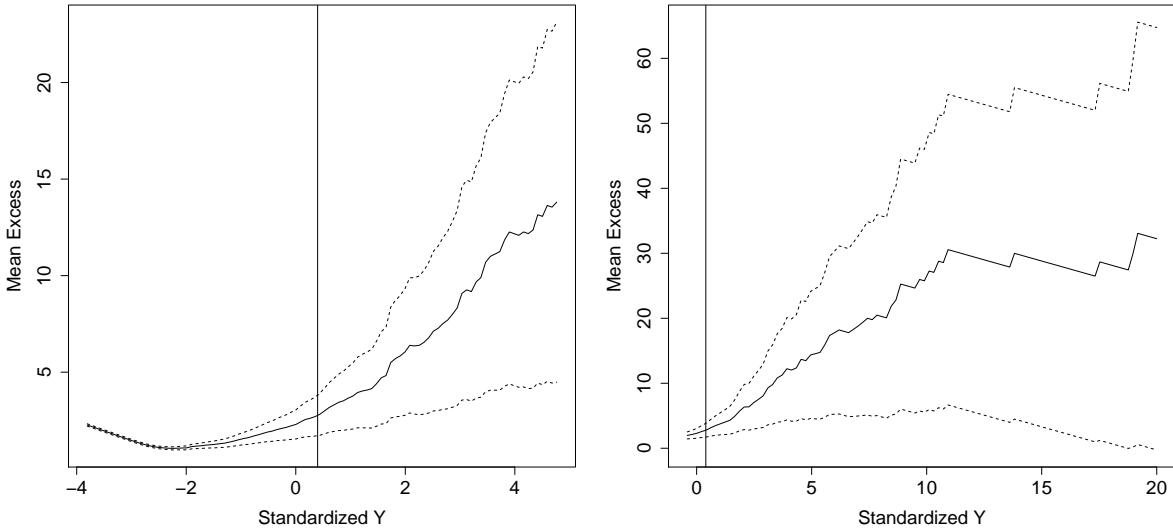


Figure 4.2 Mean residual plot for the data pooled across counties after standardizing using the county's median and interquartile range. The two panels show different ranges on the x-axis and include a vertical line at the sample 95th percentile.

The empirical basis functions for the analysis can be used to help explore spatial dependence in the extremes. The first six EBFs for the wildfire data along with the cumulative sum of the contributions for v_1, \dots, v_{25} are given in Figure 4.3. The first EBF roughly separates the Southeastern Plains (blue) from the coastal region in the southeast and mountains regions in the northeast (red). The remaining EBFs further partition the plains. As a comparison, we provide the first six

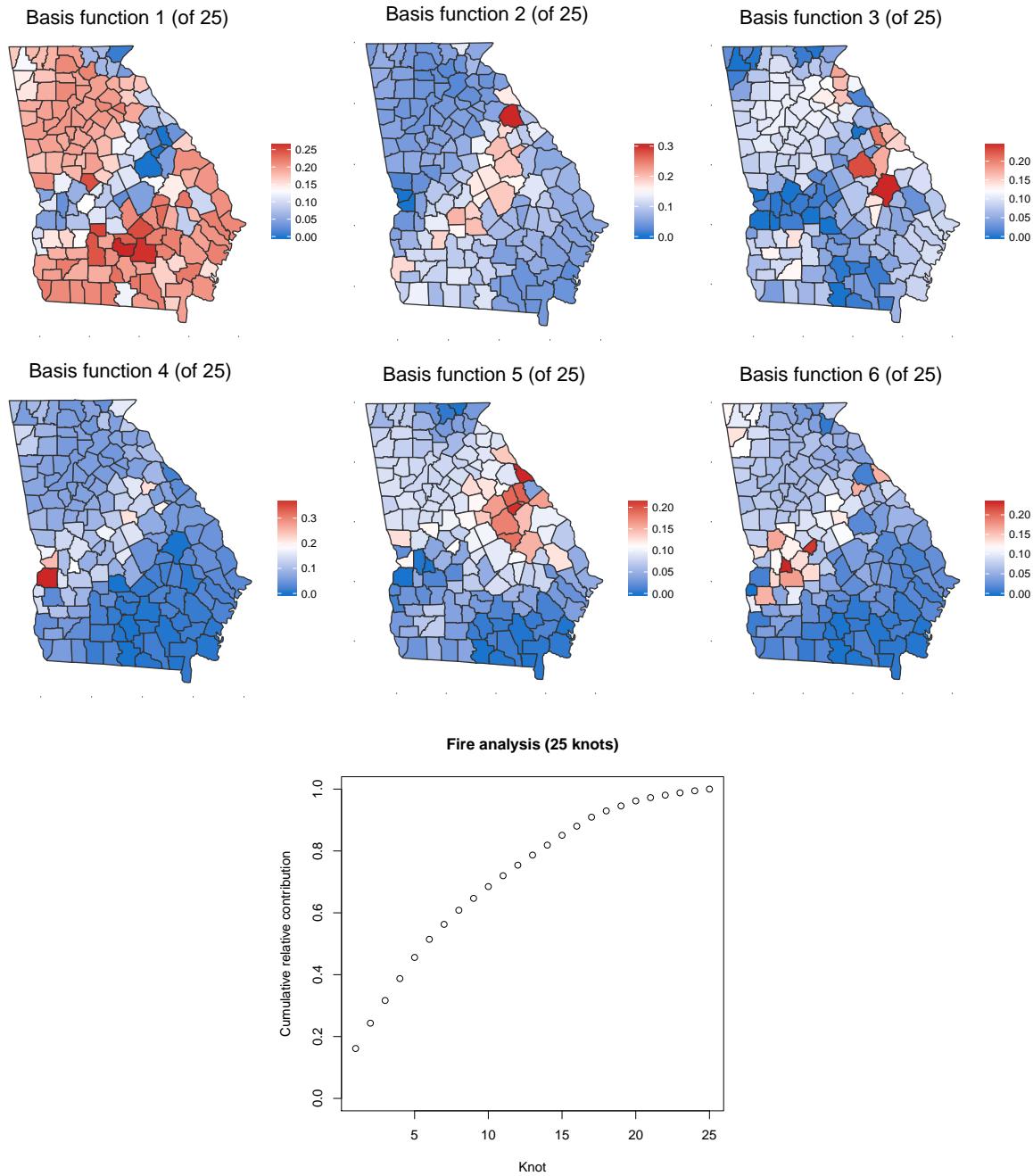


Figure 4.3 First six EBFs for the Georgia fire data and the cumulative sum of contributions v_1, \dots, v_{25} .

principal components of the fire data along with the cumulative sum of the first 25 eigenvalues in Appendix C.4.

Given the basis function estimates, we run the MCMC for 35,000 iterations using a burnin period of 25,000 iterations. We consider models fit with both EBF and GSK, and fit the model using $L = 5, 10, \dots, 40$. Timing for each setting of L for 1,000 iterations is given in Table 4.1. These timings come from a single core of an Intel Core i7-5820K Haswell-E processor, using the OpenBLAS optimized BLAS library (<http://www.openblas.net>).

4.5.3 Results for fire analysis

We use 10-fold cross-validation to assess the predictive performance of a model. For each method, we randomly select 90% of the observations across counties and years to be used as a training set to fit the model. The remaining 10% of sites and years are withheld for testing model predictions. To assess the predictions for the test set, we use quantile scores and Brier scores [GR07]. The quantile score (QS) for quantile level q^* is given by $2\{I[Y(\mathbf{s}) > \hat{q}(\mathbf{s})] - q^*\}\{\hat{q}(\mathbf{s}) - Y(\mathbf{s})\}$ where $\hat{q}(\mathbf{s})$ is the estimated q^* th quantile at site \mathbf{s} , and $I[\cdot]$ is an indicator function. The Brier score (BS) for predicting an exceedance of a level c at site \mathbf{s} is given by $\{I[Y(\mathbf{s}) > c] - \hat{P}[Y(\mathbf{s}) > c]\}^2$. For both of these methods, a lower score indicates a better fit. The Brier and quantile scores for the fire analysis are given in Table 4.1. For the data, the BS and QS are fairly similar for all number of basis functions and the EBF versus GSK. This is perhaps due to the spatial variation in the marginal distribution explaining most of the spatial variation. In fact, using the EBFs, the estimate of residual dependence for the fire data is $\hat{\alpha} = 0.861$ ($\alpha = 1$ is residual independence).

Based on the cross-validation results, we run a full analysis using all of the data with $L = 25$. Despite the GSK model having slightly smaller BS and QS values, we use the EBF method. Figure 4.4 gives posterior summaries for three quantities of interest. We provide maps of the linear time trend in the GEV location ($\beta_{1,\text{time},i}$), and GEV log scale ($\beta_{2,\text{time},i}$) and a map of $\Delta Q90_i = Q90_{i,2014} - Q90_{i,1965}$ the change in the mean of the posterior distribution of $Q90_{i,t}$ between $t_1 = 1965$ and $t_2 = 2014$. We also provide maps of the posterior probability that each of the three terms is positive.

We construct the posterior distribution of the the estimated 90th quantile $Q90_{i,t}$ using the GEV parameters as follows. Let $Q90_{i,t}^{(j)}$ be the estimated 90th quantile at site i for time t at iteration j . We first compute

$$\begin{aligned}\mu_{i,t}^{(j)} &= \beta_{1,\text{int},i}^{(j)} + \beta_{1,\text{time},i}^{(j)} t \\ \log(\sigma_{i,t}^{(j)}) &= \beta_{2,\text{int},i}^{(j)} + \beta_{2,\text{time},i}^{(j)} t.\end{aligned}\tag{4.17}$$

Table 4.1 Average Brier scores ($\times 100$), average quantile scores for $q(0.95)$ and $q(0.99)$, and time (in minutes) for 1,000 iterations for fire analysis.

	Process	Brier Scores ($\times 100$)		Quantile Scores		Time
		$q(0.95)$	$q(0.99)$	$q(0.95)$	$q(0.99)$	
$L = 5$	EBF	4.269	1.653	107.904	67.345	1.16
	GSK	4.244	1.644	103.822	64.046	1.18
$L = 10$	EBF	4.328	1.675	107.075	66.300	1.56
	GSK	4.170	1.665	104.214	64.589	1.47
$L = 15$	EBF	4.330	1.690	108.381	67.671	1.86
	GSK	4.214	1.654	104.490	65.201	1.83
$L = 20$	EBF	4.346	1.697	107.389	66.957	2.18
	GSK	4.174	1.646	104.671	65.430	2.20
$L = 25$	EBF	4.263	1.650	106.656	64.913	2.54
	GSK	4.216	1.661	104.208	64.468	2.55
$L = 30$	EBF	4.328	1.678	106.265	64.674	2.90
	GSK	4.228	1.660	104.143	64.443	2.92
$L = 35$	EBF	4.329	1.671	106.817	65.002	3.29
	GSK	4.256	1.663	105.016	64.920	3.32
$L = 40$	EBF	4.284	1.653	106.621	64.753	3.70
	GSK	4.233	1.666	105.301	64.932	3.56

Let $Q90_{i,t}^{(j)} = \mu_{i,t}^{(j)} + \sigma_{i,t}^{(j)} F^{-1}(0.90, \xi^{(j)})$ where $F^{-1}(q, \xi)$ is the inverse distribution function of the $\text{GEV}(1, 1, \xi)$ distribution evaluated at the q th quantile. Finally, let $Q90_{i,t}$ be posterior mean of $Q90_{i,t}^{(j)}$. To obtain the posterior probability of seeing an increase over time, we take the posterior distributions of each parameter of interest at the two time points. Consider two time points $t_1 < t_2$. Let $\varphi_t^{(j)}$ be the parameter of interest at iteration j and time t . We then take the posterior mean of $I[\varphi_{t_2}^{(j)} > \varphi_{t_1}^{(j)}]$, for the posterior probability of seeing an increase in φ from time t_1 to t_2 .

These results suggest that there has been a slight increase in the amount of acres burned over time. When looking at the variability of the acres burned over time, we see some patterns that generally correspond to landcover in Georgia (see <http://narsal.uga.edu/> for landcover maps). In the high intensity urban areas (e.g. the Piedmont region) and croplands, we see a decrease in variability as well as very low $P[\Delta Q90 > 0]$. However, in the more forested areas (e.g. Blue Ridge mountains, Okefenokee Swamp), we see increased variability over time as well as a high $P[\Delta Q90 > 0]$. This suggests that the 10-year return level for wildfires in regions with larger tree cover is increasing over time which corresponds with increased drought and higher temperatures.

4.5.4 Analysis of annual precipitation

We also conduct an analysis of the precipitation data presented in Reich & Shaby [RS12]. The data are climate model output from the North American Regional Climate Change Assessment Program (NARCCAP). This data consists of $n_s = 697$ grid cells at a 50km resolution in the eastern US, and includes historical data (1969 – 2000) as well as future conditions (2039 – 2070). Because the data are block maxima, we set $T = -\infty$.

For this dataset, to estimate the EBFs, we use the combined current and future data. The first six EBFs for the combined data along with the cumulative sum of the contributions for v_1, \dots, v_{25} are given in Figure 4.6. As a comparison, we provide the first six principal components of the fire data along with the cumulative sum of the first 25 eigenvalues in Appendix C.4. For the precipitation data, we run the MCMC for 25,000 iterations using a burnin period of 15,000 iterations. We consider models fit with both EBF and GSK, and fit the model using $L = 5, 10, \dots, 40$. Timing for each setting of L is given in Table 4.2 for 1,000. The timings are obtained using the same machine as for the fire analysis.

4.5.5 Results for precipitation analysis

We use 5-fold cross-validation to assess the predictive performance of a model. For each method, we randomly select 80% of the observations across counties and years to be used as a training set to fit the model. The remaining 20% of sites and years are withheld for testing model predictions. As

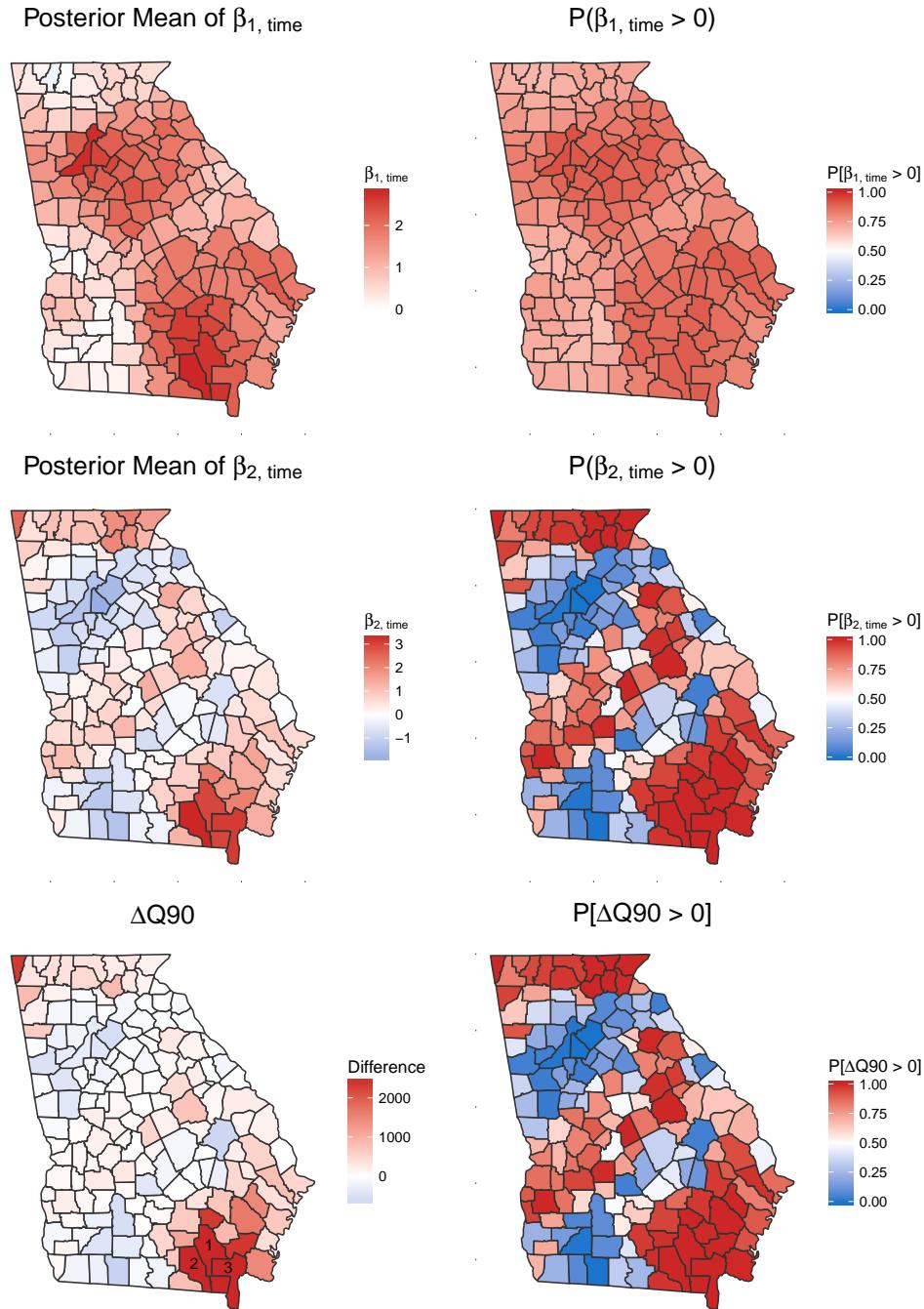


Figure 4.4 Posterior mean of $\beta_{1,\text{time}}$ (top left), posterior mean of $\beta_{2,\text{time}}$ (middle left), estimate of $\Delta Q90$ (bottom left), $P[\beta_{1,\text{time}} > 0]$ (top right), $P[\beta_{2,\text{time}} > 0]$ (middle right), and $P[\Delta Q90 > 0]$ for fire data using EBF. In three counties (labeled), $\Delta Q90 > 2500$: County 1 - Ware (11,109), County 2 - Clinch (7,128), and County 3 - Charlton (6,545)

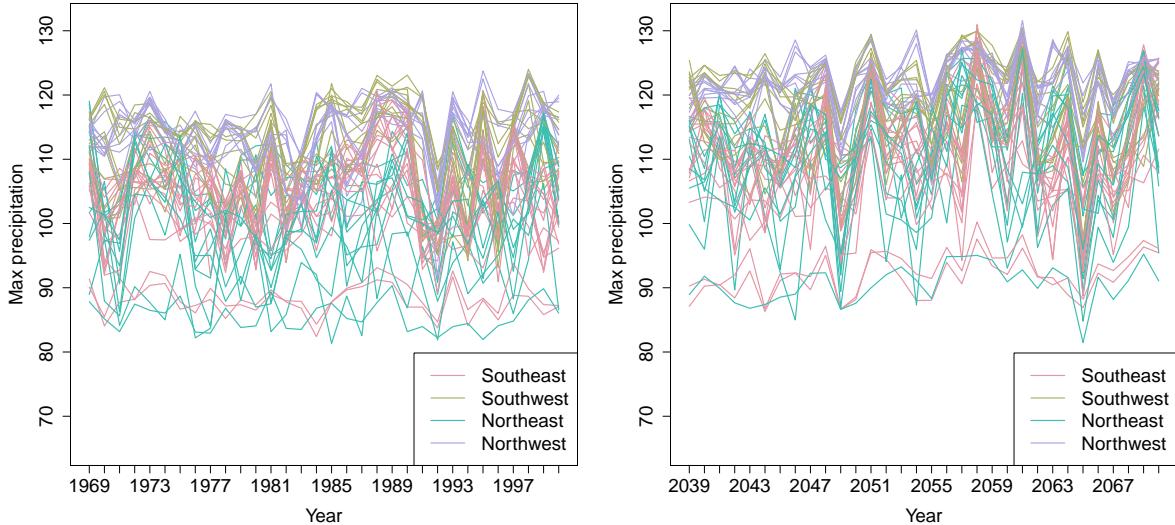


Figure 4.5 Time series of yearly max precipitation for current (1969 – 2000) (left). Time series of yearly max precipitation for future (2039 – 2070) (right).

with the fire analysis, we use Brier scores and quantile scores to compare model performance. The Brier and quantile scores for the current and future precipitation data analysis are given in Table 4.2. For these data, we observe more variation in the scores across the number of basis functions and generally an advantage in using EBF over GSK. This is in contrast to the fire analysis, and is likely due to the fact that the spatial dependence is estimated to be much stronger than for the fire analysis. When using the EBFs, the estimate of residual dependence for the precipitation data is $\hat{\alpha} = 0.280$ ($\alpha = 1$ is residual independence).

Based on the cross-validation results, we run a full analysis using all of the data with $L = 25$ and EBF. Figure 4.7 gives posterior summaries for three quantities of interest. Because we have two separate time periods, current and future, we look at the differences between the estimates for $\hat{\mu}$, $\log(\hat{\sigma})$, and $\hat{q}(0.90)$ between $t_1 = 2000$ and $t_2 = 2070$ where $\hat{\mu}$, $\log(\hat{\sigma})$, and $\hat{q}(0.90)$ are calculated as described in Section 4.5.3. We plot $\Delta\mu = \hat{\mu}_{2070} - \hat{\mu}_{2000}$, $\Delta\log(\hat{\sigma}) = \log(\hat{\sigma})_{2070} - \log(\hat{\sigma})_{2000}$, and $\Delta Q90 = \hat{q}(0.90)_{2070} - \hat{q}(0.90)_{2000}$, and the estimated probabilities that each are positive.

The results seem to suggest that the strength of extreme rain events will increase between 2000 and 2070 as well as greater variability in the northeast region of the U.S. as well as Ohio and parts of the south. There is very strong evidence to suggest that most of the eastern U.S. should expect to see an increase in the 10-year return level between 2000 and 2070. Exceptions to this trend appear in

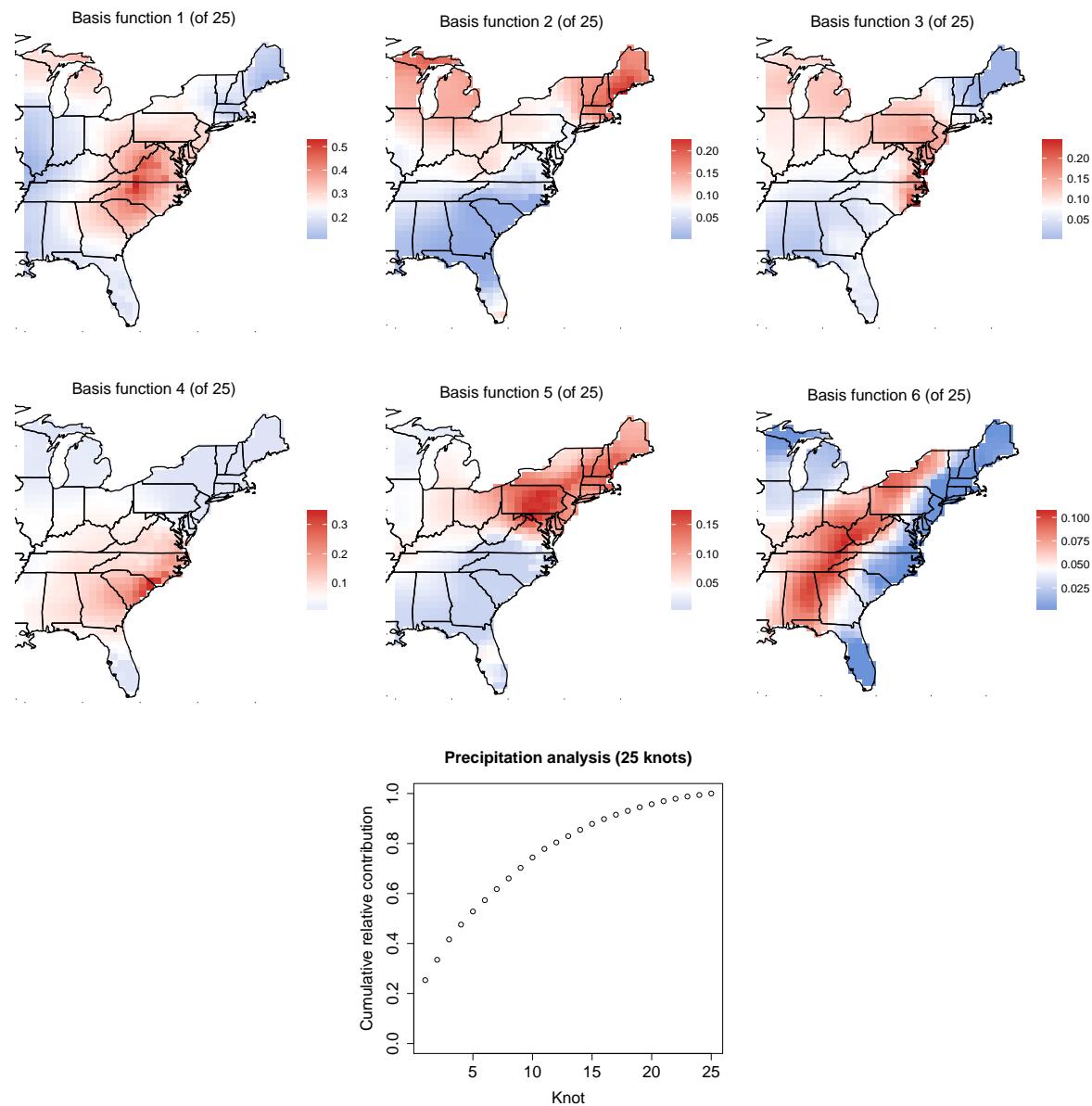


Figure 4.6 First six EBFs for the combined precipitation data and the cumulative sum of contributions v_1, \dots, v_{25} .

southern parts of Alabama and Mississippi and regions on the border between South Carolina and Georgia which will likely experience a decrease in the 10-year return level.

4.6 Discussion

In this paper we have proposed new empirical basis functions for a data-driven low-rank approximation to a max-stable process. The basis functions provide researchers with an exploratory data analysis tool to explore maps of extremal dependence over space. The functions can also be used as inputs to an MCMC algorithm for inference and predictions over space. The results from the data analysis provide evidence to suggest that in the presence of strong spatial dependence as with the precipitation data, the empirical basis functions show an improvement in quantile scores over using knots and standardized Gaussian kernel functions without an increase in the amount of time for computing.

We have used the EBF for exploratory analysis and Bayesian inference. Another possibility is to use the methods to reduce the data under consideration from the actual responses to loadings A_{kt} . That is, given the EBF, one could obtain estimates of the A_{kt} using a separate maximum likelihood estimation for each time point. Time series of the estimated A_{kt} may be used as a fast and simple method to study large-scale spatiotemporal trends.

Table 4.2 Average Brier scores ($\times 100$), average quantile scores for $q(0.95)$ and $q(0.99)$, and time (in minutes) for 1,000 iterations for precipitation analysis.

L	Process	Current				Future				
		BS ($\times 100$)		QS		Time	BS ($\times 100$)		QS	
		$q(0.95)$	$q(0.99)$	$q(0.95)$	$q(0.99)$		$q(0.95)$	$q(0.99)$	$q(0.95)$	$q(0.99)$
5	EBF	3.813	1.098	0.740	0.203	5.80	3.357	1.112	0.738	0.209
	GSK	3.854	1.074	0.745	0.204	5.54	3.338	1.101	0.742	0.210
10	EBF	3.680	1.063	0.698	0.192	6.54	3.148	1.067	0.687	0.198
	GSK	3.628	1.068	0.705	0.195	6.30	3.088	1.072	0.709	0.201
15	EBF	3.505	1.065	0.668	0.187	7.27	3.101	1.095	0.661	0.189
	GSK	3.618	1.051	0.695	0.194	7.05	3.057	1.064	0.697	0.199
20	EBF	3.400	1.026	0.651	0.182	8.14	3.101	1.087	0.649	0.189
	GSK	3.552	1.039	0.688	0.192	7.87	3.065	1.062	0.692	0.196
25	EBF	3.463	1.058	0.650	0.182	9.00	3.003	1.113	0.637	0.185
	GSK	3.643	1.063	0.679	0.190	8.74	2.039	1.054	0.686	0.196
30	EBF	3.455	1.038	0.646	0.182	9.89	2.956	1.073	0.630	0.182
	GSK	3.550	1.046	0.684	0.191	9/60	3.074	1.067	0.685	0.195
35	EBF	3.471	1.065	0.649	0.183	10.79	3.036	1.114	0.645	0.189
	GSK	3.608	1.064	0.686	0.194	10.46	3.083	1.075	0.691	0.199
40	EBF	3.551	1.114	0.652	0.184	11.70	3.050	1.131	0.645	0.189
	GSK	3.605	1.067	0.686	0.194	11.32	3.104	1.083	0.693	0.199

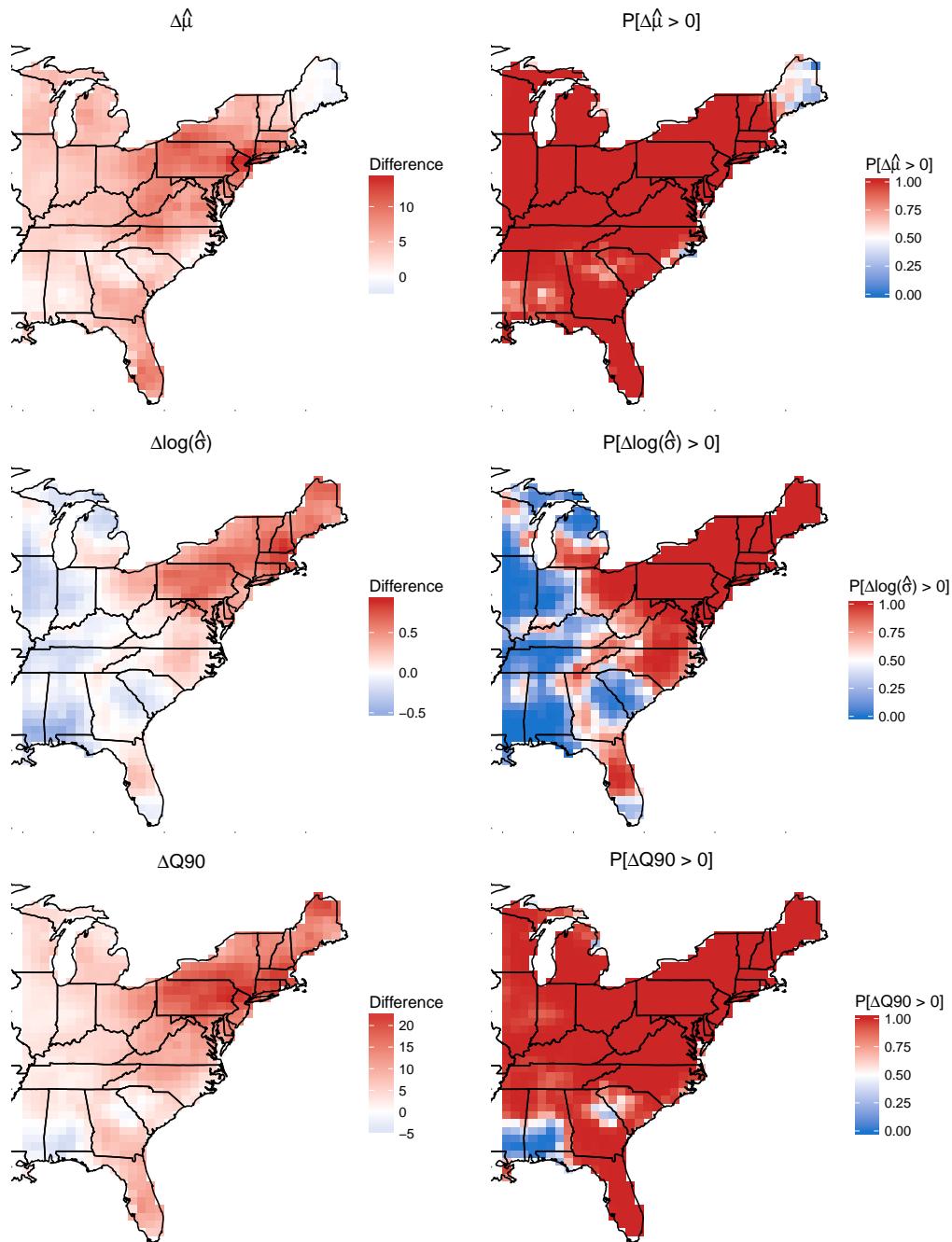


Figure 4.7 Posterior mean of $\Delta\mu$ (top left), posterior mean of $\Delta\log(\sigma)$ (middle left), estimate of $\Delta Q90$ (bottom left), $P[\Delta\mu > 0]$ (top right), $P[\Delta\log(\sigma) > 0]$ (middle right), and $P[\Delta Q90 > 0]$ between 2000 and 2070 for precipitation data using EBF.

BIBLIOGRAPHY

- [AC14] Azzalini, A. & Capitanio, A. *The Skew-Normal and Related Families*. Institute of Mathematical Statistics Monographs. Cambridge University Press, 2014.
- [AD96] Azzalini, A. & Dalla Valle, A. “The multivariate skew-normal distribution”. *Biometrika* **83**.4 (1996), pp. 715–726.
- [AC03] Azzalini, A. & Capitanio, A. “Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution”. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **65**.2 (2003), pp. 367–389.
- [Ber16] Beranger, B. et al. “Models for extremal dependence derived from skew-symmetric families”. *ArXiv e-prints* (2016). arXiv: 1507.00108.
- [BD01] Branco, M. D. & Dey, D. K. “A General Class of Multivariate Skew-Elliptical Distributions”. *Journal of Multivariate Analysis* **79**.1 (2001), pp. 99–113.
- [Coh60] Cohen, J. “A Coefficient of Agreement for Nominal Scales”. *Educational and Psychological Measurement* **20**.1 (1960), pp. 37–46.
- [Col01] Coles, S. *An Introduction to Statistical Modeling of Extreme Values*. Lecture Notes in Control and Information Sciences. London: Springer, 2001.
- [Col99] Coles, S. et al. “Dependence Measures for Extreme Value Analyses”. *Extremes* **2**.4 (1999), pp. 339–365.
- [Coo06] Cooley, D. et al. “Variograms for spatial max-stable random fields”. *Dependence in Probability and Statistics*. Ed. by Bertail, P. et al. Vol. 187. Lecture Notes in Statistics. New York, NY: Springer New York, 2006. Chap. Variograms, pp. 373–390.
- [Dav12] Davison, A. C. et al. “Statistical Modeling of Spatial Extremes”. *Statistical Science* **27**.2 (2012), pp. 161–186.
- [DS90] Davison, A. C. & Smith, R. L. “Models for exceedances over high thresholds (with Discussion)”. *Journal of the Royal Statistical Society. Series B (Methodological)* **52**.3 (1990), pp. 393–442.
- [Eng15] Engelke, S. et al. “Estimation of Hüsler-Reiss distributions and Brown-Resnick processes”. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **77** (2015), pp. 239–265. arXiv: 1207.6886v2.
- [EH08] Everitt, B. & Hothorn, T. “Principal Components Analysis”. *An introduction to applied multivariate analysis with R*. New York, NY: Springer New York, 2008, pp. 21–54. arXiv: 1011.1669v3.

- [Fin15] Finley, A. O. et al. “spBayes for Large Univariate and Multivariate Point-Referenced Spatio-Temporal Data Models”. *Journal of Statistical Software* **63**.13 (2015).
- [Gen04] Genton, M. G. *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*. Statistics (Chapman & Hall/CRC). Taylor & Francis, 2004.
- [Gen11] Genton, M. G. et al. “On the likelihood function of Gaussian max-stable processes”. *Biometrika* **98**.2 (2011), pp. 481–488.
- [GR07] Gneiting, T. & Raftery, A. E. “Strictly Proper Scoring Rules, Prediction, and Estimation”. *Journal of the American Statistical Association* **102**.477 (2007), pp. 359–378.
- [Haa84] Haan, L de. “A Spectral Representation for Max-stable Processes”. *The Annals of Probability* **12**.4 (1984), pp. 1194–1204.
- [HF06] Haan, L de & Ferreira, A. *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering. Springer, 2006.
- [Han07] Hannachi, A. et al. “Empirical orthogonal functions and related techniques in atmospheric science: A review”. *International Journal of Climatology* **27**.9 (2007), pp. 1119–1152. arXiv: 10.1002/joc.1492 [10.1002].
- [HL98] Heagerty, P. & Lele, S. “A Composite Likelihood Approach to Binary Spatial Data”. *Journal of the American Statistical Association* **145**.September (1998), pp. 1099–1111.
- [Hol14] Hollander, M. et al. *Nonparametric Statistical Methods*. 3rd. Wiley, 2014, pp. 316–322.
- [HD14] Huser, R. & Davison, A. C. “Space-time modelling of extreme events”. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **76**.2 (2014), pp. 439–461. arXiv: 1201.3245.
- [Kab09] Kabluchko, Z. et al. “Stationary max-stable fields associated to negative definite functions”. *Annals of Probability* **37**.5 (2009), pp. 2042–2065. arXiv: 0806.2780v3.
- [KM04] Kim, H. M. & Mallick, B. K. “A Bayesian prediction using the skew Gaussian distribution”. *Journal of Statistical Planning and Inference* **120**.1-2 (2004), pp. 85–101.
- [Kim05] Kim, H.-M. et al. “Analyzing Nonstationary Spatial Data Using Piecewise Gaussian Processes”. *Journal of the American Statistical Association* **100**.470 (2005), pp. 653–668.
- [LT96] Ledford, A. W. & Tawn, J. A. “Statistics for near independence in multivariate extreme values.” *Biometrika* **83** (1996), pp. 169–187.
- [LS99] Lee, D. D. & Seung, S. H. “Learning the parts of objects by non-negative matrix factorizations”. *Nature* **401** (1999), pp. 788–791.

- [Mai14] Mairal, J. et al. “Sparse modeling for image and vision processing”. *Foundations and Trends in Computer Graphics and Vision* **8** (2014), pp. 85–283.
- [Morde] Morris, S. A. et al. “A space-time skew- t model for threshold exceedances”. *Biometrics* (under review).
- [Pac16] Pacifici, K. et al. “Occupancy estimation for rare species using a spatially-adaptive sampling design”. *Methods in Ecology and Evolution* **7**.3 (2016). Ed. by McPherson, J., pp. 285–293.
- [Pad10] Padoan, S. A. et al. “Likelihood-Based Inference for Max-Stable Processes”. *Journal of the American Statistical Association* **105**.489 (2010), pp. 263–277.
- [Pad11] Padoan, S. A. “Multivariate extreme models based on underlying skew- and skew-normal distributions”. *Journal of Multivariate Analysis* **102**.5 (2011), pp. 977–991.
- [R C16] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2016.
- [RS12] Reich, B. J. & Shaby, B. A. “A hierarchical max-stable spatial model for extreme precipitation”. *The Annals of Applied Statistics* **6**.4 (2012), pp. 1430–1451.
- [Rib15] Ribatet, M. *SpatialExtremes: Modelling Spatial Extremes*. R package version 2.0-2. 2015.
- [Sam00] Samet, J. M. et al. *The National Morbidity, Mortality and Air Pollution Study Part I: Methods and Methodologic Issues*. Tech. rep. 94. 2000.
- [ST03] Schlather, M. & Tawn, J. A. “A dependence measure for multivariate and spatial extreme values: Properties and inference”. *Biometrika* **90**.1 (2003), pp. 139–156.
- [Sib60] Sibuya, M. “Bivariate extreme statistics”. *Annals of the Institute of Statistical Mathematics* **11** (1960), pp. 195–210.
- [Sin05] Sing, T. et al. “ROCR: visualizing classifier performance in R”. *Bioinformatics* **21**.20 (2005), pp. 3940–3941.
- [Smi12] Smith, D. R. et al. “Incorporating predicted species distribution in adaptive and conventional sampling designs”. *Design and Analysis of Long-term Ecological Monitoring Studies*. Ed. by Gitzen, R. A. et al. Cambridge University Press, 2012. Chap. 17, pp. 381–396.
- [Smi90] Smith, R. L. *Max-stable processes and spatial extremes*. 1990.
- [Ste09] Stephenson, A. G. “High-Dimensional Parametric Modelling of Multivariate Extreme Events”. *Australian & New Zealand Journal of Statistics* **51**.1 (2009), pp. 77–88.

- [Taw90] Tawn, J. A. "Modelling multivariate extreme value distributions". *Biometrika* **77**.2 (1990), pp. 245–253.
- [Thi13] Thibaud, E. et al. "Threshold modeling of extreme spatial rainfall". *Water Resources Research* **49**.8 (2013), pp. 4633–4644.
- [TO15] Thibaud, E. & Opitz, T. "Efficient inference and simulation for elliptical Pareto processes". *Biometrika* **102**.4 (2015), pp. 855–870. arXiv: 1401.0168v1.
- [WT14] Wadsworth, J. L. & Tawn, J. A. "Efficient inference for spatial extreme value processes associated to log-Gaussian random functions". *Biometrika* **101**.1 (2014), pp. 1–15.
- [WT12] Wadsworth, J. L. & Tawn, J. A. "Dependence modelling for spatial extremes". *Biometrika* **99**.2 (2012), pp. 253–272.
- [WD10] Wang, X. & Dey, D. K. "Generalized extreme value regression for binary response data: An application to B2B electronic payments system adoption". *The Annals of Applied Statistics* **4**.4 (2010), pp. 2000–2023.
- [WS11] Wang, Y. & Stoev, S. A. "Conditional sampling for spectrally discrete max-stable random fields". *Advances in Applied Probability* **43**.2 (2011), pp. 461–483. arXiv: 1005.0312v2.
- [ZES10] Zhang, H. & El-Shaarawi, A. "On spatial skew-Gaussian processes and applications". *Environmetrics* **21**.October 2008 (2010), pp. 33–47.

APPENDICES

APPENDIX

A

A SPACE-TIME SKEW- T MODEL FOR THRESHOLD EXCEEDANCES

A.1 MCMC details

The MCMC sampling for the model in Section 2.4 is done using R (<http://www.r-project.org>). Whenever possible, we select conjugate priors (see Appendix A.2); however, for some of the parameters, no conjugate prior distributions exist. For these parameters, we use a random walk Metropolis-Hastings update step. In each Metropolis-Hastings update, we tune the algorithm during the burn-in period to give acceptance rates near 0.40.

Spatial knot locations

For each day, we update the spatial knot locations, $\mathbf{w}_1, \dots, \mathbf{w}_K$, using a Metropolis-Hastings block update. Because the spatial domain is bounded, we generate candidate knots using the transformed knots $\mathbf{w}_1^*, \dots, \mathbf{w}_K^*$ (see Section 2.3.3) and a random walk bivariate Gaussian candidate distribution

$$\mathbf{w}_k^{*(c)} \sim N(\mathbf{w}_k^{*(r-1)}, s^2 I_2)$$

where $\mathbf{w}_k^{*(r-1)}$ is the location for the transformed knot at MCMC iteration $r-1$, s is a tuning parameter, and I_2 is an identity matrix. Let $\mathbf{Y}_t = [Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n)]$ be the vector of observed responses at each site for day t . After candidates have been generated for all K knots, the acceptance ratio is

$$R = \left\{ \frac{l[\mathbf{Y}_t | \mathbf{w}_1^{(c)}, \dots, \mathbf{w}_K^{(c)}, \dots]}{l[\mathbf{Y}_t | \mathbf{w}_1^{(r-1)}, \dots, \mathbf{w}_K^{(r-1)}, \dots]} \right\} \times \left\{ \frac{\prod_{k=1}^K \phi(\mathbf{w}_k^{(c)})}{\prod_{k=1}^K \phi(\mathbf{w}_k^{(r-1)})} \right\} \times \left\{ \frac{\prod_{k=1}^K p(\mathbf{w}_k^{*(c)})}{\prod_{k=1}^K p(\mathbf{w}_k^{*(r-1)})} \right\}$$

where l is the likelihood given in (2.16), and $p(\cdot)$ is the prior either taken from the time series (see Section 2.3.3) or assumed to be uniform over \mathcal{D} . The candidate knots are accepted with probability $\min\{R, 1\}$.

Spatial random effects

If there is no temporal dependence amongst the observations, we use a Gibbs update for z_{tk} , and the posterior distribution is given in Appendix A.2. If there is temporal dependence amongst the observations, then we update z_{tk} using a Metropolis-Hastings update. Because this model uses $|z_{tk}|$, we generate candidate random effects using the z_{tk}^* (see Section 2.3.3) and a random walk Gaussian candidate distribution

$$z_{tk}^{*(c)} \sim N(z_{tk}^{*(r-1)}, s^2)$$

where $z_{tk}^{*(r-1)}$ is the value at MCMC iteration $r-1$, and s is a tuning parameter. The acceptance ratio is

$$R = \left\{ \frac{l[\mathbf{Y}_t | z_{tk}^{(c)}, \dots]}{l[\mathbf{Y}_t | z_{tk}^{(r-1)}]} \right\} \times \left\{ \frac{p[z_{tk}^{(c)}]}{p[z_{tk}^{(r-1)}]} \right\}$$

where $p[\cdot]$ is the prior taken from the time series given in Section 2.3.3. The candidate is accepted with probability $\min\{R, 1\}$.

Variance terms

When there is more than one site in a partition, then we update σ_{tk}^2 using a Metropolis-Hastings update. First, we generate a candidate for σ_{tk}^2 using an $IG(a^*/s, b^*/s)$ candidate distribution in an independence Metropolis-Hastings update where $a^* = (n_{tk} + 1)/2 + a$, $b^* = [\mathbf{Y}_{tk}^\top \Sigma_{tk}^{-1} \mathbf{Y}_{tk} + z_{tk}^2]/2 + b$, n_{tk} is the number of sites in partition k on day t , and \mathbf{Y}_{tk} and Σ_{tk}^{-1} are the observations and precision

matrix for partition k on day t . The acceptance ratio is

$$R = \left\{ \frac{l[\mathbf{Y}_t | \sigma_{tk}^{(c)}, \dots]}{l[\mathbf{Y}_t | \sigma_{tk}^{(r-1)}]} \right\} \times \left\{ \frac{l[z_{tk} | \sigma_{tk}^{(c)}, \dots]}{l[z_{tk} | \sigma_{tk}^{(r-1)}, \dots]} \right\} \times \left\{ \frac{p[\sigma_{tk}^{(c)}]}{p[\sigma_{tk}^{(r-1)}]} \right\} \times \left\{ \frac{c[\sigma_{tk}^{(r-1)}]}{c[\sigma_{tk}^{(c)}]} \right\}$$

where $p[\cdot]$ is the prior either taken from the time series given in Section 2.3.3 or assumed to be $\text{IG}(a, b)$, and $c[\cdot]$ is the candidate distribution. The candidate is accepted with probability $\min\{R, 1\}$.

Spatial covariance parameters

We update the three spatial covariance parameters, $\log(\rho)$, $\log(\nu)$, γ , using a Metropolis-Hastings block update step. First, we generate a candidate using a random walk Gaussian candidate distribution

$$\log(\rho)^{(c)} \sim N(\log(\rho)^{(r-1)}, s^2)$$

where $\log(\rho)^{(r-1)}$ is the value at MCMC iteration $r - 1$, and s is a tuning parameter. Candidates are generated for $\log(\nu)$ and γ in a similar fashion. The acceptance ratio is

$$R = \left\{ \frac{\prod_{t=1}^{n_t} l[Y_t(\mathbf{s}) | \rho^{(c)}, \nu^{(c)}, \gamma^{(c)}, \dots]}{\prod_{t=1}^{n_t} l[Y_t(\mathbf{s}) | \rho^{(r-1)}, \nu^{(r-1)}, \gamma^{(r-1)}, \dots]} \right\} \times \left\{ \frac{p[\rho^{(c)}]}{p[\rho^{(r-1)}]} \right\} \times \left\{ \frac{p[\nu^{(c)}]}{p[\nu^{(r-1)}]} \right\} \times \left\{ \frac{p[\gamma^{(c)}]}{p[\gamma^{(r-1)}]} \right\}.$$

All three candidates are accepted with probability $\min\{R, 1\}$.

A.2 Posterior distributions

Conditional posterior of $z_{tk} | \dots$

If knots are independent over days, then the conditional posterior distribution of $|z_{tk}|$ is conjugate. For simplicity, drop the subscript t , let $\tilde{z}_k = |z_k|$, $\tilde{\mathbf{z}}_{k^c}$ be the vector of $[|z(\mathbf{s}_1)|, \dots, |z(\mathbf{s}_n)|]$ for $\mathbf{s} \notin P_k$, $\mathbf{X} = [\mathbf{X}(\mathbf{s}_1), \dots, \mathbf{X}(\mathbf{s}_n)]^\top$, let \mathbf{Y}_k and \mathbf{X}_k be the observations and covariate measurements for $\mathbf{s} \in P_k$, and let \mathbf{Y}_{k^c} and \mathbf{X}_{k^c} be the observations and covariate measurements for $\mathbf{s} \notin P_k$ and define

$$\mathbf{R} = \begin{cases} \mathbf{Y}_k - \mathbf{X}_k \boldsymbol{\beta} & \mathbf{s} \in P_k \\ \mathbf{Y}_{k^c} - \mathbf{X}_{k^c} \boldsymbol{\beta} - \lambda \tilde{\mathbf{z}}_{k^c} & \mathbf{s} \notin P_k \end{cases}$$

Let

$$\begin{aligned}\mathbf{R}_1 &= \text{the vector of } \mathbf{R} \text{ for } \mathbf{s} \in P_k \\ \mathbf{R}_2 &= \text{the vector of } \mathbf{R} \text{ for } \mathbf{s} \notin P_k \\ \Omega &= \Sigma^{-1}.\end{aligned}$$

Then

$$\begin{aligned}\pi(z_k | \dots) &\propto \exp \left\{ -\frac{1}{2} \left[\left(\begin{array}{c} \mathbf{R}_1 - \lambda \tilde{z}_k \mathbf{1} \\ \mathbf{R}_2 \end{array} \right)^T \left(\begin{array}{cc} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{array} \right) \left(\begin{array}{c} \mathbf{R}_1 - \lambda \tilde{z}_k \mathbf{1} \\ \mathbf{R}_2 \end{array} \right) + \frac{\tilde{z}_k^2}{\sigma_k^2} \right] \right\} I(z_k > 0) \\ &\propto \exp \left\{ -\frac{1}{2} [\Lambda_k \tilde{z}_k^2 - 2\mu_k \tilde{z}_k] \right\}\end{aligned}$$

where

$$\begin{aligned}\mu_k &= \lambda (\mathbf{R}_1^T \Omega_{11} + \mathbf{R}_2^T \Omega_{21}) \mathbf{1} \\ \Lambda_k &= \lambda^2 \mathbf{1}^T \Omega_{11} \mathbf{1} + \frac{1}{\sigma_k^2}.\end{aligned}$$

Then $\tilde{z}_k | \dots \sim N_{(0, \infty)}(\Lambda_k^{-1} \mu_k, \Lambda_k^{-1})$

Conditional posterior of $\boldsymbol{\beta}, \lambda | \dots$

For models that do not include a skewness parameter, we update $\boldsymbol{\beta}$ as follows. Let $\boldsymbol{\beta} \sim N_p(0, \Lambda_0)$ where Λ_0 is a precision matrix. Then

$$\begin{aligned}\pi(\boldsymbol{\beta} | \dots) &\propto \exp \left\{ -\frac{1}{2} \boldsymbol{\beta}^T \Lambda_0 \boldsymbol{\beta} - \frac{1}{2} \sum_{t=1}^{n_t} [\mathbf{Y}_t - \mathbf{X}_t \boldsymbol{\beta}]^T \Omega [\mathbf{Y}_t - \mathbf{X}_t \boldsymbol{\beta}] \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left[\boldsymbol{\beta}^T \Lambda_\beta \boldsymbol{\beta} - 2 \sum_{t=1}^{n_t} (\boldsymbol{\beta}^T \mathbf{X}_t^T \Omega \mathbf{Y}_t) \right] \right\} \\ &\propto N_p(\Lambda_\beta^{-1} \boldsymbol{\mu}_\beta, \Lambda_\beta^{-1})\end{aligned}$$

where

$$\boldsymbol{\mu}_\beta = \sum_{t=1}^{n_t} \mathbf{X}_t^\top \Omega \mathbf{Y}_t$$

$$\Lambda_\beta = \Lambda_0 + \sum_{t=1}^{n_t} \mathbf{X}_t^\top \Omega \mathbf{X}_t.$$

For models that do include a skewness parameter, a simple augmentation of the covariate matrix \mathbf{X} and parameter vector $\boldsymbol{\beta}$ allows for a block update of both $\boldsymbol{\beta}$ and λ . Let $\mathbf{X}_t^* = [\mathbf{X}_t, |\mathbf{z}_t|]$ where $\mathbf{z}_t = [z(\mathbf{s}_1), \dots, z(\mathbf{s}_n)]^\top$ and let $\boldsymbol{\beta}^* = (\beta_1, \dots, \beta_p, \lambda)^\top$. So to incorporate the $N(0, \sigma_\lambda^2)$ prior on λ , let $\boldsymbol{\beta}^* \sim N_{p+1}(0, \Lambda_0^*)$ where

$$\Lambda_0^* = \begin{pmatrix} \Lambda_0 & 0 \\ 0 & \sigma_\lambda^{-2} \end{pmatrix}.$$

Then the update for both $\boldsymbol{\beta}$ and λ is done using the conjugate prior given above with $\mathbf{X}_t = \mathbf{X}_t^*$ and $\boldsymbol{\beta} = \boldsymbol{\beta}^*$

Conditional posterior of $\sigma^2 | \dots$

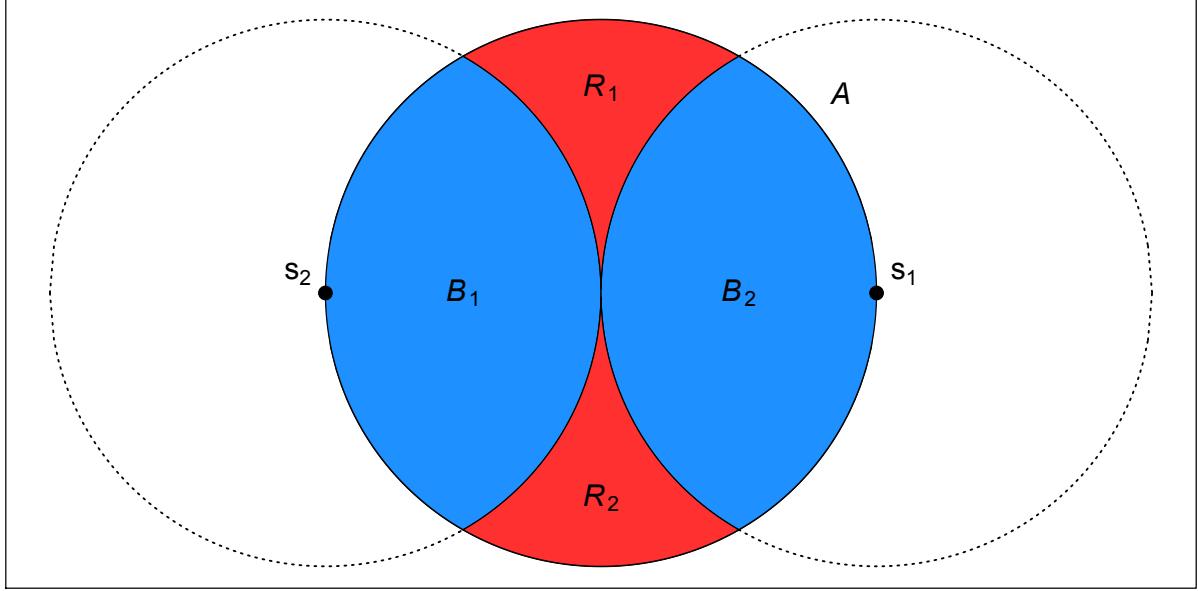
In the case where $L = 1$ and temporal dependence is negligible, then σ^2 has a conjugate posterior distribution. Let $\sigma_t^2 \stackrel{\text{iid}}{\sim} \text{IG}(\alpha_0/2, \beta_0/2)$. For simplicity, drop the subscript t . Then

$$\begin{aligned} \pi(\sigma^2 | \dots) &\propto (\sigma^2)^{-\alpha_0/2-1/2-n/2-1} \exp \left\{ -\frac{\beta_0}{2\sigma^2} - \frac{|z|^2}{2\sigma^2} - \frac{(\mathbf{Y}-\boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{Y}-\boldsymbol{\mu})}{2\sigma^2} \right\} \\ &\propto (\sigma^2)^{-(\alpha_0-1-n)/2-1} \exp \left\{ -\frac{1}{2\sigma^2} [\beta_0 + |z|^2 + (\mathbf{Y}-\boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{Y}-\boldsymbol{\mu})] \right\} \\ &\propto \text{IG}(\alpha^*, \beta^*) \end{aligned}$$

where

$$\begin{aligned} \alpha^* &= \frac{\alpha_0 + 1 + n}{2} \\ \beta^* &= \frac{1}{2} [\beta_0 + |z|^2 + (\mathbf{Y}-\boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{Y}-\boldsymbol{\mu})]. \end{aligned}$$

In the case that $K > 1$, a random walk Metropolis Hastings step will be used to update σ_{kt}^2 .

**Figure A.1** Illustration of the partition of \$A\$.**A.3 Proof that $\lim_{h \rightarrow \infty} \pi(h) = 0$**

Let \$c\$ be the midpoint of \$\mathbf{s}_1\$ and \$\mathbf{s}_2\$. Define \$A\$ as the circle centered at \$c\$ with radius \$h/2\$ where \$h = \|\mathbf{s}_1 - \mathbf{s}_2\|\$ is the distance between sites \$\mathbf{s}_1\$ and \$\mathbf{s}_2\$. Consider a homogeneous spatial Poisson process over \$A\$ with intensity \$\lambda_{\text{PP}}\$, so that

$$\mu_{\text{PP}}(A) = \lambda_{\text{PP}} |A| = \lambda_{\text{PP}} \pi \left(\frac{h}{2} \right)^2 = \lambda_{\text{PP}}^* h^2.$$

Consider a partition of \$A\$ into four regions, \$B_1, B_2, R_1, R_2\$ as seen in Figure A.1. Let \$N_i\$ be the number of knots in \$B_i\$ and \$L_i = l\$ if \$\mathbf{s}_i \in P_l\$ for \$i = 1, 2\$. Then

$$P(L_1 \neq L_2) \geq P(N_1 > 0, N_2 > 0) \quad (\text{A.1})$$

since knots in both \$B_1\$ and \$B_2\$ is sufficient, but not necessary, to ensure that \$\mathbf{s}_1\$ and \$\mathbf{s}_2\$ are in different partition sets. By definition of a Poisson process, \$N_1\$ and \$N_2\$ are independent and thus

$P(N_1 > 0, N_2 > 0) = P(N_1 > 0)^2$, and

$$\begin{aligned}\mu_{\text{PP}}(B_1) &= \lambda_{\text{PP}}|B_1| = \lambda_{\text{PP}} \frac{h^2}{4} \left(\frac{2\pi}{3} - \frac{\sqrt{3}}{2} \right) \\ &= \lambda_{\text{PPB1}}^* h^2.\end{aligned}\quad (\text{A.2})$$

So,

$$P(L_1 \neq L_2) \geq P(N_1 > 0)^2 = [1 - P(N_1 = 0)]^2 = [1 - \exp(-\lambda_{\text{PPB1}}^* h^2)]^2 \quad (\text{A.3})$$

which goes to 1 as h goes to infinity.

A.4 Skew- t distribution

Univariate skew- t distribution

We say that Y follows a univariate extended skew- t distribution with location $\xi \in \mathbb{R}$, scale $\omega > 0$, skew parameter $\alpha \in \mathbb{R}$, and degrees of freedom ν if has distribution function

$$f_{\text{EST}}(y) = 2f_T(z; \nu)F_T\left[\alpha z \sqrt{\frac{\nu+1}{\nu+z^2}}; \nu+1\right] \quad (\text{A.4})$$

where $f_T(t; \nu)$ is a univariate Student's t with ν degrees of freedom, $F_T(t; \nu) = P(T < t)$, and $z = (y - \xi)/\omega$.

Multivariate skew- t distribution

If $\mathbf{Z} \sim \text{ST}_d(0, \bar{\boldsymbol{\Omega}}, \boldsymbol{\alpha}, \eta)$ is a d -dimensional skew- t distribution, and $\mathbf{Y} = \xi + \boldsymbol{\omega}\mathbf{Z}$, where $\boldsymbol{\omega} = \text{diag}(\omega_1, \dots, \omega_d)$, then the density of Y at y is

$$f_y(\mathbf{y}) = \det(\boldsymbol{\omega})^{-1} f_z(\mathbf{z}) \quad (\text{A.5})$$

where

$$f_z(\mathbf{z}) = 2t_d(\mathbf{z}; \bar{\boldsymbol{\Omega}}, \eta)T\left[\boldsymbol{\alpha}^\top \mathbf{z} \sqrt{\frac{\eta+d}{\nu+Q(\mathbf{z})}}; \eta+d\right] \quad (\text{A.6})$$

$$\mathbf{z} = \boldsymbol{\omega}^{-1}(\mathbf{y} - \xi) \quad (\text{A.7})$$

where $t_d(\mathbf{z}; \bar{\Omega}, \eta)$ is a d -dimensional Student's t -distribution with scale matrix $\bar{\Omega}$ and degrees of freedom η , $Q(z) = \mathbf{z}^\top \bar{\Omega}^{-1} \mathbf{z}$ and $T(\cdot; \eta)$ denotes the univariate Student's t distribution function with η degrees of freedom [AC14].

Extremal dependence

For a bivariate skew- t random variable $\mathbf{Y} = [Y(\mathbf{s}), Y(\mathbf{t})]^\top$, the $\chi(h)$ statistic [Pad11] is given by

$$\chi(h) = \bar{F}_{\text{EST}} \left\{ \frac{[x_1^{1/\eta} - \varrho(h)]\sqrt{\eta+1}}{\sqrt{1-\varrho(h)^2}}; 0, 1, \alpha_1, \tau_1, \eta+1 \right\} + \bar{F}_{\text{EST}} \left\{ \frac{[x_2^{1/\eta} - \varrho(h)]\sqrt{\eta+1}}{\sqrt{1-\varrho(h)^2}}; 0, 1, \alpha_2, \tau_2, \eta+1 \right\}, \quad (\text{A.8})$$

where \bar{F}_{EST} is the univariate survival extended skew- t function with zero location and unit scale,

$$\begin{aligned} \varrho(h) &= \text{Cor}[y(\mathbf{s}), y(\mathbf{t})] \\ \alpha_j &= \alpha_i \sqrt{1-\varrho^2} \\ \tau_j &= \sqrt{\eta+1}(\alpha_j + \alpha_i \varrho) \\ x_j &= \frac{F_T(\bar{\alpha}_i \sqrt{\eta+1}; 0, 1, \eta)}{F_T(\bar{\alpha}_j \sqrt{\eta+1}; 0, 1, \eta)}, \end{aligned}$$

with $j = 1, 2$ and $i = 2, 1$ and where $\bar{\alpha}_j = (\alpha_j + \alpha_i \varrho) / \sqrt{1 + \alpha_i^2 [1 - \varrho(h)^2]}$.

Proof that $\lim_{h \rightarrow \infty} \chi(h) > 0$

Consider the bivariate distribution of $\mathbf{Y} = [Y(\mathbf{s}), Y(\mathbf{t})]^\top$, with $\varrho(h)$ given by (2.2). So, $\lim_{h \rightarrow \infty} \varrho(h) = 0$. Then

$$\lim_{h \rightarrow \infty} \chi(h) = \bar{F}_{\text{EST}} \left\{ \sqrt{\eta+1}; 0, 1, \alpha_1, \tau_1, \eta+1 \right\} + \bar{F}_{\text{EST}} \left\{ \sqrt{\eta+1}; 0, 1, \alpha_2, \tau_2, \eta+1 \right\}. \quad (\text{A.9})$$

Because the extended skew- t distribution is not bounded above, for all $\bar{F}_{\text{EST}}(x) = 1 - F_{\text{EST}(x)} > 0$ for all $x < \infty$. Therefore, for a skew- t distribution, $\lim_{h \rightarrow \infty} \chi(h) > 0$.

A.5 Comparisons with other parameterizations

Various forms of multivariate skew-normal and skew- t distributions have been proposed in the literature. In this section, we make a connection between our parameterization in (2.1) of the main

text and another popular version. Azzalini & Capitanio [AC14] and Beranger et al. [Ber16] define a skew-normal process as

$$\tilde{X}(\mathbf{s}) = \tilde{\lambda}|z| + (1 - \tilde{\lambda}^2)^{1/2} v(\mathbf{s}) \quad (\text{A.10})$$

where $\tilde{\lambda} \in (-1, 1)$, $z \sim N(0, 1)$, and $v(\mathbf{s})$ is a Gaussian process with mean zero, variance one, and spatial correlation function ρ . To extend this to the skew- t distribution, Azzalini & Capitanio [AC03] take $\tilde{Y}(\mathbf{s}) = W\tilde{X}(\mathbf{s})$ where $W^{-2} \sim \text{Gamma}(a/2, a/2)$. Returning to the proposed parameterization (with $\beta = 0$), let $W^{-2} = \frac{b}{a}\sigma^{-2}$ so that (2.1) in the manuscript becomes

$$Y(\mathbf{s}) = W \left[\lambda \left(\frac{b}{a} \right)^{1/2} |z| + \left(\frac{b}{a} \right)^{1/2} v(\mathbf{s}) \right]. \quad (\text{A.11})$$

Clearly setting $b/a = (1 - \tilde{\lambda}^2) > 0$, and $\lambda = \tilde{\lambda}/(1 - \tilde{\lambda}^2)^{1/2} \in (-\infty, \infty)$ resolves the difference in parameterizations. We note that our parameterization has three parameters (a, b, λ) compared to the two parameters of the alternative parameterization $(a, \tilde{\lambda})$. Since we have assumed that both $v(\mathbf{s})$ and z have unit scale, the additional b parameter in our parameterization is required to control the precision.

A.6 Temporal dependence

It is very challenging to derive an analytical expression the temporal extremal dependence at a single site \mathbf{s} . However, using simulated data, we have evidence to suggest that the model does exhibits temporal extremal dependence. To demonstrate that our model maintains temporal extremal dependence, we generate lag- m observations for $m = 1, 3, 5, 10$ from our model setting $\phi_w = \phi_z = \phi_\sigma = \varphi$, for $\varphi = 0, 0.02, 0.04, \dots, 1$. To estimate the lag- m chi-statistic $\chi(m)$ we first estimate the lag- m F -madogram $\nu_F(m)$ [Coo06] using $\hat{\nu}_F(m) = \frac{1}{2n} \sum_{i=1}^n |\hat{F}(y_0) - \hat{F}(y_m)|$ where $\hat{F}(\cdot)$ represents an empirical CDF and y_m is the lag- m observation. The F -madogram is related to the χ statistic as follows

$$\chi = 2 - \frac{1 + 2\nu_F}{1 - 2\nu_F}. \quad (\text{A.12})$$

Figure A.2 suggests that the extremal dependence increases as $\varphi \rightarrow 1$, and that the extremal dependence decreases as m increases.

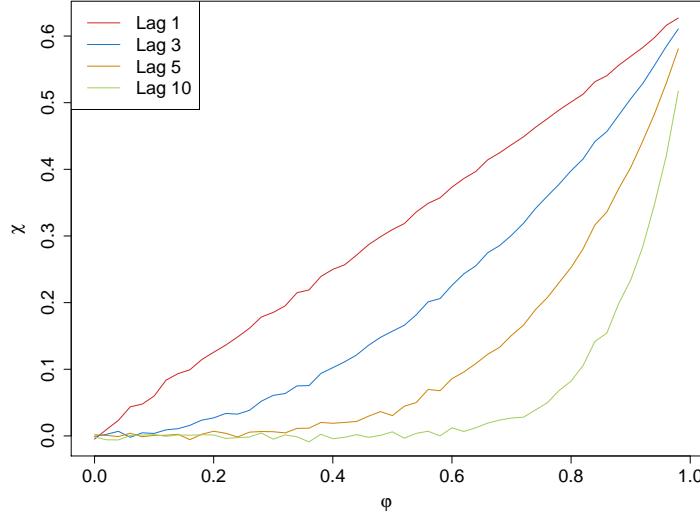


Figure A.2 Simulated lag- m χ for varying levels of φ .

A.7 Brier scores for ozone prediction

Because typical ozone concentration varies throughout the US, we have included Brier scores for exceedance of the 99th quantile by site for two model fits (Gaussian and Symmetric- t , $K = 10$ knots, $T = 75$, time series) in Figure A.3. As we can see in these plots, both models seem to perform similarly across the US with the poorest performance in California. Other methods have similar Brier score maps to these.

A.8 Simulation study results

The following tables show the methods that have significantly different Brier scores when using a Wilcoxon-Nemenyi-McDonald-Thompson test. In each column, different letters signify that the methods have significantly different Brier scores. For example, there is significant evidence to suggest that method 1 and method 4 have different Brier scores at $q(0.90)$, whereas there is not significant evidence to suggest that method 1 and method 2 have different Brier scores at $q(0.90)$. In each table group A represents the group with the lowest Brier scores. Groups are significant with a familywise error rate of $\alpha = 0.05$.

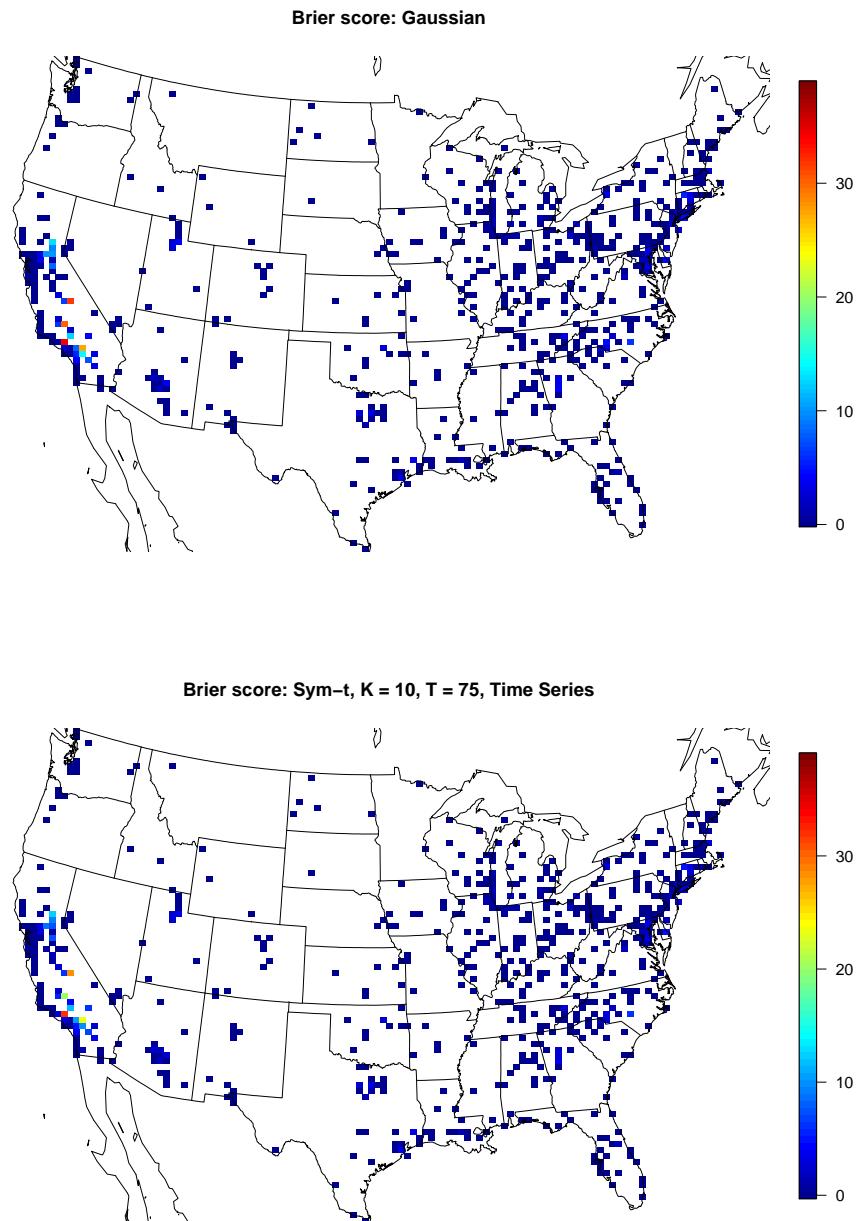


Figure A.3 Map of Brier scores for Gaussian (top) vs Symmetric- t , $K = 10$ knots, $T = 75$, time series (bottom).

Table A.1 Setting 1 – Gaussian marginal, $K = 1$ knot

Method	$q(0.90)$	$q(0.95)$	$q(0.98)$	$q(0.99)$
1	A	A	A	A
2	A	A	A	A
3	B	B	C	A
4	A	A	A B	A
5	B	B	B C	A
6	C	C	D	B

Table A.2 Setting 2 – Skew- t marginal, $K = 1$ knot

Method	$q(0.90)$	$q(0.95)$	$q(0.98)$	$q(0.99)$
1	B	B	B	B
2	A	A	A	A
3	A B	A B	A B	A B
4	A B	A B	A B	A B
5	C	C	C	C
6	D	D	D	C

Table A.3 Setting 3 – Skew- t marginal, $K = 5$ knots

Method	$q(0.90)$	$q(0.95)$	$q(0.98)$	$q(0.99)$
1	C	C	B	B
2	C	C	B	B
3	B	B	A	A
4	A	A	A	A
5	A	A	A	A
6	D	D	C	C

Table A.4 Setting 4 – Max-stable, Asymmetric logistic

Method	$q(0.90)$	$q(0.95)$	$q(0.98)$	$q(0.99)$
1	A B	B	B	C
2	B	B	B	B C
3	C D	C	B	B
4	D	D	C	C
5	C	B C	B	B C
6	A	A	A	A

Table A.5 Setting 5 – Max-stable, Brown-Resnick

Method	$q(0.90)$		$q(0.95)$		$q(0.98)$		$q(0.99)$	
	1	D	2	C	3	A B	4	B
1		D		C		A	B	C
2		D		C		C	B	C
3	A	B		A		A B		B
4		C		B		B		B
5	A			A		A		A B
6		B C		A		A		A

APPENDIX

B

SPATIAL MODEL FOR RARE BINARY EVENTS

B.1 Binary regression using the GEV link

Here, we provide a brief review of the the GEV link of Wang & Dey [WD10]. Let $Y_i \in \{0, 1\}$, $i = 1, \dots, n$ be a collection of i.i.d. binary responses. It is assumed that $Y_i = I(z_i > 0)$ where $I(\cdot)$ is an indicator function, $z_i = [1 - \xi \mathbf{X}_i \boldsymbol{\beta}]^{1/\xi}$ is a latent variable following a $\text{GEV}(1, 1, 1)$ distribution, \mathbf{X}_i is the associated p -vector of covariates with first element equal to one for the intercept, and $\boldsymbol{\beta}$ is a p -vector of regression coefficients. Then, $Y_i \stackrel{\text{ind}}{\sim} \text{Bern}(\pi_i)$ where $\pi_i = 1 - \exp(-\frac{1}{z_i})$.

B.2 Derivation of the likelihood

We use the hierarchical max-stable spatial model given by Reich & Shaby [RS12]. If at each margin, $Z_i \sim \text{GEV}(1, 1, 1)$, then $Z_i | \theta_i \stackrel{\text{ind}}{\sim} \text{GEV}(\theta, \alpha\theta, \alpha)$. We reorder the data such that $Y_1 = \dots = Y_K = 1$, and $Y_{K+1} = \dots = Y_n = 0$. Then the joint likelihood conditional on the random effect θ is

$$\begin{aligned}
P(Y_1 = y_1, \dots, Y_n = y_n) &= \prod_{i \leq K} \left\{ 1 - \exp \left[- \left(\frac{\theta_i}{z_i} \right)^{1/\alpha} \right] \right\} \prod_{i > K} \exp \left[- \left(\frac{\theta_i}{z_i} \right)^{1/\alpha} \right] \\
&= \exp \left[- \sum_{i=K+1}^n \left(\frac{\theta_i}{z_i} \right)^{1/\alpha} \right] - \exp \left[- \sum_{i=K+1}^n \left(\frac{\theta_i}{z_i} \right)^{1/\alpha} \right] \sum_{i=1}^K \exp \left[- \left(\frac{\theta_i}{z_i} \right)^{1/\alpha} \right] \\
&\quad + \exp \left[- \sum_{i=K+1}^n \left(\frac{\theta_i}{z_i} \right)^{1/\alpha} \right] \sum_{1 < i < j \leq K} \left\{ \exp \left[- \left(\frac{\theta_i}{z_i} \right)^{1/\alpha} - \left(\frac{\theta_j}{z_j} \right)^{1/\alpha} \right] \right\} \\
&\quad + \dots + (-1)^K \exp \left[- \sum_{i=1}^n \left(\frac{\theta_i}{z_i} \right)^{1/\alpha} \right]
\end{aligned} \tag{B.1}$$

Finally marginalizing over the random effect, we obtain

$$\begin{aligned}
P(Y_1 = y_1, \dots, Y_n = y_n) &= \int G(\mathbf{z}|\mathbf{A}) p(\mathbf{A}|\boldsymbol{\alpha}) d\mathbf{A} \\
&= \int \exp \left[- \sum_{i=K+1}^n \left(\frac{\theta_i}{z_i} \right)^{1/\alpha} \right] - \exp \left[- \sum_{i=K+1}^n \left(\frac{\theta_i}{z_i} \right)^{1/\alpha} \right] \sum_{i=1}^K \exp \left[- \left(\frac{\theta_i}{z_i} \right)^{1/\alpha} \right] \\
&\quad + \exp \left[- \sum_{i=K+1}^n \left(\frac{\theta_i}{z_i} \right)^{1/\alpha} \right] \sum_{1 < i < j \leq K} \left\{ \exp \left[- \left(\frac{\theta_i}{z_i} \right)^{1/\alpha} - \left(\frac{\theta_j}{z_j} \right)^{1/\alpha} \right] \right\} \\
&\quad + \dots + (-1)^K \exp \left[- \sum_{i=1}^n \left(\frac{\theta_i}{z_i} \right)^{1/\alpha} \right] p(\mathbf{A}|\boldsymbol{\alpha}) d\mathbf{A}.
\end{aligned} \tag{B.2}$$

Consider the first term in the summation,

$$\begin{aligned}
\int \exp \left\{ - \sum_{i=K+1}^n \left(\frac{\theta_i}{z_i} \right)^{1/\alpha} \right\} p(\mathbf{A}|\boldsymbol{\alpha}) d\mathbf{A} &= \int \exp \left\{ - \sum_{i=K+1}^n \left(\frac{\left[\sum_{l=1}^L A_l w_l(\mathbf{s}_i)^{1/\alpha} \right]^\alpha}{z_i} \right)^{1/\alpha} \right\} p(\mathbf{A}|\boldsymbol{\alpha}) d\mathbf{A} \\
&= \int \exp \left\{ - \sum_{i=K+1}^n \sum_{l=1}^L A_l \left(\frac{w_l(\mathbf{s}_i)}{z_i} \right)^{1/\alpha} \right\} p(\mathbf{A}|\boldsymbol{\alpha}) d\mathbf{A} \\
&= \exp \left\{ - \sum_{l=1}^L \left[\sum_{i=K+1}^n \left(\frac{w_l(\mathbf{s}_i)}{z_i} \right)^{1/\alpha} \right]^\alpha \right\}.
\end{aligned} \tag{B.3}$$

The remaining terms in equation (B.2) are straightforward to obtain, and after integrating out the random effect, the joint density for $K = 0, 1, 2$ is given by

$$P(Y_1 = y_1, \dots, Y_n = y_n) = \begin{cases} G(\mathbf{z}), & K = 0 \\ G(\mathbf{z}_{(1)}) - G(\mathbf{z}), & K = 1 \\ G(\mathbf{z}_{(12)}) - G(\mathbf{z}_{(1)}) - G(\mathbf{z}_{(2)}) + G(\mathbf{z}), & K = 2 \end{cases} \quad (\text{B.4})$$

where

$$\begin{aligned} G[\mathbf{z}_{(1)}] &= P[Z(\mathbf{s}_2) < z(\mathbf{s}_2), \dots, Z(\mathbf{s}_n) < z(\mathbf{s}_n)] \\ G[\mathbf{z}_{(2)}] &= P[Z(\mathbf{s}_1) < z(\mathbf{s}_1), Z(\mathbf{s}_3) < z(\mathbf{s}_3), \dots, Z(\mathbf{s}_n) < z(\mathbf{s}_n)] \\ G[\mathbf{z}_{(12)}] &= P[Z(\mathbf{s}_3) < z(\mathbf{s}_3), \dots, Z(\mathbf{s}_n) < z(\mathbf{s}_n)]. \end{aligned}$$

Similar expressions can be derived for all K , but become cumbersome for large K .

APPENDIX

C EMPIRICAL BASIS FUNCTIONS FOR MAX-STABLE SPATIAL DEPENDENCE

C.1 Extreme value distributions

The cumulative distribution function for the GEV is $F(y) = \exp\{-t(y)\}$ where

$$t(y) = \begin{cases} \left[1 + \xi \frac{y - \mu}{\sigma}\right]^{-1/\xi}, & \xi \neq 0 \\ \exp\left\{-\frac{y - \mu}{\sigma}\right\}, & \xi = 0. \end{cases} \quad (\text{C.1})$$

The probability density function for the GEV is given by $f(y) = \frac{1}{\sigma} t(y)^{\xi+1} \exp\{-t(y)\}$ where $t(y)$ is defined in (C.1).

C.2 Grid approximation to PS density

The $\text{PS}(\alpha)$ density can be challenging to use because it does not have a closed form. From Section 2 of [Ste09], the density can be expressed as

$$g_1(A) = \int_0^1 g_1(A, B) dB, \quad (\text{C.2})$$

where

$$g_1(A, B) = \frac{\alpha}{1-\alpha} \left(\frac{1}{A} \right)^{1/(1-\alpha)} c(\pi B) \exp \left\{ - \left(\frac{1}{A} \right)^{\alpha/(1-\alpha)} c(\pi B) \right\}, \quad (\text{C.3})$$

with

$$c(\psi) = \left[\frac{\sin(\alpha\psi)}{\sin(\psi)} \right]^{1/(1-\alpha)} \frac{\sin[(1-\alpha)\psi]}{\sin(\alpha\psi)}. \quad (\text{C.4})$$

Stephenson [Ste09] presents an auxiliary variable technique to deal with the integral in the density function, but we opt to numerically evaluate the integral because it is only one-dimensional. To evaluate the integral, we use 50 evenly spaced quantiles of a Beta(0.5, 0.5) distribution as the midpoints B_1, \dots, B_{50} , and then use the midpoint rule to evaluate $\int_0^1 g_1(A, B) dB$.

C.3 Standardized Gaussian kernel functions

Reich & Shaby [RS12] use standardized Gaussian kernel functions as their spatial basis functions in the low-rank max-stable model. Consider a set of $\mathbf{k}_1, \dots, \mathbf{k}_L$ spatial knot locations in \mathcal{D}^2 , the region of interest. Then

$$\hat{B}_{il} = \frac{\exp \left\{ - \frac{\|\mathbf{s}_i - \mathbf{k}_l\|^2}{2\rho^2} \right\}}{\sum_{j=1}^L \exp \left\{ - \frac{\|\mathbf{s}_i - \mathbf{k}_j\|^2}{2\rho^2} \right\}} \quad (\text{C.5})$$

where $\|\cdot\|$ is the Euclidean distance between a site and a knot location.

C.4 Principal components

As a comparison to the EBFs, Figure C.1 gives the first six principal components for the fire data, and Figure C.2 gives the first six principal components for the precipitation data. These figures show that the EBFs resemble the EOFs for the uncensored precipitation data, but are quite different than the EOF from the censored fire data. This is not surprising considering that the censoring results in only a small subset of data being used to estimate dependence.

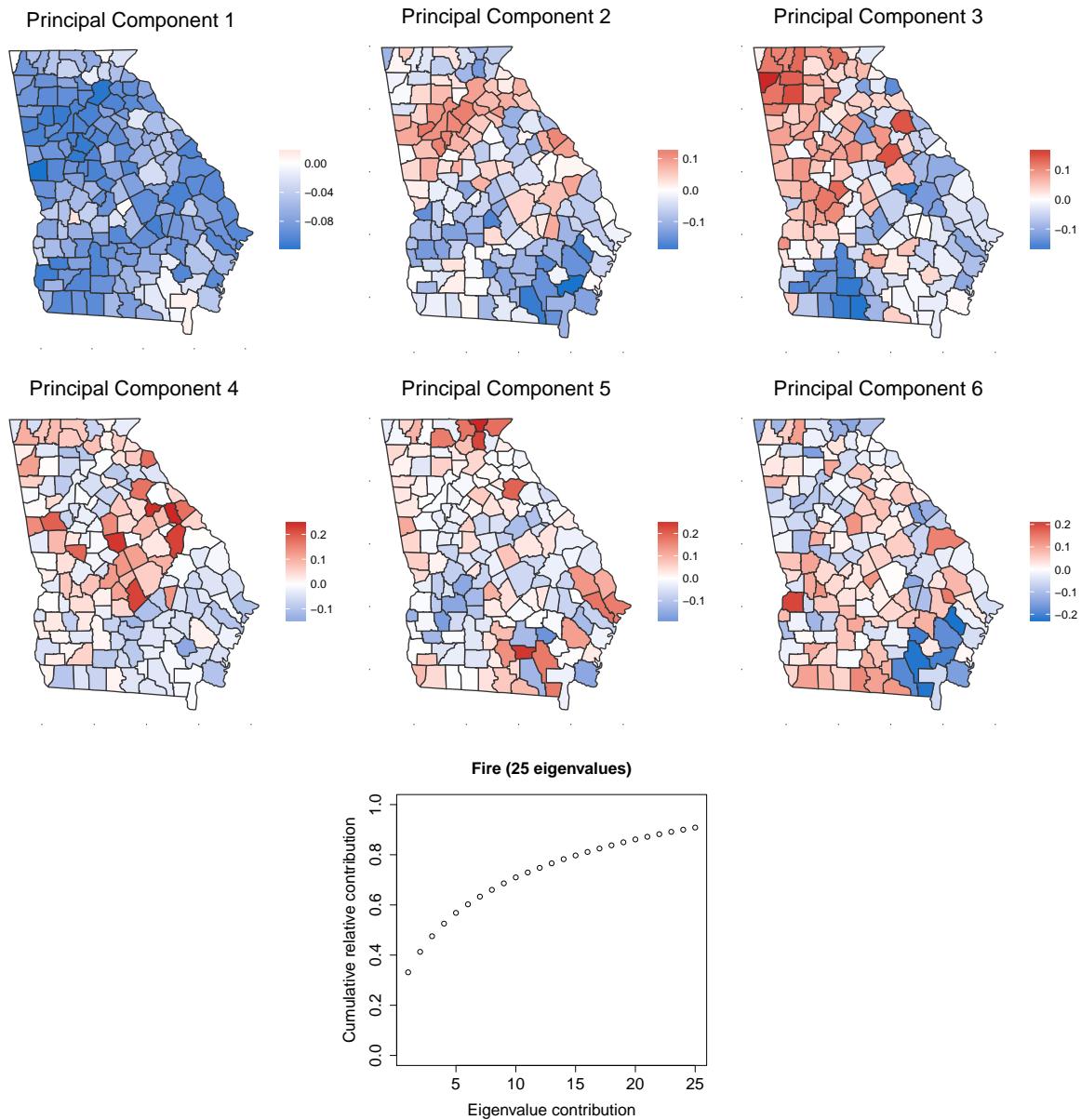


Figure C.1 First six principal components and the cumulative sum of the first 25 eigenvalues for the Georgia fire data.

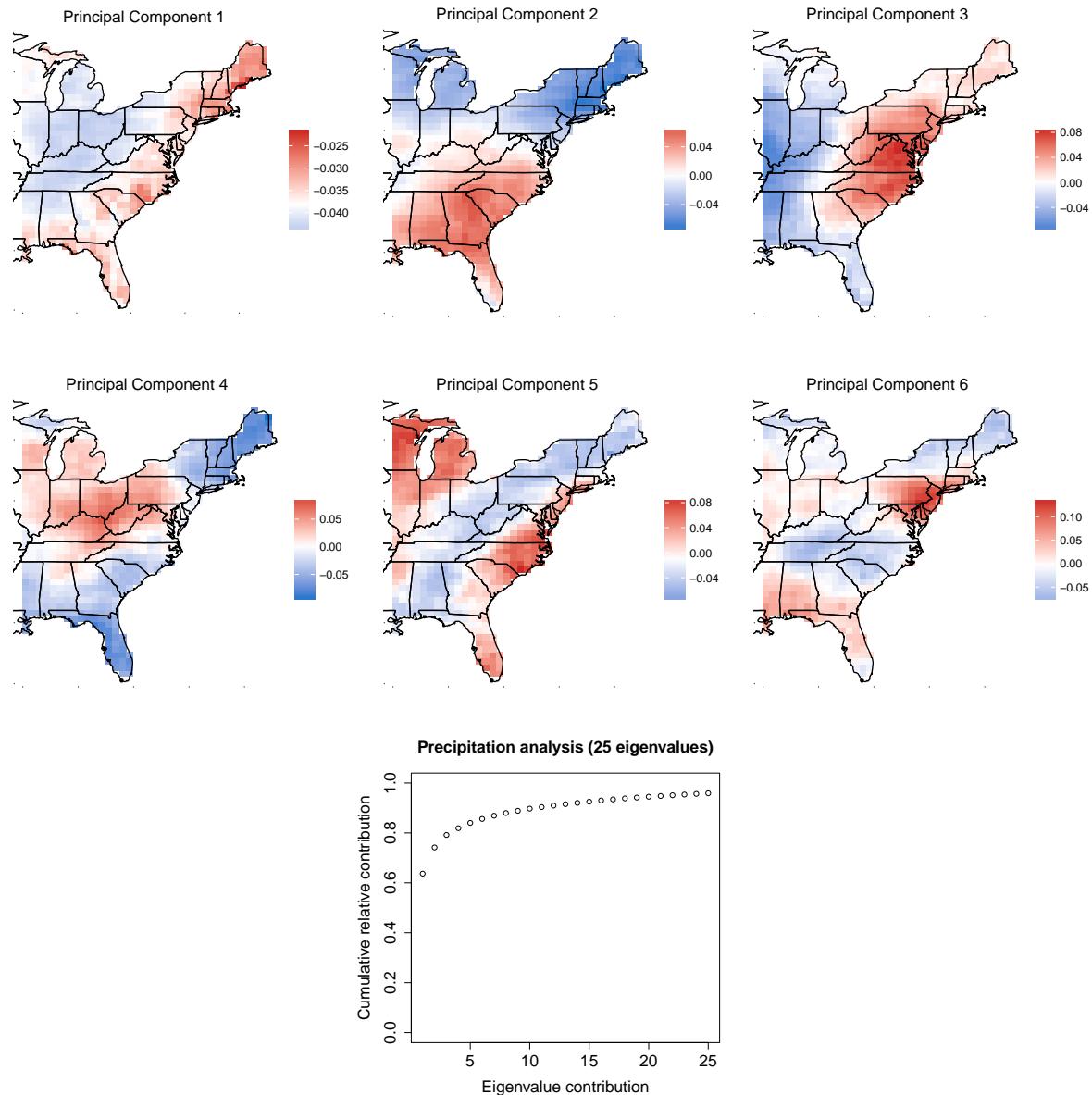


Figure C.2 First six principal components and the cumulative sum of the first 25 eigenvalues for the precipitation data.