

1 **A spatial skew-*t* model for threshold exceedances**

2 July 23, 2015

3 **1 Introduction**

4 In many climatological applications, researchers are interested in learning about the average behavior of
5 different climate variables (e.g. ozone, temperature, rainfall). Our study is motivated by an air pollution
6 application where the focus is not on the average behavior, but instead the behavior over a fixed threshold
7 determined by government regulation. More specifically, we consider consider the case of compliance for
8 ozone. A site is said to be in compliance if the fourth highest daily maximum 8-hour concentration averaged
9 over 3 years does not exceed 75 parts per billion (ppb).

10 Traditional geostatistical modeling is based upon the assumption that observations come from a Gaussian
11 process, a process that is fully defined by its mean and covariance functions. In the limit of the Gaussian
12 distribution, all observations are independent regardless of the strength of the correlation in the bulk of the
13 data. Furthermore, the Gaussian distribution is light-tailed and symmetric. Therefore, it is inappropriate to
14 use standard geostatistical methods when trying to describe dependence in the tails of the distribution.

15 Threshold modeling is popular in the field of extreme value statistics where extreme events are naturally
16 defined in terms of exceedances over a high threshold. Davison and Smith (1990) considered modeling
17 threshold exceedances of univariate time series by the generalized Pareto distribution. Bivariate threshold
18 models for extreme value distributions were considered by Ledford and Tawn (1996) who introduced a
19 censored approach that provides a way to deal with different types of exceedances of a bivariate threshold in
20 terms of only one or both components. These threshold models were extended to spatial models for extremes
21 by Wadsworth and Tawn (2012) and Thibaud et al. (2013) who fit various models to spatial extremes using a
22 censored pairwise likelihood (Padoan et al., 2010) based on the approach of Ledford and Tawn (1996). Huser

23 and Davison (2014) further extended this to spate-time modeling. Engelke et al. (2014), Wadsworth and
24 Tawn (2014), and Thibaud and Opitz (2013) introduced more efficient inference for threshold exceedances
25 of extremal spatial processes with full likelihood methods. The previous approaches to threshold modeling
26 are motivated by extreme value theory and assume the threshold is high enough to assume extremal models
27 are valid for the data, and for extrapolation beyond the range of observed values. Moreover, these approaches
28 are computationally intensive and limited to rather small datasets. Our application with ozone data does not
29 fit into this framework because we do not focus on exceedances of a very high threshold, but on exceedances
30 of a fixed threshold.

31 Instead, we propose a new spatiotemporal threshold exceedance model based on the skew-*t* process
32 (Padoan, 2011). Our model is a threshold exceedance model for the multivariate skew-*t* distribution that
33 uses imputation for values below a fixed threshold. We use a skew-*t* distribution because of its flexibility to
34 model asymmetry and heavy-tailed data with the aim of predicting the probability of exceeding a high fixed
35 threshold at an unobserved location.

36 In a spatial setting, the multivariate skew-*t* distribution demonstrates asymptotic dependence between
37 observations at all sites regardless of the distance between the sites. In order to address this concern, we
38 introduce a random spatial partition similar to the method used by Kim et al. (2005) for non-stationary
39 Gaussian data. This partition alleviates the asymptotic spatial dependence present in the skew-*t* distribution
40 for sites that are far apart. Finally, our model allows for inference and predictions using the full likelihood
41 with computing on the order of Gaussian models for large space-time datasets.

42 The paper is organized as follows. Section 2 is a brief review of the spatial skew-*t* process. In Section
43 3.3, we build upon the traditional skew-*t* by incorporating censoring to focus on tails, partitioning to remove
44 long-range asymptotic dependence, and extending the model to space-time data. The computing is described
45 in Section 4.1. In Section 5, we present a simulation study that examines the predictive capabilities of this

46 model compared with a naïve Gaussian method. We then compare our method to Gaussian and max-stable
 47 methods with a data analysis of ozone measurements from throughout the US in section 6. The final section
 48 provides brief discussion and direction for future research.

49 **2 Spatial skew processes**

50 Many types of data demonstrate some level of skewness and therefore should be modeled with distributions
 51 that allow for asymmetry. The skew-elliptical family of distributions provides models that are mathemati-
 52 cally tractable while introducing a slant parameter to account for asymmetric data (Azzalini and Capitanio,
 53 2014). A brief review of the additive process by which a skew- t process is created is given here.

54 **2.1 Skew- t process**

55 Let $Y(\mathbf{s})$ be the observation at spatial location $\mathbf{s} = (s_1, s_2)$ in a spatial domain of interest $\mathcal{D} \in \mathbb{R}^2$. The
 56 spatial skew- t process can be written

$$Y(\mathbf{s}) = \mathbf{X}(\mathbf{s})^T \boldsymbol{\beta} + \lambda \sigma |z| + \sigma v(\mathbf{s}) \quad (1)$$

57 where $\mathbf{X}(\mathbf{s})$ is a set of spatial covariates at site \mathbf{s} , $\boldsymbol{\beta}$ is the vector of regression parameters, $\lambda \in \mathcal{R}$ is a
 58 parameter controlling skew, $z \sim N(0, 1)$, $\sigma^2 \sim \text{IG}(a, b)$ is an inverse gamma random variable, and $v(\mathbf{s})$ is
 59 a spatial Gaussian process with mean zero, variance one, and a positive definite correlation function.

60 For a finite collection of locations $\mathbf{s}_1, \dots, \mathbf{s}_n$, denote the vector of observations $\mathbf{Y} = [Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n)]^T$.

61 After marginalizing over both z and σ ,

$$\mathbf{Y} \sim \text{ST}_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, 2a), \quad (2)$$

62 that is, \mathbf{Y} follows an n -dimensional skew- t distribution with location $\mathbf{X}\boldsymbol{\beta}$, correlation matrix $\boldsymbol{\Omega}$, slant parameters $\boldsymbol{\alpha}$ and degrees of freedom $2a$, where $\mathbf{X} = [\mathbf{X}(\mathbf{s}_1)^T, \dots, \mathbf{X}(\mathbf{s}_n)^T]$, $\boldsymbol{\Omega} = \boldsymbol{\omega}\bar{\boldsymbol{\Omega}}\boldsymbol{\omega}$, $\boldsymbol{\omega} = \text{diag}\left(\frac{1}{\sqrt{ab}}, \dots, \frac{1}{\sqrt{ab}}\right)$,
 63 $\bar{\boldsymbol{\Omega}} = (\boldsymbol{\Sigma} + \lambda^2 \mathbf{1}\mathbf{1}^T)$, $\boldsymbol{\alpha} = \lambda(1 + \lambda^2 \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1})^{-1/2} \mathbf{1}^T \boldsymbol{\Sigma}^{-1}$ is a vector of slant parameters, and $\boldsymbol{\Sigma}$ is the positive
 64 definite correlation matrix of $[v(\mathbf{s}_1), \dots, v(\mathbf{s}_n)]$. Although any positive definite correlation function could
 65 be used, we choose to use the stationary isotropic Matérn correlation with
 66

$$\text{cor}[v(\mathbf{s}), v(\mathbf{t})] = \gamma I(\mathbf{s} = \mathbf{t}) + (1 - \gamma) \frac{1}{\Gamma(\nu) 2^{\nu-1}} \left(\sqrt{2\nu} \frac{h}{\rho} \right)^\nu K_\nu \left(\sqrt{2\nu} \frac{h}{\rho} \right) \quad (3)$$

67 where ρ is the spatial range, ν is the smoothness, γ is the proportion of variance accounted for by the
 68 spatial variation, K_ν is a modified Bessel function of the second kind, and $h = \|\mathbf{s} - \mathbf{t}\|$. This process is
 69 desirable because of its flexible tail that is controlled by the skewness parameter λ and degrees of freedom
 70 $2a$. Furthermore, the marginal distributions at each location also follow a univariate skew- t distribution
 71 (Azzalini and Capitanio, 2014).

72 2.2 Extremal dependence

73 Our interest lies in spatial dependence in the tail of the skew- t process. One measure of extremal dependence
 74 is the χ statistic (Coles et al., 1999). For a stationary and isotropic spatial process, the χ statistic for locations
 75 \mathbf{s} and \mathbf{t} separated by distance $h = \|\mathbf{s} - \mathbf{t}\|$ with identical marginal distributions is

$$\chi(h) = \lim_{c \rightarrow c^*} \Pr[Y(\mathbf{s}) > c | Y(\mathbf{t}) > c] \quad (4)$$

76 where c^* is the upper limit of the support of Y . If $\chi(h) = 0$, then observations are asymptotically indepen-
 77 dent at distance h . For Gaussian processes, $\chi(h) = 0$ regardless of the distance h , so they are not suitable for
 78 modeling asymptotically dependent extremes. Unlike the Gaussian process, the skew- t process is asymptot-

79 ically dependent (the explicit expression for $\chi(h)$ is given in Appendix A.4). However, one problem with
 80 the spatial skew- t process is that $\lim_{h \rightarrow \infty} \chi(h) > 0$. This occurs because all observations, both near and
 81 far, share the same z and σ terms. Therefore, this long-range dependence feature of the skew- t process is
 82 not ideal for spatial analysis of large geographic regions where we expect only local spatial dependence.

83 **3 Spatiotemporal skew- t model for extremes**

84 In this section, we propose extensions to the skew- t process to model spatial extremes over a large geo-
 85 graphic region by introducing censoring to focus on tail behavior and a random partition to remove long-
 86 range asymptotic dependence. For notational convenience, we introduce the model for a single replication,
 87 and then extend this model to the spatiotemporal setting in Section 3.3.

88 **3.1 Censoring to focus on the tails**

89 Because one of our goals is to model the dependence of the distribution in the tails of the data, we choose to
 90 censor values below threshold. Let

$$\tilde{Y}(\mathbf{s}) = \begin{cases} Y(\mathbf{s}) & \delta(\mathbf{s}) = 1 \\ T & \delta(\mathbf{s}) = 0 \end{cases} \quad (5)$$

91 be the censored observation at site \mathbf{s} where $Y(\mathbf{s})$ is the uncensored observation, $\delta(\mathbf{s}) = I[Y(\mathbf{s}) > T]$, and T
 92 is a pre-specified threshold value. Then, assuming the uncensored data $Y(\mathbf{s})$ are observations from a skew- t
 93 process, we update values censored below the threshold using standard Bayesian missing data methods as
 94 described in Section 4.1.

95 **3.2 Partitioning to remove long-range asymptotic dependence**

96 The motivation for the partition is that for a large spatial domain, it may not be reasonable to assume sites
 97 that are far apart demonstrate asymptotic dependence. Modeling different levels of asymptotic dependence
 98 was discussed by Wadsworth and Tawn (2012) with a hybrid spatial dependence model. Huser and Davison
 99 (2014) also allow for asymptotic dependence across both space and time with a partition structure repre-
 100 sented by random discs moving across the space for a random duration with a random velocity and random
 101 radius. We handle the problem of long-range asymptotic dependence with a random partition. As dis-
 102 cussed in Section 2, the source of long-range dependence is the shared z and σ . Therefore, to alleviate this
 103 dependence, we allow z and σ to vary by site. The model becomes

$$Y(\mathbf{s}) = \mathbf{X}(\mathbf{s})^T \boldsymbol{\beta} + \lambda \sigma(\mathbf{s})|z(\mathbf{s})| + \sigma(\mathbf{s})v(\mathbf{s}). \quad (6)$$

104 Let $\mathbf{w} = (w_1, w_2)$ be the location of a spatial knot. To model spatial variation, consider a set of spatial knots
 105 $\mathbf{w}_1, \dots, \mathbf{w}_K$ from a homogeneous Poisson process with intensity μ over spatial domain $\mathcal{D} \in \mathbb{R}^2$. The knots
 106 define a random partition of \mathcal{D} by subregions P_1, \dots, P_K defined as

$$P_k = \{\mathbf{s} : k = \arg \min_\ell \|\mathbf{s} - \mathbf{w}_\ell\|\}. \quad (7)$$

107 All $z(\mathbf{s})$ and $\sigma(\mathbf{s})$ for sites in subregion k are assigned common values

$$z(\mathbf{s}) = z_k \quad \text{and} \quad \sigma(\mathbf{s}) = \sigma_k \quad (8)$$

108 and the z_k and σ_k^2 are distributed as $z_k \stackrel{iid}{\sim} N(0, 1)$ and $\sigma^2 \stackrel{iid}{\sim} \text{IG}(a, b)$ where IG is the distribution function
 109 of an inverse gamma random variable. So, within each partition, $Y(\mathbf{s})$ follows the spatial skew- t process

110 defined in Section 2. Across partitions, the $Y(\mathbf{s})$ remain correlated via the correlation function for $v(\mathbf{s})$
111 because $v(\mathbf{s})$ spans all partitions.

112 When incorporating the random partition, conditional on knots $\mathbf{w}_1, \dots, \mathbf{w}_K$, the χ statistic for two sites
113 \mathbf{s} and \mathbf{t} in partitions k_s and k_t respectively is

$$\begin{aligned}\chi(h) &= I(k_s = k_t)\chi_{\text{skew-}t}(h) + I(k_s \neq k_t)\chi_{\text{Gaus}}(h) \\ &= I(k_s = k_t)\chi_{\text{skew-}t}(h)\end{aligned}\tag{9}$$

114 where $I(\cdot)$ is an indicator function, $\chi_{\text{skew-}t}(h)$ is the χ statistic for a skew- t process given in (28), $\chi_{\text{Gaus}}(h)$
115 is the χ statistic for a Gaussian process, and $h = \|\mathbf{s} - \mathbf{t}\|$. Therefore, sites in different subregions are
116 asymptotically independent because $\chi_{\text{Gaus}}(h) = 0$ for all h . Marginally, over the knots $\mathbf{w}_1, \dots, \mathbf{w}_K$, $\chi(h) =$
117 $\pi(h)\chi_{\text{skew-}t}(h)$, where $\pi(h) = \Pr(k_s = k_t)$ is the probability that two sites separated by distance h are in
118 the same partition. So, to show that $\lim_{h \rightarrow \infty} \chi(h) = 0$, we need only know that $\lim_{h \rightarrow \infty} \pi(h) = 0$. A proof
119 of this is given in Appendix A.3.

120 In Figure 1, we give $\chi(h)$ for $K = 1, 3, 5, 10$ partitions for a skew- t distribution with $\alpha = 10$, and
121 3 degrees of freedom. To estimate $\pi(h)$, we generate 500 sites uniformly over the unit-square. We then
122 randomly generate 400 different sets of partitions using $K = 3, 5$, and 10. For each set of knots, we
123 take $\pi(h)$ to be the proportion of sites in the same partition that are separated by distance h . This plot
124 demonstrates how partitioning helps to reduce extremal dependence as h increases.

125 3.3 Extension to space-time data

126 When using daily measurements, the assumption of temporal independence is often inappropriate. In this
127 section, we extend (6) to the spatiotemporal setting. There are several places where temporal dependence
128 could be incorporated in the model, including the residuals $v_t(\mathbf{s})$. However, we choose to allow for temporal

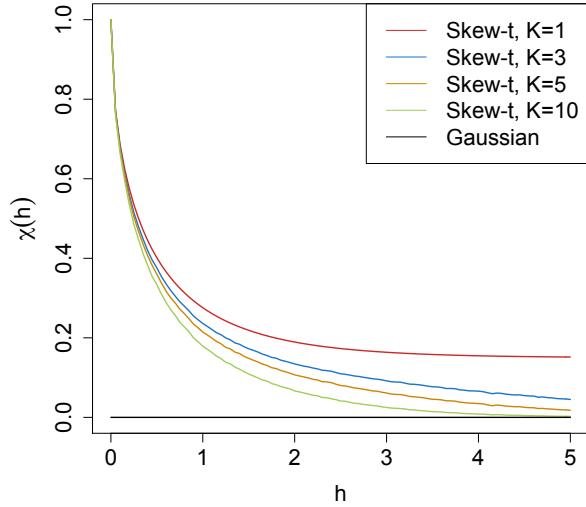


Figure 1: $\chi(h)$ for $K = 1, 3, 5$, and 10 knots as a function of distance.

129 dependence in the \mathbf{w} , z , and σ terms because these terms dictate the tail behavior which is our primary focus.

130 Let

$$Y_t(\mathbf{s}) = \mathbf{X}_t(\mathbf{s})^T \boldsymbol{\beta} + \lambda \sigma_t(\mathbf{s}) |z_t(\mathbf{s})| + \sigma_t(\mathbf{s}) v_t(\mathbf{s}), \quad (10)$$

131 where $t \in \{1, \dots, T\}$ denotes the day of each observation. Let $\mathbf{w}_{tk} = (w_{tk1}, w_{tk2})$ be a spatial knot on day
 132 t , and let w_{t1}, \dots, w_{tK} be a collection of spatial knots on day t . As in Section 3.2, these knots define a daily
 133 partition P_{t1}, \dots, P_{tK} , and for $\mathbf{s} \in P_{tk}$,

$$z_t(\mathbf{s}) = z_{tk} \quad \text{and} \quad \sigma_t(\mathbf{s}) = \sigma_{tk}. \quad (11)$$

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

136 We use an AR(1) time series model for w_{tk} , z_{tk} , and σ_{tk} . The time series model must be specified after

137 a transformation to preserve the skew- t process at each time point. For each time-varying parameter, we
 138 transform to obtain a standard normal marginal distribution, place a Gaussian prior with autocorrelation on
 139 the transformed parameter, and then transform back to obtain the marginal distribution required to preserve
 140 the skew- t process. We first transform the spatial knots from \mathcal{D} to \mathcal{R}^2 as follows. Let

$$w_{tki}^* = \Phi^{-1} \left[\frac{w_{tki} - \min(\mathbf{s}_i)}{\max(\mathbf{s}_i) - \min(\mathbf{s}_i)} \right], \quad i = 1, 2 \quad (12)$$

141 where Φ is a univariate standard normal density function, and $\mathbf{s}_i = [s_{1i}, \dots, s_{ni}]$. Then the transformed
 142 knots $\mathbf{w}_{tk}^* \in \mathcal{R}^2$. We use a copula on $\sigma_t^2(\mathbf{s})$ to ensure that the marginal distributions of $\sigma_t^2(\mathbf{s})$ are inverse
 143 gamma. Let

$$\sigma_t^{2*}(\mathbf{s}) = \Phi^{-1} \{ \text{IG}[\sigma_t^2(\mathbf{s})] \} \quad (13)$$

144 where IG is defined as before. We also use a copula on $z_t(\mathbf{s})$ to ensure that the marginal distributions of
 145 $z_t(\mathbf{s})$ are half-normal. Let

$$z_t^*(\mathbf{s}) = \Phi^{-1} \{ \text{HN}[z_t(\mathbf{s})] \} \quad (14)$$

146 where HN is the distribution function of a half-normal random variable. The AR(1) process for each tail
 147 parameter is $\mathbf{w}_{1k}^* \sim N_w(0, 1)$, $z_{1k}^* \sim N(0, \sigma_{1k}^2)$, $\sigma_{1k}^{2*} \sim N(0, 1)$, and for $t > 1$ the time series is modeled as

$$\mathbf{w}_{tk}^* | \mathbf{w}_{t-1,k}^* \sim N_2 [\phi_w \mathbf{w}_{t-1,k}^*, (1 - \phi_w^2)] \quad (15)$$

$$z_{tk}^* | z_{t-1,k}^* \sim N [\phi_z z_{t-1,k}^*, \sigma_{tk}^2 (1 - \phi_z^2)] \quad (16)$$

$$\sigma_{tk}^{2*} | \sigma_{t-1,k}^{2*} \sim N [\phi_\sigma \sigma_{t-1,k}^{2*}, (1 - \phi_\sigma^2)] \quad (17)$$

148 where $|\phi_w|, |\phi_z|, |\phi_\sigma| < 1$. These are stationary time series models with marginal distributions $\mathbf{w}_k^* \sim N_2(0, 1)$,
149 $z_k^* \sim N(0, \sigma_k^2)$, and $\sigma_k^{2*} \sim N(0, 1)$. After transformation back to the original space, $\mathbf{w}_{tk} \sim \text{Unif}(\mathcal{D})$,
150 $z_{tk} \sim HN(0, \sigma_{tk}^2)$ $\sigma_{tk}^2 \sim \text{IG}(a, b)$. For each day, the model is identical to the spatial-only model in (6)
151 by construction.

152 4 Hierarchical model

153 Conditioned on $z_{tk}(\mathbf{s})$, $\sigma_{tk}^2(\mathbf{s})$, and P_{tk} , the marginal distributions are Gaussian and the joint distribution
154 multivariate Gaussian. However, we do not fix the partitions, they are treated as unknown and updated in the
155 MCMC. We model this with a Bayesian hierarchical model as follows. Let $\mathbf{w}_{t1}, \dots, \mathbf{w}_{tK}$ be a set of daily
156 spatial knots in a spatial domain of interest, \mathcal{D} , and P_{tk} as defined in (7). In practice, we fix K at many

¹⁵⁷ different levels, and assess the impact of fit as described in 5.2. Then

$$Y_t(\mathbf{s}) \mid z_t(\mathbf{s}), \sigma_t^2(\mathbf{s}), P_{tk}, \Theta = \mathbf{X}_t(\mathbf{s})^T \beta + \lambda |z_t(\mathbf{s})| + \sigma_t(\mathbf{s}) v_t(\mathbf{s}) \quad (18)$$

$$z_t(\mathbf{s}) = z_{tk} \text{ if } \mathbf{s} \in P_{tk}$$

$$\sigma_t^2(\mathbf{s}) = \sigma_{tk}^2 \text{ if } \mathbf{s} \in P_{tk}$$

$$\lambda = \lambda_1 \lambda_2$$

$$\lambda_1 = \begin{cases} +1 & \text{w.p. 0.5} \\ -1 & \text{w.p. 0.5} \end{cases}$$

$$\lambda_2^2 \sim IG(a, b)$$

$$v_t(\mathbf{s}) \mid \Theta \sim \text{Matérn}(0, \Sigma)$$

$$z_{tk}^* \mid z_{t-1,k}^*, \sigma_{tk}^2 \sim N(\phi_z z_{t-1,k}^*, \sigma_{tk}^2(1 - \phi_z^2))$$

$$\sigma_{tk}^{2*} \mid \sigma_{t-1,k}^{2*} \sim N(\phi_\sigma \sigma_{t-1,k}^{2*}, (1 - \phi_\sigma^2))$$

$$\mathbf{w}_{tk}^* \mid \mathbf{w}_{t-1,k}^* \sim N_2(\phi_w \mathbf{w}_{t-1,k}^*, (1 - \phi_w^2))$$

¹⁵⁸ where $\Theta = \{\rho, \nu, \gamma, \lambda, \beta\}$, and Σ is a Matérn covariance matrix as described in Section 2.1. We parameterize

¹⁵⁹ $\lambda = \lambda_1 \lambda_2$ to help with convergence in the MCMC.

¹⁶⁰ 4.1 Computation

¹⁶¹ We use Markov chain Monte Carlo methods to explore the posterior. At each MCMC iteration, we first

¹⁶² impute values below the threshold conditional on observations above the threshold. This is feasible for large

¹⁶³ datasets with our model because for a single day, conditional on the model parameters, we only need to draw

¹⁶⁴ from a truncated multivariate normal distribution. We can use Gibbs sampling to update $Y_t(\mathbf{s})$ for censored

¹⁶⁵ observations that are below the threshold T . After conditioning on λ , $z_t(\mathbf{s})$ and non-censored observations,

166 $Y_t(\mathbf{s})$ has truncated normal full conditionals. So we sample $Y_t(\mathbf{s}) \sim N_{(-\infty, T)}(\mathbf{X}_t^T(\mathbf{s})\beta + \lambda|z_t(\mathbf{s})|, \Sigma)$.

167 Then, we update model parameters, Θ , using Metropolis Hastings or Gibbs sampling when appropriate.

168 The final step of the computation is to use Bayesian Kriging to generate a predictive distribution for $Y_t(\mathbf{s}^*)$

169 at prediction location \mathbf{s}^* . This step is similar to the imputation for censored observations except that the full

170 conditionals are no longer truncated at T . See Appendices A.1 and A.2 for details regarding the MCMC.

171 5 Simulation study

172 In this section, we conduct a simulation study to investigate how the number of partitions and the level of

173 thresholding impact the accuracy of predictions made by the model.

174 5.1 Design

175 For all simulation designs, we generate data from the model in Section 3.2 using $n_s = 144$ sites and

176 $n_t = 50$ independent days. The sites are generated Uniform($[0, 10] \times [0, 10]$). We generate data from 4

177 different simulation designs:

178 1. Gaussian marginal, $K = 1$ knot

179 2. Skew- t marginal, $K = 1$ knots

180 3. Skew- t marginal, $K = 5$ knots

181 4. Max-stable

182 In the first three designs, the $v_t(\mathbf{s})$ terms are generated using a Matérn covariance with smoothness parameter

183 $\nu = 0.5$ and spatial range $\rho = 1$. For the covariance matrices in designs 1 – 3, the proportion of the variance

184 accounted for by the spatial variation is $\gamma = 0.9$ while the proportion of the variance accounted for by the

185 nugget effect is 0.1. In the first design, $\sigma^2 = 2$ is used for all days which results in a Gaussian distribution.

186 For designs 2 and 3, $\sigma_{tk}^2 \stackrel{iid}{\sim} \text{IG}(3, 8)$ to give a t distribution with 6 degrees of freedom. For design 1,

187 we set $\lambda = 0$. For designs 2 and 3, $\lambda = 3$ was used as to simulate moderate skewness, and the z_t are
188 generated as described in (8). In designs 1 – 3, the mean $\mathbf{X}^T \boldsymbol{\beta} = 10$ is assumed to be constant across space.

189 In the fourth design, we generate from a spatial max-stable distribution (Reich and Shaby, 2012). In this
190 design, data have marginal distributions that follow a generalized extreme value distribution with parameters
191 $\mu = 1, \sigma = 1, \xi = 0.2$. In this model, a random effect is used to induce spatial dependence using 144 spatial
192 knots on a regular lattice in the square $[1, 9] \times [1, 9]$. For this setting, we set $\gamma = 0.5$.

193 $M = 50$ data sets are generated for each design. For each data set we fit the data using six models

194 1. Gaussian marginal, $K = 1$ knots

195 2. Skew- t marginal, $K = 1$ knots, $T = -\infty$

196 3. Symmetric- t marginal, $K = 1$ knots, $T = q(0.80)$

197 4. Skew- t marginal, $K = 5$ knots, $T = -\infty$

198 5. Symmetric- t marginal, $K = 5$ knots, $T = q(0.80)$

199 6. A max-stable model based on Reich and Shaby (2012) thresholded at $T = q(0.80)$

200 where $q(0.80)$ is the 80th sample quantile of the data. The design matrix \mathbf{X} includes an intercept with a first-
201 order spatial trend with priors of $\beta_{\text{int}}, \beta_{\text{lat}}, \beta_{\text{long}}, \stackrel{iid}{\sim} N(0, 10)$. The spatial covariance parameters have priors
202 $\log(\nu) \sim N(-1.2, 1), \gamma \sim \text{Unif}(0, 1), \rho \sim \text{Unif}(15)$. The skewness parameter has prior $\lambda_2 \sim \text{IG}(0.1, 0.1)$.

203 The residual variance terms have priors $\sigma_t^2(\mathbf{s}) \sim \text{IG}(0.1, 0.1)$. The knots have priors $\mathbf{w} \sim \text{Unif}(\mathcal{D})$. We tried
204 also fitting the skew- t marginals for the thresholded models, but it is very challenging for the MCMC to
205 properly identify the skewness parameter with a censored left tail. Each chain of the MCMC ran for 20,000
206 iterations with a burn-in period of 10,000 iterations. Parameters appear to converge properly; however, in
207 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
208 $\sigma^2(\mathbf{s})$ because of partition label switching throughout the MCMC.

209 **5.2 Cross validation**

210 Models were compared using cross validation with 100 sites used as training sites and 44 sites withheld for
211 testing. The model was fit using the training set, and predictions were generated at the testing site locations.
212 Because one of the primary goals of this model is to predict exceedances over a fixed threshold, we use Brier
213 scores to select the model that best fits the data (Gneiting and Raftery, 2007). The Brier score for predicting
214 exceedance of a threshold c is given by $[e(c) - P(c)]^2$ where $e(c) = I[y > c]$ is an indicator function
215 indicating that a test set value, y , has exceeded the threshold, c , and $P(c)$ is the predicted probability of
216 exceeding c . We average the Brier scores over all test sites and days. For the Brier score, a lower score
217 indicates a better fit.

218 **5.3 Results**

219 We compared the Brier scores for exceeding 4 different thresholds for each dataset. The thresholds used for
220 the Brier scores are extreme quantiles from the simulated data for $q(0.90)$, $q(0.95)$, $q(0.98)$, and $q(0.99)$.
221 Figure 2 gives the Brier score relative to the Brier score for the Gaussian method calculated as

$$BS_{\text{rel}} = \frac{BS_{\text{method}}}{BS_{\text{Gaussian}}}. \quad (19)$$

222 We analyzed the results for the simulation study using a Friedman test at $\alpha = 0.05$ to see if at least one
223 method had a significantly different Brier score. For Friedman tests that came back with a significant p-
224 value, we conducted a Wilcoxon-Nemenyi-McDonald-Thompson test to see which of the methods had dif-
225 ferent results. The full results for the Wilcoxon-Nemenyi-McDonald-Thompson tests are given in Appendix
226 A.5.

227 A plot of the relative Brier scores is given in Figure 2. We find that when the data are generated from a

228 Gaussian process, our method performs comparably to a Gaussian approach. In general, when the underly-
229 ing process is not Gaussian, our method results in an improvement over both the max-stable and Gaussian
230 methods. One exception to this is the case when the generative process is max-stable. In this case, the
231 max-stable method outperforms our method; however, for predictions further out in the tail, the differences
232 between the max-stable method and our method diminish. The non-thresholded methods tend to outperform
233 the thresholded methods, but this is not surprising given that in most cases, the data are generated directly
234 from the model used in the method. In summary, our method provides more flexibility for data that demon-
235 strate some level of asymmetry or heavy tails, while still performing comparably to Gaussian methods when
236 the data are symmetric and have light tails.

237 **6 Data analysis**

238 To illustrate this method, we consider 31 daily observations of maximum 8-hour ozone measurements for
239 July 2005 at 1089 Air Quality System (AQS) monitoring sites in the United States as the response (see Figure
240 3). For each site, we also have covariate information containing the estimated ozone from the Community
241 Multi-scale Air Quality (CMAQ) modeling system. Initially, we fit a linear regression assuming a mean
242 function of

$$E[Y_i(\mathbf{s})] = \beta_0 + \beta_1 \cdot \text{CMAQ}_t(\mathbf{s}). \quad (20)$$

243 The data from July 10 are shown in Figure 3 along with a Q-Q plot of the residuals compared to a skew-*t*
244 distribution with 10 d.f. and $\alpha = 1$.

245 Standard exploratory data analysis techniques for extremal dependence are very challenging with only
246 31 days worth of data because it is difficult to estimate extreme quantiles at each site to obtain empirical

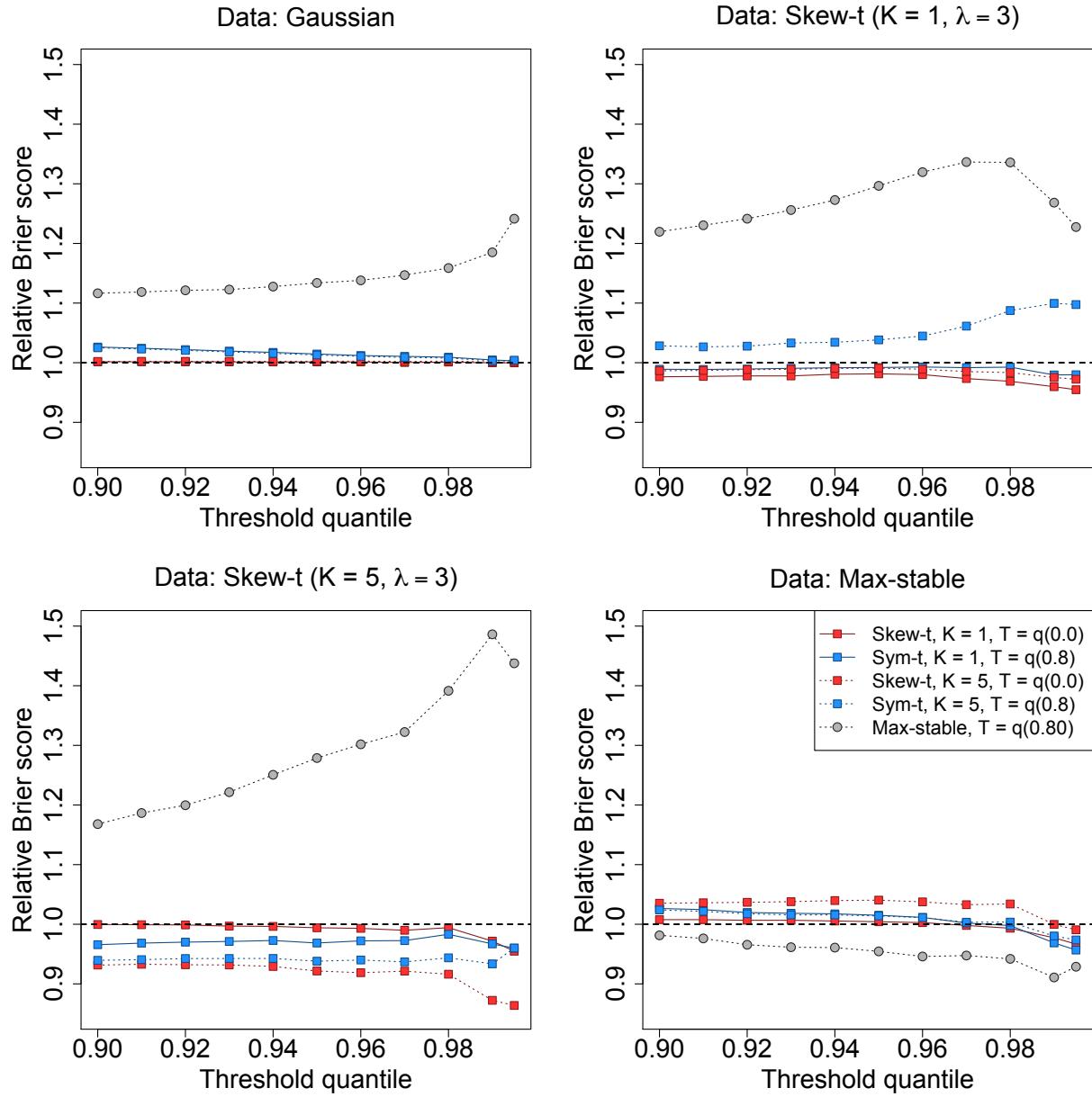


Figure 2: Brier scores relative to the Gaussian method for simulation study results. A ratio lower than 1 indicates that the method outperforms the Gaussian method.

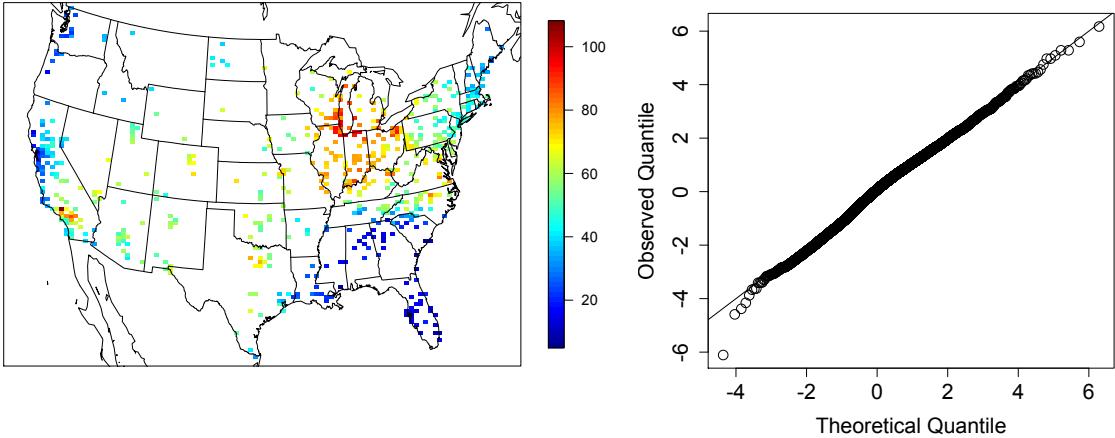


Figure 3: Ozone values on 10 July 2005 (left) Q-Q plot of the residuals for a skew- t distribution with 10 d.f. and $\alpha = 1$ (right)

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. Figure 6 suggests suggests there is more agreement between sites that are close to one another than sites that are far from one another.

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 choose to censor $Y(\mathbf{s})$ at $T = 0, 50$ (0.42 sample quantile), and 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 use a symmetric- t marginal distribution. We also compare models with no time series to models that include the time series. Finally, as a comparison to max-stable methods, we fit the model using the hierarchical max-stable model of Reich and Shaby (2012) with the data thresholded at $T = 75$. All methods assume the mean function given in (20). To ensure that the max-stable method runs in a reasonable amount of time, we take a stratified sample of the sites to get 800 sites and consider this

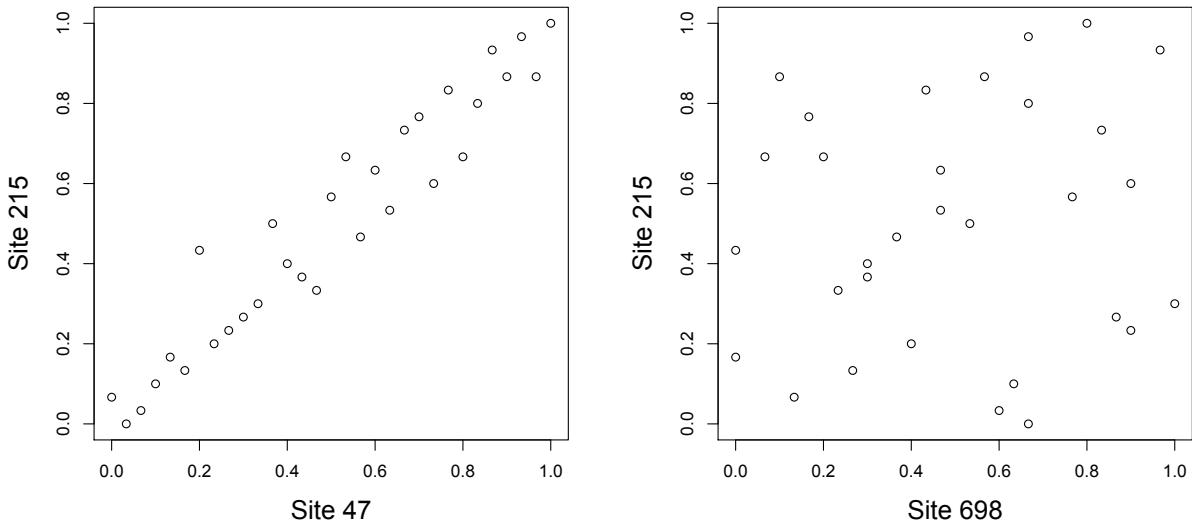


Figure 4: Daily quantiles for sites that are close (left) and far apart (right)

260 our new dataset. We conduct two-fold cross validation using 400 training sites and 400 validation sites as
 261 described in Section 5.2

262 Each chain of the MCMC ran for 30,000 iterations with a burn-in period of 25,000 iterations. Parameters
 263 appear to converge properly; however, as before, for models with multiple partitions it is hard to assess the
 264 convergence of \mathbf{w} , $z(\mathbf{s})$, and $\sigma^2(\mathbf{s})$ because of partition label switching throughout the MCMC. For each
 265 model, Brier scores were averaged over all sites and days to obtain a single Brier score for each dataset. At
 266 a particular threshold or quantile level, the model that fits the best is the one with the lowest score. We then
 267 compute the relative (to Gaussian) Brier scores (see Section 5.3) to compare each model.

268 **6.2 Results**

269 The results suggest that the skew- t , thresholded, partitioned, and time series models all give an improvement
 270 in predictions over the Gaussian model, whereas the max-stable method results in relative Brier scores
 271 between 1.07 and 1.15 indicating poorer performance than the Gaussian model. The plots in Figure 5

Table 1: Top two performing models for ozone analysis at extreme quantiles with Relative Brier score

		1st				2nd		
$q(0.90)$	No time series	$K = 7$	$T = 0$	BS: 0.980	No time series	$K = 9$	$T = 0$	BS: 0.980
$q(0.95)$	No time series	$K = 15$	$T = 50$	BS: 0.970	No time series	$K = 9$	$T = 50$	BS: 0.970
$q(0.98)$	No time series	$K = 5$	$T = 50$	BS: 0.945	No time series	$K = 10$	$T = 50$	BS: 0.946
$q(0.99)$	Time series	$K = 10$	$T = 75$	BS: 0.912	Time series	$K = 6$	$T = 75$	BS: 0.913
$q(0.995)$	Time series	$K = 6$	$T = 75$	BS: 0.917	Time series	$K = 10$	$T = 75$	BS: 0.918

272 show the relative Brier scores for time-series and non-time-series models, using $K = 1, 7$, and 15 knots at
 273 thresholds $T = 0, 50$, and 75 ppb. Most of the models perform similarly across all the Brier scores; however,
 274 for single-partition models without thresholding, performance tends to diminish in the extreme quantiles.
 275 The results also suggest that thresholding improves performance for estimates in the extreme quantiles. Both
 276 plots have similar features suggesting that most settings do reasonably well. In particular, for all extreme
 277 quantiles, selecting a moderate number of knots (e.g. $K = 5, \dots, 10$) tends to give the best results. Table 1
 278 shows the best two models for selected extreme quantiles.

279 We illustrate the predictive capability of our model in Figure 6 by plotting the 99th quantile for South
 280 Carolina and Georgia, a subset of the spatial domain, in order to study local features. The four methods used
 281 are

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

286 In the bottom two plots, we plot the differences between method 4 and methods 1 and 2. The most noticeable
 287 differences between the reference methods and the comparison methods is that the comparison methods tend
 288 to give higher estimates of the 99th quantile along the I-85 corridor between Charlotte and Atlanta.

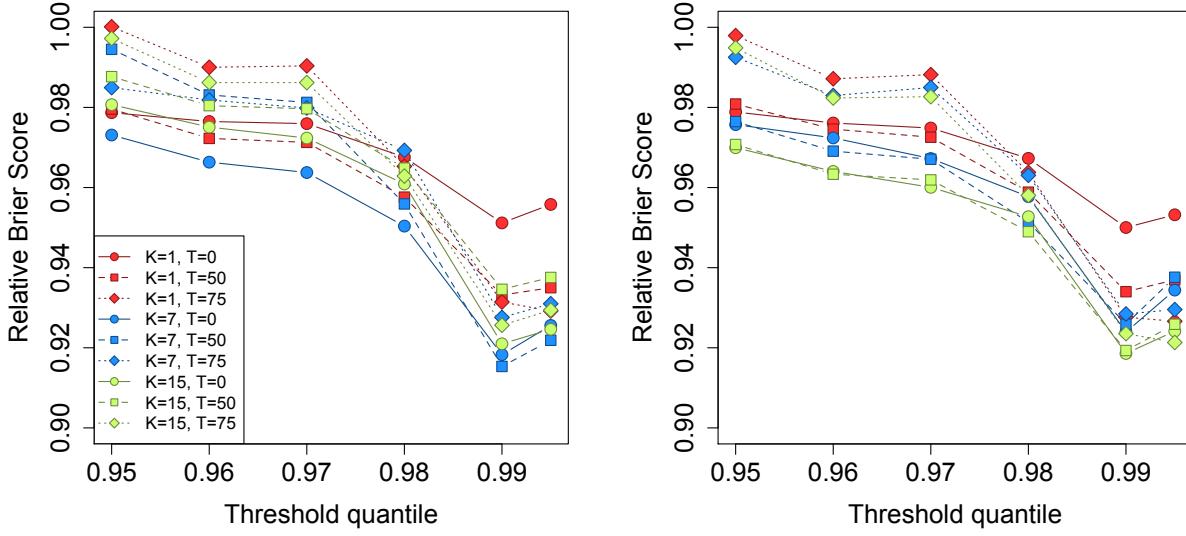


Figure 5: Relative Brier scores for time-series models (left) and non-time-series models (right). Relative brier score for the max-stable model is between 1.07 and 1.15

289 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 model shows statistically significant improvements over a naïve Gaussian approach. In both the simulation study, and the application to ozone data, we find that incorporating a partition in the model improves extreme prediction. 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

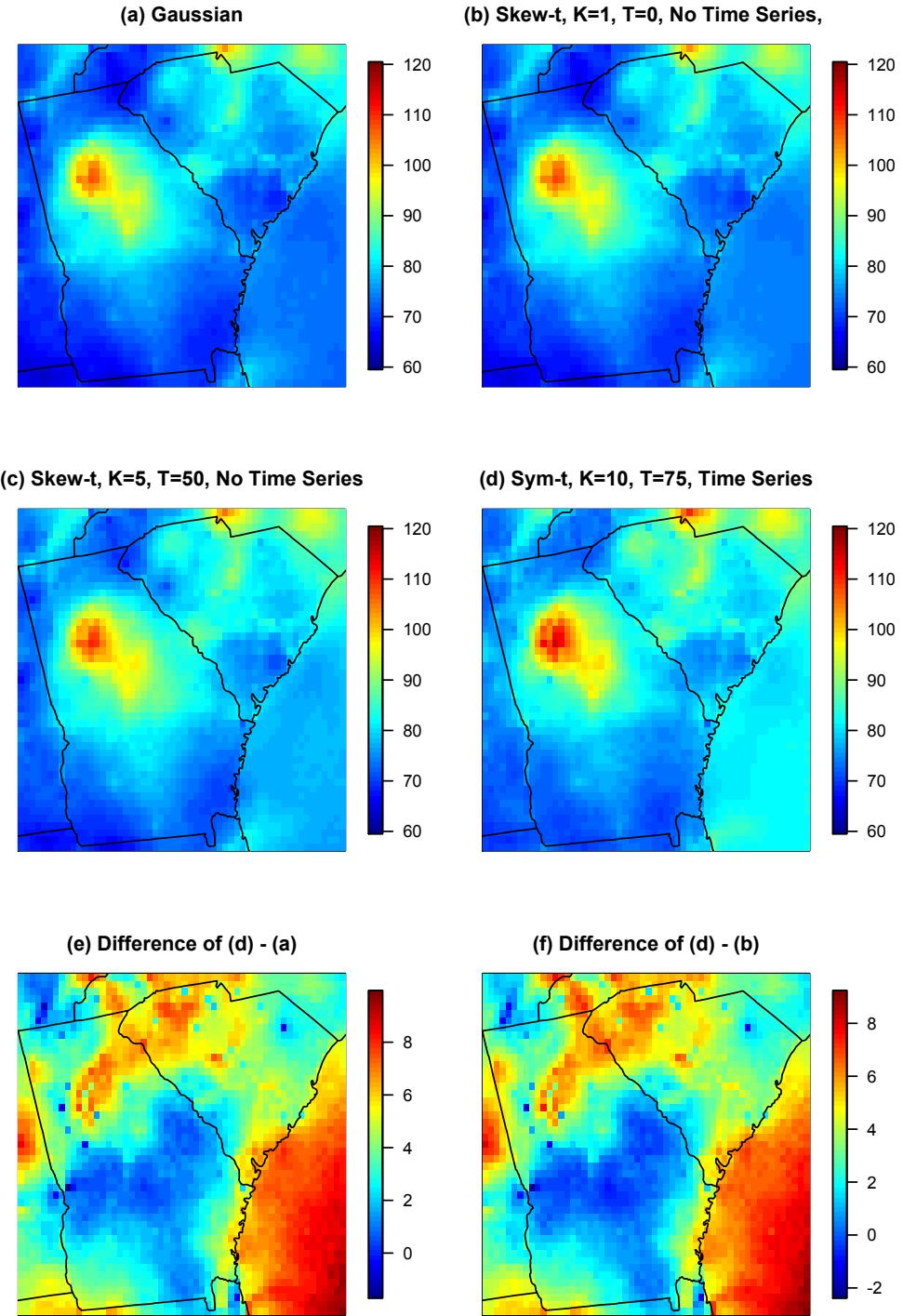


Figure 6: Panels (a) – (d) give the posterior predictive $\hat{q}(0.99)$ for the month of July under four different models, panel (e) gives the difference between $\hat{q}(0.99)$ in panels (d) and (a), panel (f) gives the difference between $\hat{q}(0.99)$ in panels (d) and (b).

301 for multiple knots to impact each site, with the weight of each knot to be determined by some characteristic
302 such as distance. Another area that should be explored is the temporal dependence in the model. Instead of
303 implementing a time series on the random effects, a three-dimensional covariance structure on the residuals
304 could be implemented to address temporal dependence. Finally, we acknowledge that by specifying the
305 number of knots, we may be underestimating the uncertainty in the model. This could be incorporated by
306 treating the number of knots as a model parameter instead of fixing it to be a specific value.

307 **Acknowledgments**

308 **A Appendices**

309 **A.1 MCMC details**

310 The MCMC sampling for the model 4 is done using R (<http://www.r-project.org>). Whenever possible,
311 we select conjugate priors (see Appendix A.2); however, for some of the parameters, no conjugate prior
312 distributions exist. When no conjugate prior distribution exists, we use a random walk Metropolis Hastings
313 update step. In each Metropolis Hastings update, we tune the algorithm to give acceptance rates near 0.40.

314 **Spatial knot locations**

315 For each day, we update the spatial knot locations, $\mathbf{w}_1, \dots, \mathbf{w}_K$, using a Metropolis Hastings block up-
316 date. Because the spatial domain is bounded, we generate candidate knots using the transformed knots
317 $\mathbf{w}_1^*, \dots, \mathbf{w}_K^*$ (see section 3.3) and a random walk bivariate Gaussian candidate distribution

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

³¹⁸ where $\mathbf{w}_k^{*(r-1)}$ is the location for the transformed knot at MCMC iteration $r - 1$, s is a tuning parameter,
³¹⁹ and I_2 is an identity matrix. After candidates have been generated for all K knots, the acceptance ratio is

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

³²⁰ where l is the likelihood given in (18), and $p(\cdot)$ is the prior either taken from the time series given in (3.3)
³²¹ or assumed to be uniform over \mathcal{D} . The candidate knots are accepted with probability $\min\{R, 1\}$.

³²² Spatial random effects

³²³ If there is no temporal dependence amongst the observations, we use a Gibbs update for z_{tk} , and the posterior
³²⁴ distribution is given in A.2. If there is temporal dependence amongst the observations, then we update z_{tk}
³²⁵ using a Metropolis Hastings update. Because this model uses $|z_{tk}|$, we generate candidate random effects
³²⁶ using the z_{tk}^* (see Section 3.3) and a random walk Gaussian candidate distribution

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

³²⁷ where $z_{tk}^{*(r-1)}$ is the value at MCMC iteration $r - 1$, and s is a tuning parameter. The acceptance ratio is

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

³²⁸ where $p[\cdot]$ is the prior taken from the time series given in Section 3.3. The candidate is accepted with
³²⁹ probability $\min\{R, 1\}$.

330 **Variance terms**

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

$$R = \left\{ \frac{l[Y_t(\mathbf{s}) | \sigma_{tk}^{2(c)}, \dots]}{l[Y_t(\mathbf{s}) | \sigma_{tk}^{2(r-1)}]} \right\} \times \left\{ \frac{l[z_{tk} | \sigma_{tk}^{2(c)}, \dots]}{l[z_{tk} | \sigma_{tk}^{2(r-1)}, \dots]} \right\} \times \left\{ \frac{p[\sigma_{tk}^{2(c)}]}{p[\sigma_{tk}^{2(r-1)}]} \right\} \times \left\{ \frac{c[\sigma_{tk}^{2(r-1)}]}{c[\sigma_{tk}^{2(c)}]} \right\}$$

336 where $p[\cdot]$ is the prior either taken from the time series given in Section 3.3 or assumed to be $IG(a, b)$, and
 337 $c[\cdot]$ is the candidate distribution. The candidate is accepted with probability $\min\{R, 1\}$.

338 **Spatial covariance parameters**

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

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

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

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

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

³⁴⁴ **A.2 Posterior distributions**

³⁴⁵ **Conditional posterior of $z_{tk} | \dots$**

³⁴⁶ If knots are independent over days, then the conditional posterior distribution of $|z_{tk}|$ is conjugate. For

³⁴⁷ simplicity, drop the subscript t , let $\tilde{z}_{tk} = |z_{tk}|$, and define

$$R(\mathbf{s}) = \begin{cases} Y(\mathbf{s}) - X(\mathbf{s})\beta & s \in P_l \\ Y(\mathbf{s}) - X(\mathbf{s})\beta - \lambda \tilde{z}(\mathbf{s}) & s \notin P_l \end{cases}$$

³⁴⁸ Let

$R_1 = \text{the vector of } R(\mathbf{s}) \text{ for } s \in P_l$

$R_2 = \text{the vector of } R(\mathbf{s}) \text{ for } s \notin P_l$

$$\Omega = \Sigma^{-1}.$$

³⁴⁹ Then

$$\begin{aligned} \pi(z_l | \dots) &\propto \exp \left\{ -\frac{1}{2} \left[\begin{pmatrix} R_1 - \lambda \tilde{z}_l \mathbf{1} \\ R_2 \end{pmatrix}^T \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} \begin{pmatrix} R_1 - \lambda \tilde{z}_l \mathbf{1} \\ R_2 \end{pmatrix} + \frac{\tilde{z}_l^2}{\sigma_l^2} \right] \right\} I(z_l > 0) \\ &\propto \exp \left\{ -\frac{1}{2} [\Lambda_l \tilde{z}_l^2 - 2\mu_l \tilde{z}_l] \right\} \end{aligned}$$

³⁵⁰ where

$$\mu_l = \lambda(R_1^T \Omega_{11} + R_2^T \Omega_{21})\mathbf{1}$$

$$\Lambda_l = \lambda^2 \mathbf{1}^T \Omega_{11} \mathbf{1} + \frac{1}{\sigma_l^2}.$$

³⁵¹ Then $\tilde{Z}_l | \dots \sim N_{(0,\infty)}(\Lambda_l^{-1} \mu_l, \Lambda_l^{-1})$

³⁵² **Conditional posterior of β | ...**

³⁵³ Let $\beta \sim N_p(0, \Lambda_0)$ where Λ_0 is a precision matrix. Then

$$\begin{aligned} \pi(\beta | \dots) &\propto \exp \left\{ -\frac{1}{2} \beta^T \Lambda_0 \beta - \frac{1}{2} \sum_{t=1}^T [\mathbf{Y}_t - X_t \beta - \lambda |z_t|]^T \Omega [\mathbf{Y}_t - X_t \beta - \lambda |z_t|] \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left[\beta^T \Lambda_\beta \beta - 2 \sum_{t=1}^T [\beta^T X_t^T \Omega (\mathbf{Y}_t - \lambda |z_t|)] \right] \right\} \\ &\propto N(\Lambda_\beta^{-1} \mu_\beta, \Lambda_\beta^{-1}) \end{aligned}$$

³⁵⁴ where

$$\begin{aligned} \mu_\beta &= \sum_{t=1}^T [X_t^T \Omega (\mathbf{Y}_t - \lambda |z_t|)] \\ \Lambda_\beta &= \Lambda_0 + \sum_{t=1}^T X_t^T \Omega X_t. \end{aligned}$$

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

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

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

358 where

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

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

360 **Conditional posterior of $\lambda | \dots$**

361 For convergence purposes we model $\lambda = \lambda_1 \lambda_2$ where

$$\lambda_1 = \begin{cases} +1 & \text{w.p.0.5} \\ -1 & \text{w.p.0.5} \end{cases} \quad (21)$$

$$\lambda_2^2 \sim IG(\alpha_\lambda, \beta_\lambda). \quad (22)$$

$$(23)$$

³⁶² Then

$$\begin{aligned}\pi(\lambda_2 | \dots) &\propto \lambda_2^{2(-\alpha_\lambda - 1)} \exp\left\{-\frac{\beta_\lambda}{\lambda_2^2}\right\} \prod_{t=1}^T \prod_{k=1}^K \frac{1}{\lambda_2} \exp\left\{-\frac{z_{tk}^2}{2\lambda_2^2 \sigma_{tk}^2}\right\} \\ &\propto \lambda_2^{2(-\alpha_\lambda - kt - 1)} \exp\left\{-\frac{1}{\lambda_2^2} \left[\beta_\lambda + \frac{z^2}{2\sigma_{tk}^2}\right]\right\}\end{aligned}$$

³⁶³ Then $\lambda_2 | \dots \sim IG\left(\alpha_\lambda + kt, \beta_\lambda + \frac{z^2}{2\sigma_{tk}^2}\right)$

³⁶⁴ **A.3 Proof that** $\lim_{h \rightarrow \infty} \pi(h) = 0$

³⁶⁵ Consider a homogeneous spatial Poisson process with intensity μ . Define A as the circle with center

³⁶⁶ $(\mathbf{s}_1 + \mathbf{s}_2)/2$ and radius $h/2$. Then \mathbf{s}_1 and \mathbf{s}_2 are in different partitions almost surely if two or more points are

³⁶⁷ in A . Let $N(A)$ be the number of points in A , and let

$$\mu(A) = \mu|A| = \mu\pi\left(\frac{h}{2}\right)^2 = \lambda h^2.$$

³⁶⁸ Then

$$\begin{aligned}P[N(A) \geq 2] &= 1 - P[N(A) = 0] - P[N(A) = 1] \\ &= 1 - \exp\{-\lambda h^2\} - \lambda h^2 \exp\{-\lambda h^2\} \\ &= 1 - (1 + \lambda h^2) \exp\{-\lambda h^2\}\end{aligned}$$

³⁶⁹ which goes to one as $h \rightarrow \infty$.

³⁷⁰ **A.4 Skew-*t* distribution**

³⁷¹ **Univariate extended skew-*t* distribution**

³⁷² We say that Y follow a univariate extended skew-*t* distribution with location $\xi \in \mathcal{R}$, scale $\omega > 0$, skew
³⁷³ parameter $\alpha \in \mathcal{R}$, extended parameter $\tau \in \mathcal{R}$, and degrees of freedom ν if has distribution function

$$f_{\text{EST}}(y) = \omega^{-1} \frac{f_T(z; \nu)}{F_T(\tau/\sqrt{1+\alpha^2}; \nu)} F_T \left[(\alpha z + \tau) \sqrt{\frac{\nu+1}{\nu+z^2}}; 0, 1, \nu+1 \right] \quad (24)$$

³⁷⁴ where $f_T(t; \nu)$ is a univariate Student's *t* with ν degrees of freedom, $F_T(t; \nu) = P(T < t)$, and $z = (y - \xi)/\omega$.

³⁷⁵ In the case that $\tau = 0$, then Y follows a univariate skew-*t* distribution.

³⁷⁶ **Multivariate skew-*t* distribution**

³⁷⁷ If $\mathbf{Z} \sim \text{ST}_d(0, \bar{\Omega}, \boldsymbol{\alpha}, \eta)$ is a d -dimensional skew-*t* distribution, and $\mathbf{Y} = \xi + \boldsymbol{\omega}\mathbf{Z}$, where $\boldsymbol{\omega} = \text{diag}(\omega_1, \dots, \omega_d)$,
³⁷⁸ then the density of Y at y is

$$f_y(\mathbf{y}) = \det(\boldsymbol{\omega})^{-1} f_z(\mathbf{z}) \quad (25)$$

³⁷⁹ where

$$f_z(\mathbf{z}) = 2t_d(\mathbf{z}; \bar{\Omega}, \eta) T \left[\boldsymbol{\alpha}^T \mathbf{z} \sqrt{\frac{\eta+d}{\nu+Q(\mathbf{z})}}; \eta+d \right] \quad (26)$$

$$\mathbf{z} = \boldsymbol{\omega}^{-1}(\mathbf{y} - \xi) \quad (27)$$

³⁸⁰ 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
³⁸¹ η , $Q(\mathbf{z}) = \mathbf{z}^T \bar{\Omega}^{-1} \mathbf{z}$ and $T(\cdot; \eta)$ denotes the univariate Student's *t* distribution function with η degrees of
³⁸² freedom (Azzalini and Capitanio, 2014).

383 **Extremal dependence**

384 For a bivariate skew- t random variable $\mathbf{Y} = [Y(\mathbf{s}), Y(\mathbf{t})]^T$, the $\chi(h)$ statistic (Padoan, 2011) is given by

$$\chi(h) = \bar{F}_{\text{EST}} \left\{ \frac{[x_1^{1/\eta} - \varrho(h)]\sqrt{\eta+1}}{\sqrt{1-\varrho(h)^2}}; 0, 1, \alpha_1, \tau_1, \eta + 1 \right\} + \bar{F}_{\text{EST}} \left\{ \frac{[x_2^{1/\eta} - \varrho(h)]\sqrt{\eta+1}}{\sqrt{1-\varrho(h)^2}}; 0, 1, \alpha_2, \tau_2, \eta + 1 \right\}, \quad (28)$$

385 where \bar{F}_{EST} is the univariate survival extended skew- t function with zero location and unit scale, $\varrho(h) = \text{cor}[y(\mathbf{s}), y(\mathbf{t})]$,

386 $\alpha_j = \alpha_i \sqrt{1 - \varrho^2}$, $\tau_j = \sqrt{\eta+1}(\alpha_j + \alpha_i \varrho)$, and $x_j = F_T(\bar{\alpha}_i \sqrt{\eta+1}; 0, 1, \eta)/F_T(\bar{\alpha}_j \sqrt{\eta+1}; 0, 1, \eta)$ with

387 $j = 1, 2$ and $i = 2, 1$ and where $\bar{\alpha}_j = (\alpha_j + \alpha_i \varrho)/\sqrt{1 + \alpha_i^2[1 - \varrho(h)^2]}$.

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

389 Consider the bivariate distribution of $\mathbf{Y} = [Y(\mathbf{s}), Y(\mathbf{t})]^T$, with $\varrho(h)$ given by (3). So, $\lim_{h \rightarrow \infty} \varrho(h) = 0$.

390 Then

$$\lim_{h \rightarrow \infty} \chi(h) = \bar{F}_{\text{EST}} \left\{ \sqrt{\eta+1}; 0, 1, \alpha_1, \tau_1, \eta + 1 \right\} + \bar{F}_{\text{EST}} \left\{ \sqrt{\eta+1}; 0, 1, \alpha_2, \tau_2, \eta + 1 \right\}. \quad (29)$$

391 Because the extended skew- t distribution is not bounded above, for all $\bar{F}_{\text{EST}}(x) = 1 - F_{\text{EST}} > 0$ for all

392 $x < \infty$. Therefore, for a skew- t distribution, $\lim_{h \rightarrow \infty} \chi(h) > 0$.

393 **A.5 Simulation study pairwise difference results**

394 The following tables show the methods that have significantly different Brier scores when using a Wilcoxon-

395 Nemenyi-McDonald-Thompson test. In each column, different letters signify that the methods have signifi-

396 cantly different Brier scores. For example, there is significant evidence to suggest that method 1 and method

397 4 have different Brier scores at $q(0.90)$, whereas there is not significant evidence to suggest that method 1

398 and method 2 have different Brier scores at $q(0.90)$. In each table group A represents the group with the
 399 lowest Brier scores. Groups are significant with a familywise error rate of $\alpha = 0.05$.

Table 2: Setting 1 – Gaussian marginal, $K = 1$ knot

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

Table 3: Setting 2 – Skew- t marginal, $K = 1$ knot

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

Table 4: Setting 3 – Skew- t marginal, $K = 5$ knots

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

400 References

- 401 Azzalini, A. and Capitanio, A. (2014) *The Skew-Normal and Related Families*. Institute of Mathematical Statistics Monographs. Cambridge University Press.
 402 URL <https://books.google.com/books?id=-tkaAgAAQBAJ>.
- 403
- 404 Coles, S., Heffernan, J. and Tawn, J. (1999) Dependence Measures for Extreme Value Analyses. *Extremes*, 2, 339–365. URL <http://link.springer.com/article/10.1023/A:1009963131610>.

Table 5: Setting 4 – Max-stable

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

Table 6: Setting 5 – Transformation below $T = q(0.80)$

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

- 406 Davison, A. C. and Smith, R. L. (1990) Models for exceedances over high thresholds (with Discussion).
 407 *Journal of the Royal Statistical Society. Series B (Methodological)*, **52**, 393–442.
- 408 Engelke, S., Malinowski, A., Kabluchko, Z. and Schlather, M. (2014) Estimation of Hüsler-Reiss distributions
 409 and Brown-Resnick processes. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 239–265.
- 410
- 411 Gneiting, T. and Raftery, A. E. (2007) Strictly Proper Scoring Rules, Prediction,
 412 and Estimation. *Journal of the American Statistical Association*, **102**, 359–378.
 413 URL <http://www.tandfonline.com/doi/abs/10.1198/016214506000001437>.
- 414 Huser, R. and Davison, A. C. (2014) Space-time modelling of extreme
 415 events. *Journal of the Royal Statistical Society: Series B (Statistical
 416 Methodology)*, **76**, 439–461. URL <http://arxiv.org/abs/1201.3245>
 417 <http://doi.wiley.com/10.1111/rssb.12035>.
- 418 Kim, H.-M., Mallick, B. K. and Holmes, C. C. (2005) Analyzing Nonstationary Spatial Data Using Piece-
 419 wise Gaussian Processes. *Journal of the American Statistical Association*, **100**, 653–668.
- 420 Ledford, A. W. and Tawn, J. a. (1996) Statistics for near independence in multivariate extreme values.
 421 *Biometrika*, **83**, 169–187.
 422 URL <http://biomet.oxfordjournals.org/cgi/content/abstract/83/1/169>.
- 423 Padoan, S. A. (2011) Multivariate extreme models based on underlying skew-
 424 and skew-normal distributions. *Journal of Multivariate Analysis*, **102**, 977–991.
 425 URL <http://linkinghub.elsevier.com/retrieve/pii/S0047259X11000157>.
- 426 Padoan, S. A., Ribatet, M. and Sisson, S. A. (2010) Likelihood-Based Inference for

- 427 Max-Stable Processes. *Journal of the American Statistical Association*, **105**, 263–277.
428 URL`http://www.tandfonline.com/doi/abs/10.1198/jasa.2009.tm08577.`
- 429 Reich, B. J. and Shaby, B. A. (2012) A hierarchical max-stable spatial model
430 for extreme precipitation. *The Annals of Applied Statistics*, **6**, 1430–1451.
431 URL`http://projecteuclid.org/euclid.aoas/1356629046.`
- 432 Thibaud, E., Mutzner, R. and Davison, a. C. (2013) Threshold modeling of extreme spatial rainfall. *Water
433 Resources Research*, **49**, 4633–4644.
- 434 Thibaud, E. and Opitz, T. (2013) Efficient inference and simulation for elliptical Pareto processes. 1–19.
- 435 Wadsworth, J. L. and Tawn, J. a. (2012) Dependence modelling for spatial extremes. *Biometrika*, **99**, 253–
436 272.
- 437 — (2014) Efficient inference for spatial extreme value processes associated to log-Gaussian random functions. *Biometrika*, **101**, 1–15.
438 URL`http://biomet.oxfordjournals.org/cgi/doi/10.1093/biomet/ast042.`