

1            **A space-time skew- $t$  model for threshold exceedances**

2            Samuel A Morris<sup>1</sup>, Brian J Reich<sup>1</sup>, Emeric Thibaud<sup>2</sup>, and Daniel Cooley<sup>2</sup>

3            September 2, 2015

4            **Abstract**

5            To assess the compliance of air quality regulations, the Environmental Protection Agency (EPA) must  
6            know if a site exceeds a pre-specified threshold. In the case of ozone, the threshold for compliance is  
7            fixed at 75 parts per billion, which is high, but not extreme at all locations. We present a new method  
8            based on the spatial skew- $t$  process. Our method incorporates a random partition to permit long-distance  
9            asymptotic independence while allowing for sites that are near one another to be asymptotically depen-  
10          dent, and we incorporate thresholding to allow the tails of the data to speak for themselves. We also  
11          introduce a transformed AR(1) time-series to allow for temporal dependence. Finally, our model allows  
12          for high-dimensional Bayesian inference that is comparable in speed to traditional geostatistical meth-  
13          ods for large datasets. We apply our method to an ozone analysis for July 2005, and find that our model  
14          improves over both Gaussian and max-stable methods in terms of predicting exceedances over a fixed  
15          threshold.

16          **Key words:** Skew- $t$ , random partition, MCMC, extreme value analysis, spatio-temporal

---

<sup>1</sup>North Carolina State University

<sup>2</sup>Colorado State University

17 **1 Introduction**

18 Epidemiological studies have linked air quality to public health concerns regarding morbidity and mortality  
19 (Samet et al., 2000). As a results, the Environmental Protection Agency (EPA) has developed a set of stan-  
20 dards to help reduce air pollution thereby improving air quality. Our study is motivated by an air pollution  
21 application where the focus is not on the average behavior, but instead the behavior over a fixed threshold  
22 determined by government regulation. More specifically, we consider the case of compliance for ozone. A  
23 site is said to be in compliance if the fourth highest daily maximum 8-hour concentration averaged over 3  
24 years does not exceed 75 parts per billion (ppb). Figure 1 shows the ozone levels from July 10, 2005, at 1089  
25 stations across the United States. We see a large area above the compliance level in the midwest covering  
26 Ohio, Indiana, Illinois, and parts of the surrounding states.

27 A spatial model for threshold exceedances warrants special consideration and standard spatial methods  
28 are likely to perform poorly. First, because we are interested only in high values, we want to “let the tail  
29 speak for itself”. That is, if we fit a model to the entire data set, low-to-moderate values would influence  
30 the fit of the overall model. As there are more of these values, they can unduly influence the distribution at  
31 the higher levels about which we are interested. Our inference method will only use data which exceed a  
32 threshold, and will impute data below the threshold, thereby tailoring the fit to the levels of interest. Second,  
33 likelihood-based spatial modeling typically assumes a Gaussian process, which is appropriate when mean  
34 behavior is of interest. However, the Gaussian distribution is light-tailed and symmetric, and therefore  
35 may be inappropriate for modeling data which does not share this tail behavior. Third, we aim to capture  
36 the dependence structure when ozone is at high levels, and dependence at these levels may not be well-  
37 represented by covariances which focus again on mean behavior. Asymptotic dependence/independence (see  
38 Section 2.2) are notions which describe how two random variables’ dependence behaves as one looks further  
39 into the joint tail. The Gaussian distribution always exhibits asymptotic independence, except in the case

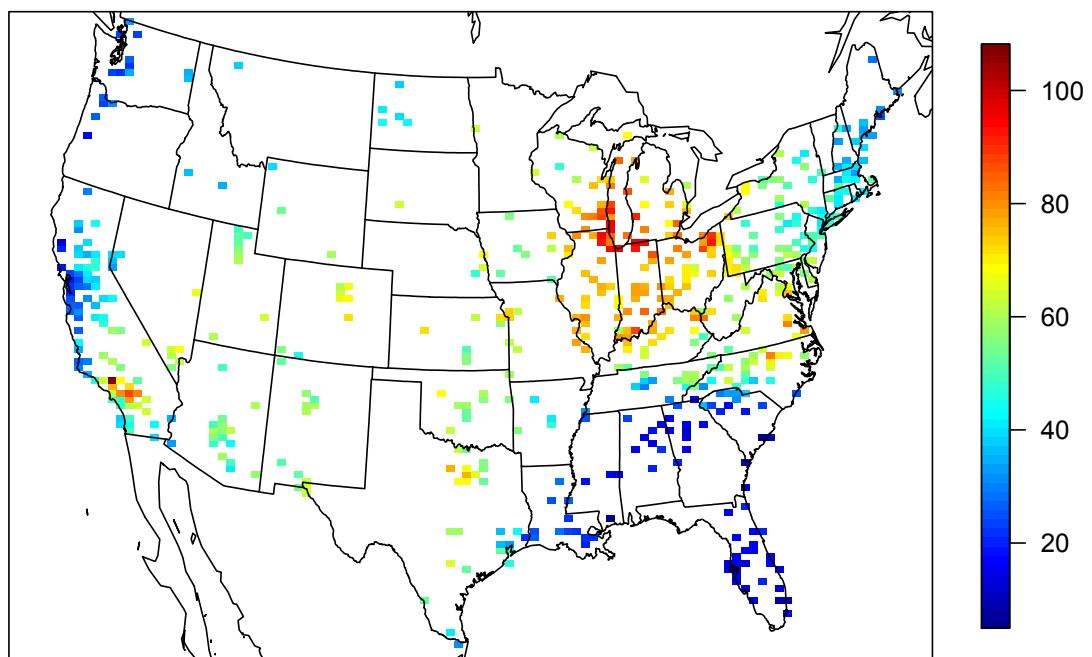


Figure 1: Ozone values on July 10, 2005

40 of perfect dependence, thus is an inappropriate model for data which exhibits asymptotic dependence. To  
41 allow for more flexibility in the marginal tail and to allow for asymptotic dependence, the skew-*t* distribution  
42 forms the basis for our model.

43 Our approach differs from threshold modeling approaches based on extreme value distributions. There  
44 has been extensive work on threshold modeling in the field of extreme value statistics where extreme events  
45 are naturally defined in terms of exceedances over a high threshold. Davison and Smith (1990) considered  
46 modeling threshold exceedances of univariate time series by the generalized Pareto distribution. Bivariate  
47 threshold models for extreme value distributions were considered by Ledford and Tawn (1996) who intro-  
48 duced a censored approach that provides a way to deal with different types of exceedances of a bivariate  
49 threshold in terms of only one or both components. These threshold models were extended to spatial models  
50 for extremes by Wadsworth and Tawn (2012) and Thibaud et al. (2013) who fit various models to spatial  
51 extremes using a censored pairwise likelihood (Padoan et al., 2010) based on the approach of Ledford and  
52 Tawn (1996). Huser and Davison (2014) further extended this to space-time modeling. Thibaud and Opitz  
53 (2013), Engelke et al. (2014), and Wadsworth and Tawn (2014), introduced more efficient inference for  
54 threshold exceedances of extremal spatial processes with full likelihood methods. The previous approaches  
55 to threshold modeling are motivated by extreme value theory and assume the threshold is high enough that  
56 extremal models are valid for the data, and for extrapolation beyond the range of observed values. More-  
57 over, these approaches are computationally intensive and limited to rather small datasets. For example,  
58 Wadsworth and Tawn (2014) present a simulation study with observations at 16 sites on a regular grid, and  
59 Engelke et al. (2014) analyze a dataset with observations at 35 meteorological stations. Our application  
60 with ozone data does not fit into this framework because we do not focus on exceedances of a very high  
61 threshold, but on exceedances of a fixed threshold. Furthermore, in our application, we have observations at  
62 over 1,000 ozone monitoring locations.

63 We propose a new spatiotemporal threshold exceedance model based on the skew-*t* process (Padoan,  
64 2011). Our model is a threshold exceedance model for the multivariate skew-*t* distribution for a fixed  
65 threshold. In this setting, we describe the threshold as fixed because it is specified in advance by regulatory  
66 compliance. This differs from the more traditional extremes literature where a threshold is selected to be  
67 the value beyond which an extremal model is appropriate for the data. We use a skew-*t* distribution because  
68 of its flexibility to model asymmetry and heavy-tailed data with the aim of predicting the probability of  
69 exceeding a high fixed threshold at an unobserved location.

70 Our model allows for inference and predictions using the full likelihood with computing on the order  
71 of Gaussian models for large space-time datasets. This allows us to use Bayesian methods to impute data  
72 below the threshold as well as make predictions at unobserved locations. In a spatial setting, the multivariate  
73 skew-*t* distribution demonstrates asymptotic dependence between observations at all sites regardless of the  
74 distance between the sites. In order to address this concern, we introduce a random spatial partition similar  
75 to the method used by Kim et al. (2005) for non-stationary Gaussian data. This partition alleviates the  
76 asymptotic spatial dependence present in the skew-*t* distribution for sites that are far apart.

77 The paper is organized as follows. Section 2 is a brief review of the spatial skew-*t* process. In Section  
78 3, we build upon the traditional skew-*t* process by incorporating censoring to focus on tails, partitioning to  
79 remove long-range asymptotic dependence, and extending the model to space-time data. The computing is  
80 described in Section 4. In Section 5, we present a simulation study that examines the predictive capabilities  
81 of this model compared Gaussian and max-stable methods. We then compare our method to Gaussian  
82 and max-stable methods with a data analysis of ozone measurements from 800 sites throughout the US in  
83 Section 6. The final section provides brief discussion and direction for future research.

84 **2 Spatial skew processes**

85 Many types of data demonstrate some level of skewness and therefore should be modeled with distributions  
 86 that allow for asymmetry. The skew-elliptical family of distributions provides models that are mathemati-  
 87 cally tractable while introducing a slant parameter to account for asymmetric data (Azzalini and Capitanio,  
 88 2014). A brief review of the additive process by which a skew- $t$  process is created is given here.

89 **2.1 Skew- $t$  process**

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

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

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

96 For a finite collection of locations  $\mathbf{s}_1, \dots, \mathbf{s}_n$ , denote the vector of observations  $\mathbf{Y} = [Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n)]^T$ .  
 97 After marginalizing over both  $z$  and  $\sigma$ ,

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

98 that is,  $\mathbf{Y}$  follows an  $n$ -dimensional skew- $t$  distribution with location  $\mathbf{X}\boldsymbol{\beta}$ , correlation matrix  $\boldsymbol{\Omega}$ , slant param-  
 99 eters  $\boldsymbol{\alpha}$  and degrees of freedom  $2a$ , where  $\mathbf{X} = [\mathbf{X}(\mathbf{s}_1)^T, \dots, \mathbf{X}(\mathbf{s}_n)^T]$ ,  $\boldsymbol{\Omega} = \boldsymbol{\omega} \bar{\boldsymbol{\Omega}} \boldsymbol{\omega}$ ,  $\boldsymbol{\omega} = \text{diag} \left( \frac{1}{\sqrt{ab}}, \dots, \frac{1}{\sqrt{ab}} \right)$ ,  
 100  $\bar{\boldsymbol{\Omega}} = (\boldsymbol{\Sigma} + \lambda^2 \mathbf{1} \mathbf{1}^T)$ ,  $\boldsymbol{\alpha} = \lambda(1 + \lambda^2 \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1})^{-1/2} \mathbf{1}^T \boldsymbol{\Sigma}^{-1}$ , and  $\boldsymbol{\Sigma}$  is the positive definite correlation matrix

101 of  $[v(\mathbf{s}_1), \dots, v(\mathbf{s}_n)]$ . This process is desirable because of its flexible tail that is controlled by the skewness  
 102 parameter  $\lambda$  and degrees of freedom  $2a$ . Furthermore, the marginal distributions at each location also follow  
 103 a univariate skew- $t$  distribution (Azzalini and Capitanio, 2014).

104 Although any positive definite correlation function could be used, we choose to use the stationary  
 105 isotropic Matérn correlation with

$$\text{cor}[v(\mathbf{s}_1), v(\mathbf{s}_2)] = \gamma I(\mathbf{s}_1 = \mathbf{s}_2) + (1 - \gamma) \frac{1}{\Gamma(\nu)^{2\nu-1}} \left( \sqrt{2\nu} \frac{h}{\rho} \right)^\nu K_\nu \left( \sqrt{2\nu} \frac{h}{\rho} \right) \quad (3)$$

106 where  $\rho$  is the spatial range,  $\nu$  is the smoothness,  $\gamma$  is the proportion of variance accounted for by the  
 107 spatial variation,  $K_\nu$  is a modified Bessel function of the second kind, and  $h = \|\mathbf{s}_1 - \mathbf{s}_2\|$ . We use this  
 108 parameterization for spatial correlation because the  $\gamma$  parameter permits the inclusion of a nugget effect to  
 109 account for non-spatial variability due to issues like measurement error.

## 110 2.2 Extremal dependence

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

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

114 where  $c^*$  is the upper limit of the support of  $Y$ . If  $\chi(h) = 0$ , then observations are asymptotically indepen-  
 115 dent at distance  $h$ . For Gaussian processes,  $\chi(h) = 0$  regardless of the distance  $h$ , so they are not suitable for  
 116 modeling asymptotically dependent extremes. Unlike the Gaussian process, the skew- $t$  process is asymptom-  
 117 ically dependent (the explicit expression for  $\chi(h)$  is given in Appendix A.4). However, one problem with

118 the spatial skew- $t$  process is that  $\lim_{h \rightarrow \infty} \chi(h) > 0$ . This occurs because all observations, both near and  
 119 far, share the same  $z$  and  $\sigma$  terms. Therefore, this long-range dependence feature of the skew- $t$  process is  
 120 not ideal for spatial analysis of large geographic regions where we expect only local spatial dependence. We  
 121 propose a solution to this in Section 3.2.

### 122 **3 Spatiotemporal skew- $t$ model for threshold exceedances**

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

#### 127 **3.1 Censoring to focus on the tails**

128 As mentioned previously, we propose to use a censored approach because we are interested in high values  
 129 and do not want the low-to-moderate values to influence the fit of the overall model. The censored observa-  
 130 tions below the threshold give information on the marginal probabilities to exceed the threshold and on the  
 131 dependence, but their values are not used to fit the model. Let

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

132 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  $T$   
 133 is a pre-specified threshold value. We impute the censored values as a step in the MCMC algorithm used to  
 134 fit the model described in Section 4.1.

<sup>135</sup> **3.2 Partitioning to remove long-range asymptotic dependence**

<sup>136</sup> The motivation for the partition is that for a large spatial domain, it may not be reasonable to assume sites  
<sup>137</sup> that are far apart demonstrate asymptotic dependence. Modeling different levels of asymptotic dependence  
<sup>138</sup> was discussed by Wadsworth and Tawn (2012) with a hybrid spatial dependence model. Huser and Davison  
<sup>139</sup> (2014) also allow for varying asymptotic dependence across both space and time with a partition structure  
<sup>140</sup> represented by random discs moving across the space for a random duration with a random velocity and  
<sup>141</sup> random radius. We handle the problem of long-range asymptotic dependence with a random partition. As  
<sup>142</sup> discussed in Section 2, the source of long-range dependence is the shared  $z$  and  $\sigma$ . Therefore, to alleviate  
<sup>143</sup> this dependence, we allow  $z$  and  $\sigma$  to vary by site. The model becomes

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

<sup>144</sup> To model spatial variation, consider a set of spatial knots  $\mathbf{w}_1, \dots, \mathbf{w}_K$  from a homogeneous Poisson process  
<sup>145</sup> with intensity  $\mu$  over spatial domain  $\mathcal{D} \in \mathbb{R}^2$ . The knots define a random partition of  $\mathcal{D}$  by subregions  
<sup>146</sup>  $P_1, \dots, P_K$  defined as

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

<sup>147</sup> All  $z(\mathbf{s})$  and  $\sigma(\mathbf{s})$  for sites in subregion  $k$  are assigned common values

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

<sup>148</sup> and the  $z_k$  and  $\sigma_k^2$  are distributed as  $z_k \stackrel{iid}{\sim} N(0, 1)$  and  $\sigma_k^2 \stackrel{iid}{\sim} \text{IG}(a, b)$ . So, within each partition,  $Y(\mathbf{s})$   
<sup>149</sup> follows the spatial skew- $t$  process defined in Section 2. Across partitions, the  $Y(\mathbf{s})$  remain correlated via the

150 correlation function for  $v(\mathbf{s})$  because  $v(\mathbf{s})$  spans all partitions.

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

152 for two sites  $\mathbf{s}$  and  $\mathbf{t}$  in partitions  $k_s$  and  $k_t$  respectively is

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

153 where  $I(\cdot)$  is an indicator function,  $\chi_{\text{skew-}t}(h)$  is the  $\chi$  statistic for a skew- $t$  process given in (28),  $\chi_{\text{Gaus}}(h)$

154 is the  $\chi$  statistic for a Gaussian process, and  $h = \|\mathbf{s} - \mathbf{t}\|$ . Therefore, sites in different subregions are asymptotically independent because  $\chi_{\text{Gaus}}(h) = 0$  for all  $h$ . Marginally, over the knots,  $\chi(h) = \pi(h)\chi_{\text{skew-}t}(h)$ ,

156 where  $\pi(h) = \Pr(k_s = k_t)$  is the probability that two sites separated by distance  $h$  are in the same partition.

157 In Appendix A.3, we show that  $\lim_{h \rightarrow \infty} \pi(h) = 0$ , implying  $\lim_{h \rightarrow \infty} \chi(h) = 0$ . In Figure 2, we give  $\chi(h)$

158 for  $K = 1, 3, 5, 10$  partitions for a skew- $t$  distribution with  $\alpha = 10$ , and 3 degrees of freedom. We estimate

159  $\pi(h)$  through simulation.

### 160 3.3 Extension to space-time data

161 When using daily measurements, the assumption of temporal independence is often inappropriate. In this

162 section, we extend (6) to the spatiotemporal setting. There are several places where temporal dependence

163 could be incorporated in the model, including the residuals  $v_t(\mathbf{s})$ . However, we choose to allow for temporal

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

165 Let

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

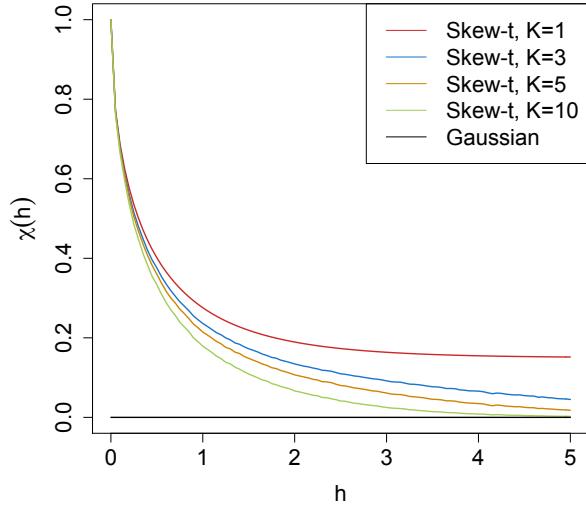


Figure 2:  $\chi$ , as a function of distance, for  $K = 1, 3, 5$ , and 10 knots.

166 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  
 167  $t$ , 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  
 168 partition  $P_{t1}, \dots, P_{tK}$ , and for  $\mathbf{s} \in P_{tk}$ ,

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

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

171 We use an AR(1) time series model for  $w_{tk}$ ,  $z_{tk}$ , and  $\sigma_{tk}$ . The time series model must be specified after  
 172 a transformation to preserve the skew- $t$  process at each time point. For each time-varying parameter, we  
 173 transform to obtain a standard normal marginal distribution, place a Gaussian prior with autocorrelation on  
 174 the transformed parameter, and then transform back to the appropriate marginal distribution for the skew- $t$

<sup>175</sup> process. We first transform the spatial knots from  $\mathcal{D}$  to  $\mathcal{R}^2$  as follows. Let

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

<sup>176</sup> where  $\Phi$  is a univariate standard normal density function, and  $\mathbf{s}_i = [s_{1i}, \dots, s_{ni}]$ . Then the transformed  
<sup>177</sup> 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  
<sup>178</sup> gamma. Let

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

<sup>179</sup> where IG is defined as before. We also use a copula on  $z_t(\mathbf{s})$  to ensure that the marginal distributions of  
<sup>180</sup>  $z_t(\mathbf{s})$  are half-normal. Let

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

<sup>181</sup> where HN is the distribution function of a half-normal random variable. The AR(1) process for each tail  
<sup>182</sup> 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

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

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

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

<sup>183</sup> where  $|\phi_w|, |\phi_z|, |\phi_\sigma| < 1$ . These are stationary time series models with marginal distributions  $\mathbf{w}_k^* \sim N_2(0, 1)$ ,  
<sup>184</sup>  $z_k^* \sim N(0, \sigma_k^2)$ , and  $\sigma_k^{2*} \sim N(0, 1)$ . After transformation back to the original space,  $\mathbf{w}_{tk} \sim \text{Unif}(\mathcal{D})$ ,

<sup>185</sup>  $z_{tk} \sim HN(0, \sigma_{tk}^2)$   $\sigma_{tk}^2 \sim IG(a, b)$ . For each day, the model is identical to the spatial-only model in (6)  
<sup>186</sup> by construction.

## <sup>187</sup> 4 Hierarchical model

<sup>188</sup> Conditioned on  $z_t(\mathbf{s})$ ,  $\sigma_t^2(\mathbf{s})$ , and  $P_{tk}$ , the marginal distributions are Gaussian and the joint distribution  
<sup>189</sup> multivariate Gaussian. However, we do not fix the partitions, they are treated as unknown and updated in  
<sup>190</sup> the MCMC algorithm. We model this with a Bayesian hierarchical model as follows. Let  $\mathbf{w}_{t1}, \dots, \mathbf{w}_{tK}$  be  
<sup>191</sup> a set of daily spatial knots in a spatial domain of interest,  $\mathcal{D}$ , and  $P_{tk}$  as defined in (7). In practice, we fix  $K$   
<sup>192</sup> at different levels, and assess its impact on prediction as described in 5.2. Then

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

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

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

$$\lambda = \lambda_1 \lambda_2$$

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

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

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

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

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

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

193 where  $\Theta = \{\rho, \nu, \gamma, \lambda, \beta\}$ , and  $\Sigma$  is a Matérn covariance matrix as described in Section 2.1. We parameterize  
194  $\lambda = \lambda_1 \lambda_2$  to help with convergence in the MCMC.

195 **4.1 Computation**

196 We use MCMC methods to explore the posterior. At each MCMC iteration, we first impute values below  
197 the threshold conditional on observations above the threshold. This is feasible for large datasets with our  
198 model because for a single day, conditional on the model parameters, we only need to draw from a truncated  
199 multivariate normal distribution. We can use Gibbs sampling to update  $Y_t(\mathbf{s})$  for censored observations that  
200 are below the threshold  $T$ . After conditioning on  $\lambda$ ,  $z_t(\mathbf{s})$  and non-censored observations,  $Y_t(\mathbf{s})$  has truncated  
201 normal full conditionals. So we sample  $Y_t(\mathbf{s}) \sim N_{(-\infty, T)}(\mathbf{X}_t^T(\mathbf{s})\beta + \lambda|z_t(\mathbf{s})|, \Sigma)$ .

202 Then, we update model parameters,  $\Theta$ , using a Metropolis-Hastings algorithm with Gibbs sampling  
203 when needed. The final step of the computation is to use Bayesian Kriging to generate a predictive distri-  
204 bution for  $Y_t(\mathbf{s}^*)$  at prediction location  $\mathbf{s}^*$ . This step is similar to the imputation for censored observations  
205 except that the full conditionals are no longer truncated at  $T$ . See Appendices A.1 and A.2 for details  
206 regarding the MCMC algorithm.

207 **5 Simulation study**

208 In this section, we present the results from a simulation study to investigate how the number of partitions  
209 and the level of thresholding impact the accuracy of predictions made by the model.

210 **5.1 Design**

211 For all simulation designs, we generated data from model (6) in Section 3.2 using  $n_s = 144$  sites and  
212  $n_t = 50$  independent days. The sites were generated Uniform([0, 10]  $\times$  [0, 10]). We generated data from 4

213 different simulation designs:

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

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

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

217 4. Max-stable

218 In the first three designs, the  $v_t(\mathbf{s})$  terms were generated using a Matérn covariance with smoothness param-  
219 eter  $\nu = 0.5$  and spatial range  $\rho = 1$ . For the covariance matrices in designs 1 – 3, the proportion of the  
220 variance accounted for by the spatial variation was  $\gamma = 0.9$  while the proportion of the variance accounted  
221 for by the nugget effect was 0.1. In the first design,  $\sigma^2 = 2$  was used for all days which results in a Gaussian  
222 distribution. 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  
223 design 1, we set  $\lambda = 0$ . For designs 2 and 3,  $\lambda = 3$  was used as to simulate moderate skewness, and the  
224  $z_t$  were generated as described in (8). In designs 1 – 3, the mean  $\mathbf{X}^T \boldsymbol{\beta} = 10$  was assumed to be constant  
225 across space. In the fourth design, we generated from a spatial max-stable distribution (Reich and Shaby,  
226 2012). In this design, data have marginal distributions that follow a generalized extreme value distribution  
227 with location parameter 1, scale parameter 1, and shape parameter 0.2. In this model, a random effect was  
228 used to induce spatial dependence using 144 spatial knots on a regular lattice in the square  $[1, 9] \times [1, 9]$ .  
229 For this setting,  $\gamma_{RS} \in (0, 1)$ , is similar to  $\gamma$  in that it controls the relative strength of the nugget effect. We  
230 set  $\gamma_{RS} = 0.5$ , which represents moderate spatial dependence.

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

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

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

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

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

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

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

238 where  $q(0.80)$  is the 80th sample quantile of the data. The design matrix  $\mathbf{X}$  includes an intercept with a first-

239 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

240  $\log(\nu) \sim N(-1.2, 1)$ ,  $\gamma \sim \text{Unif}(0, 1)$ ,  $\rho \sim \text{Unif}(15)$ . The skewness parameter has prior  $\lambda_2 \sim \text{IG}(0.1, 0.1)$ .

241 The residual variance terms have priors  $\sigma_t^2(\mathbf{s}) \sim \text{IG}(a, b)$ , where  $a$  has a  $\text{Gamma}(0.1, 0.1)$  prior and  $b$  has a

242 discrete uniform prior on a mesh from 0.1 to 10 with spacing of 0.1. The knots have priors  $\mathbf{w} \sim \text{Unif}(\mathcal{D})$ .

243 We tried also fitting the skew- $t$  marginals for the thresholded models, but it is very challenging for the

244 MCMC to properly identify the skewness parameter with a censored left tail. Each chain of the MCMC ran

245 for 20,000 iterations with a burn-in period of 10,000 iterations. Parameters appear to converge properly;

246 however, in the models with multiple partitions (i.e. models 4 and 5) it is hard to assess the convergence of

247  $\mathbf{w}$ ,  $z(\mathbf{s})$ , and  $\sigma^2(\mathbf{s})$  because of partition label switching throughout the MCMC.

248 **5.2 Cross validation**

249 Models were compared using cross validation, with 100 sites used as training sites to fit the models, and

250 44 sites withheld for testing the predictions. Because one of the primary goals of this model is to predict

251 exceedances over a fixed threshold, we use Brier scores to compare the models (Gneiting and Raftery, 2007).

252 The Brier score for predicting exceedance of a threshold  $c$  is given by  $[e(c) - P(c)]^2$  where  $e(c) = I[y > c]$

253 is an indicator function indicating that a test set value,  $y$ , has exceeded the threshold,  $c$ , and  $P(c)$  is the

254 predicted probability of exceeding  $c$ . We average the Brier scores over all test sites and days. For the Brier

255 score, a lower score indicates a better fit.

256 **5.3 Results**

257 We compared the Brier scores for exceeding 4 different thresholds for each dataset. The thresholds used for  
258 the Brier scores are extreme quantiles from the simulated data for  $q(0.90)$ ,  $q(0.95)$ ,  $q(0.98)$ , and  $q(0.99)$ .  
259 Figure 3 gives the Brier score relative to the Brier score for the Gaussian method calculated as

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

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

265 The results show that when the data are generated from a Gaussian process, our method performs com-  
266 parably to a Gaussian approach. In general, when the underlying process is not Gaussian, our method results  
267 in an improvement over both the max-stable and Gaussian methods. One exception to this is the case when  
268 the generative process is max-stable. In this case, the max-stable method outperforms our method; however,  
269 for predictions at high quantile levels, the differences between the max-stable method and our method de-  
270 crease. The non-thresholded methods tend to outperform the thresholded methods, but this is not surprising  
271 given that in most cases, the data are generated directly from the model used in the method. In summary, our  
272 method provides more flexibility for data that demonstrate some level of asymmetry or heavy tails, while  
273 still performing comparably to Gaussian methods when the data are symmetric and have light tails.

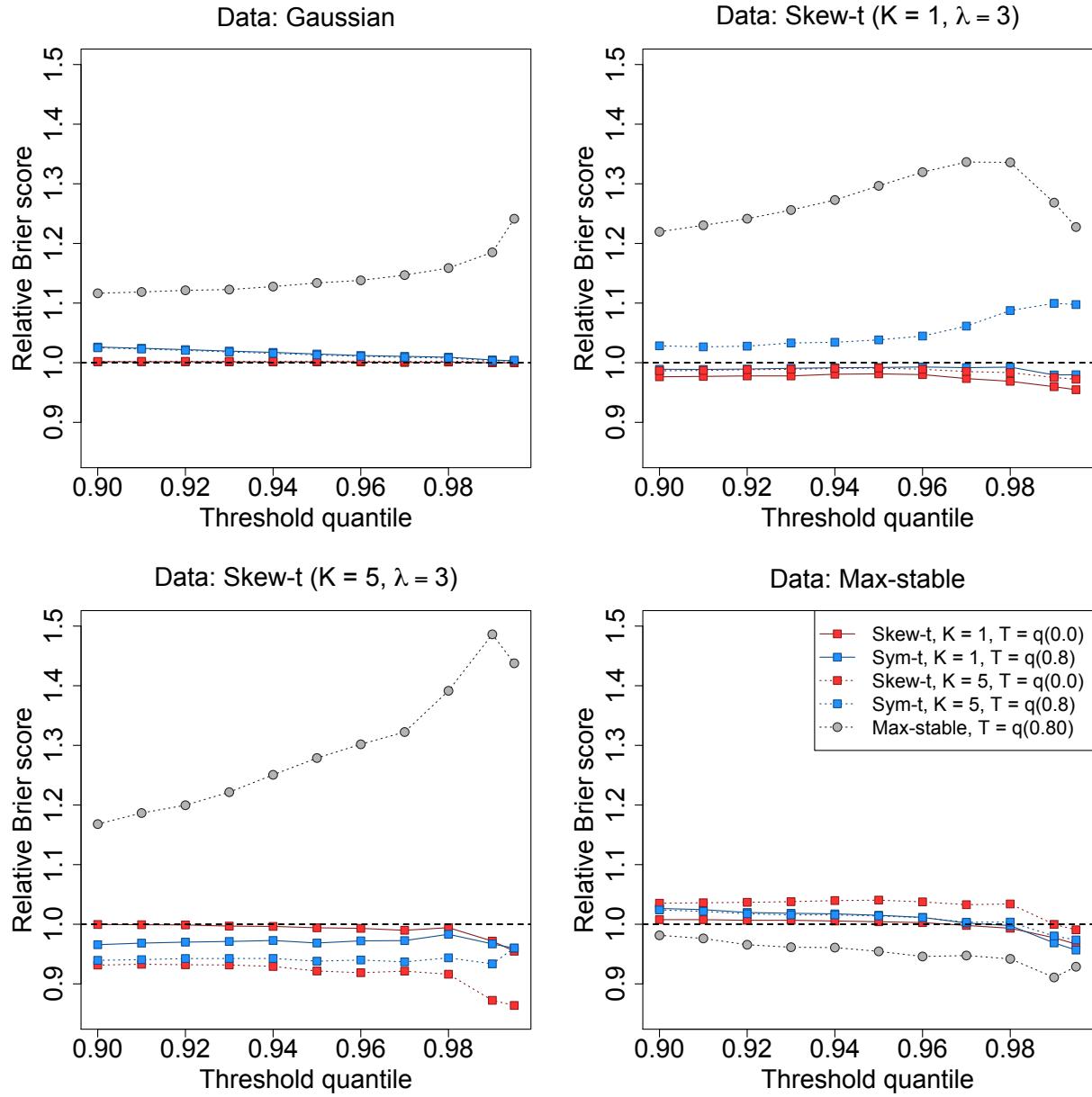


Figure 3: Brier scores relative to the Gaussian method for simulation study results. A ratio lower than 1 indicates that the method outperforms the Gaussian method.

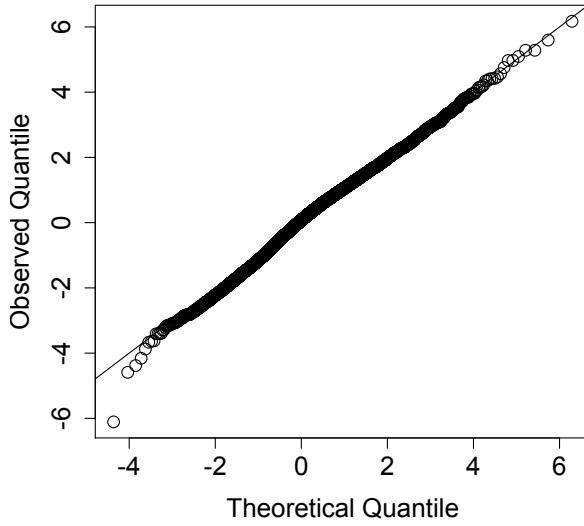


Figure 4: Q-Q plot of the residuals for a skew- $t$  distribution with 10 d.f. and  $\alpha = 1$  (right)

## 274 6 Data analysis

275 We consider daily observations of maximum 8-hour ozone measurements for the 31 days of July 2005 at  
 276 1,089 Air Quality System (AQS) monitoring sites in the United States as the response (see Figure 1). For  
 277 each site, we also have covariate information containing the estimated ozone from the Community Multi-  
 278 scale Air Quality (CMAQ) modeling system. Initially, we fit a linear regression assuming a mean function  
 279 of

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

280 Figure 4 shows a Q-Q plot of the residuals compared to a skew- $t$  distribution with 10 d.f. and  $\alpha = 1$ .  
 281 Standard exploratory data analysis techniques for extremal dependence are very challenging with only  
 282 31 days worth of data because it is difficult to estimate extreme quantiles at each site to obtain empirical  
 283 estimates of  $\chi$ . Despite the fact that there is only one month of data, we can get some sense of extremal

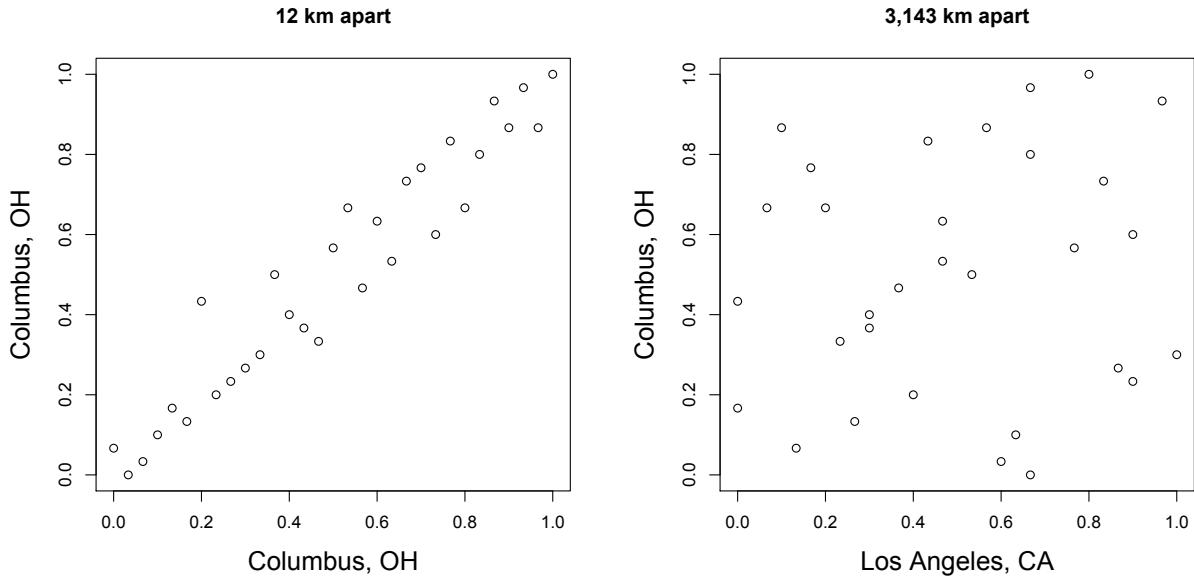


Figure 5: Daily quantiles for two monitoring locations near Columbus, OH (left) and daily quantiles for a monitoring location in Los Angeles, CA and Columbus, OH (right)

dependence between sites by looking at joint occurrences of high sample quantiles. For example, Figure 5 suggests there is more agreement between sites that are close to one another than sites that are far from one another. Another aspect that distinguishes our approach from more traditional extremes analyses, is how the threshold is selected. In our example, a threshold of 75 ppb which corresponds to  $q(92)$  for all observations, but marginally it represents anywhere from  $q(0.06)$  to  $q(1)$ .

## 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  $T = 0$ ,  $T = 50$  (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 use a symmetric- $t$  marginal distribution. We also compare

295 models with no time series to models that include the time series. Finally, as a comparison to max-stable  
296 methods, we fit the model using the hierarchical max-stable model of Reich and Shaby (2012) with the data  
297 thresholded at  $T = 75$ . All methods assume the mean function given in (20). To ensure that the max-  
298 stable method runs in a reasonable amount of time, we use a stratified sub-sample of 800 sites. We conduct  
299 two-fold cross validation using 400 training sites and 400 validation sites as described in Section 5.2

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

## 306 6.2 Results

307 The results suggest that the skew- $t$ , thresholded, partitioned, and time series models all give an improvement  
308 in predictions over the Gaussian model, whereas the max-stable method results in relative Brier scores  
309 between 1.07 and 1.15 indicating poorer performance than the Gaussian model. The plots in Figure 6  
310 show the relative Brier scores for time-series and non-time-series models, using  $K = 1, 7$ , and 15 knots at  
311 thresholds  $T = 0, 50$ , and 75 ppb. Most of the models perform similarly across all the Brier scores; however,  
312 for single-partition models without thresholding, performance tends to diminish in the extreme quantiles.  
313 The results also suggest that thresholding improves performance for estimates in the extreme quantiles. Both  
314 plots have similar features suggesting that most settings do reasonably well. In particular, for all extreme  
315 quantiles, selecting a moderate number of knots (e.g.  $K = 5, \dots, 10$ ) tends to give the best results. Table 1  
316 shows the best two models for selected extreme quantiles.

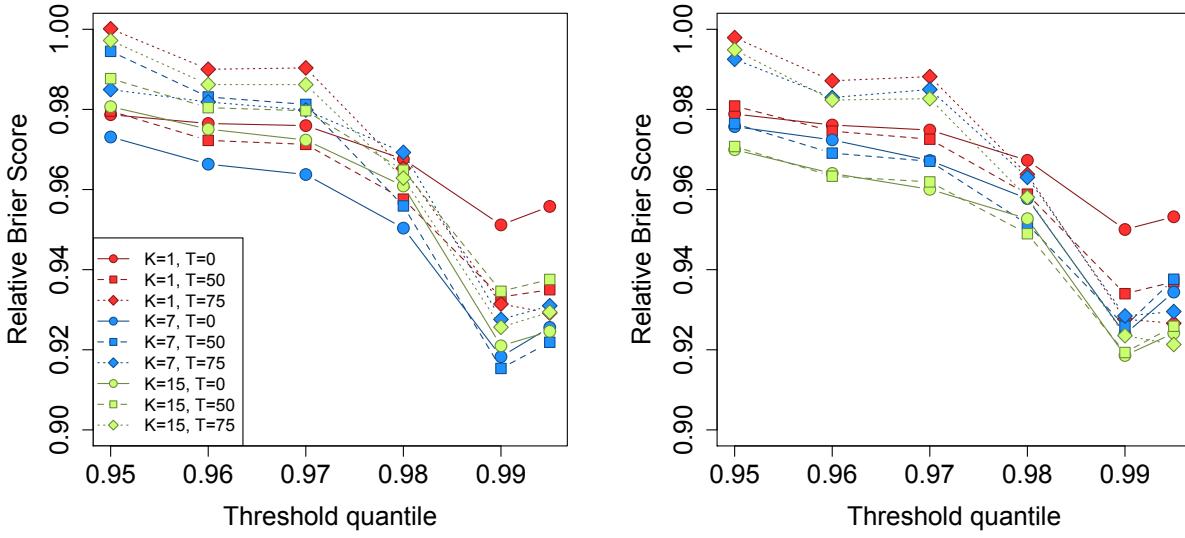


Figure 6: Relative Brier scores for time-series models (left) and non-time-series models (right). Relative brier score for the max-stable model is between 1.07 and 1.15

317 We illustrate the predictive capability of our model in Figure 7 by plotting the 99th quantile for South

318 Carolina and Georgia, a subset of the spatial domain, in order to study local features. The four methods used

319 are

320 1. Gaussian

321 2. Skew- $t$ ,  $K = 1$  knot,  $T = 0$ , no time series

322 3. Skew- $t$ ,  $K = 5$  knots,  $T = 50$ , no time series

323 4. Symmetric- $t$ ,  $K = 10$  knots,  $T = 75$ , time series.

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

325 differences between the reference methods and the comparison methods is that the comparison methods tend

326 to give higher estimates of the 99th quantile along the I-85 corridor between Charlotte and Atlanta.

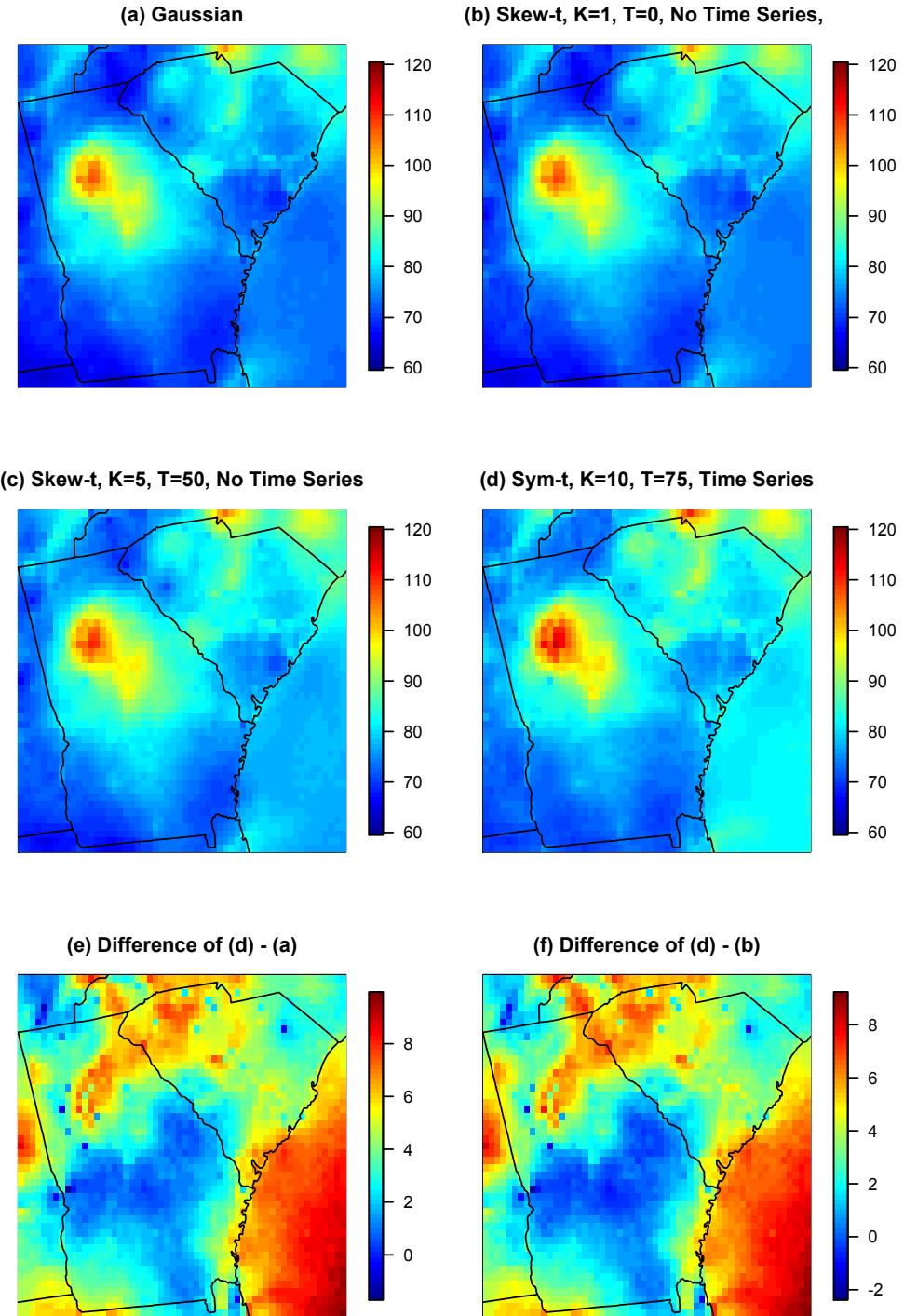


Figure 7: Panels (a) – (d) give the posterior predictive  $\hat{q}(0.99)$  for the month of July under four different models, panel (e) gives the difference between  $\hat{q}(0.99)$  in panels (d) and (a), panel (f) gives the difference between  $\hat{q}(0.99)$  in panels (d) and (b).

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

## 327 7 Discussion

328 In this paper we propose a new threshold exceedance approach for spatiotemporal modeling based on the  
 329 skew- $t$  process. The proposed model gives flexible tail behavior, demonstrates asymptotic dependence for  
 330 observations at sites that are near to one another, and has computation on the order of Gaussian models  
 331 for large space-time datasets. In the simulation study, we demonstrate that this model shows statistically  
 332 significant improvements over a naïve Gaussian approach and in most cases, a max-stable approach. In both  
 333 the simulation study, and the application to ozone data, we find that incorporating a partition in the model  
 334 can improve extreme predictions. Furthermore the results from the data analysis suggest that thresholding  
 335 can improve performance when predicting in the extreme tails of the data.

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

345 **Acknowledgments**

346 The authors' work was partially supported by grants from the Department of the Interior (14-1-04-9), Na-  
347 tional Institutes of Health (R21ES022795-01A1), the US Environmental Protection Agency (R835228), the  
348 National Science Foundation (1107046), and the Research Network for Statistical Methods for Atmospheric  
349 and Oceanic Sciences (STATMOS). The authors also received high-performance computing support from  
350 Yellowstone (ark:/85065/d7wd3xhc) provided by NCAR's Computational and Information Systems Labo-  
351 ratory, sponsored by the National Science Foundation.

352 **A Appendices**

353 **A.1 MCMC details**

354 The MCMC sampling for the model 4 is done using R (<http://www.r-project.org>). Whenever possible,  
355 we select conjugate priors (see Appendix A.2); however, for some of the parameters, no conjugate prior  
356 distributions exist. For these parameters, we use a random walk Metropolis-Hastings update step. In each  
357 Metropolis-Hastings update, we tune the algorithm during the burn-in period to give acceptance rates near  
358 0.40.

359 **Spatial knot locations**

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

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

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

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

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

### 367 Spatial random effects

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

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

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

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

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

375 **Variance terms**

376 When there is more than one site in a partition, then we update  $\sigma_{tk}^2$  using a Metropolis-Hastings update.  
 377 First, we generate a candidate for  $\sigma_{tk}^2$  using an  $IG(a^*/s, b^*/s)$  candidate distribution in an independence  
 378 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  
 379 of sites in partition  $k$  on day  $t$ , and  $Y_{tk}$  and  $\Sigma_{tk}^{-1}$  are the observations and precision matrix for partition  $k$  on  
 380 day  $t$ . The acceptance ratio is

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

381 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  
 382  $c[\cdot]$  is the candidate distribution. The candidate is accepted with probability  $\min\{R, 1\}$ .

383 **Spatial covariance parameters**

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

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

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

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

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

389 **A.2 Posterior distributions**

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

391 If knots are independent over days, then the conditional posterior distribution of  $|z_{tk}|$  is conjugate. For  
392 simplicity, drop the subscript  $t$ , let  $\tilde{z}_{tk} = |z_{tk}|$ , and define

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

393 Let

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

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

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

394 Then

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

<sup>395</sup> where

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

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

<sup>396</sup> Then  $\tilde{Z}_l | \dots \sim N_{(0,\infty)}(\Lambda_l^{-1} \mu_l, \Lambda_l^{-1})$

<sup>397</sup> **Conditional posterior of  $\beta$  | ...**

<sup>398</sup> Let  $\beta \sim N_p(0, \Lambda_0)$  where  $\Lambda_0$  is a precision matrix. Then

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

<sup>399</sup> where

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

400 **Conditional posterior of  $\sigma^2$  | ...**

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

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

403 where

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

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

405 **Conditional posterior of  $\lambda$  | ...**

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

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

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

$$(23)$$

<sup>407</sup> Then

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

<sup>408</sup> Then  $\lambda_2 | \dots \sim IG\left(\alpha_\lambda + kt, \beta_\lambda + \frac{z^2}{2\sigma_{tk}^2}\right)$

<sup>409</sup> **A.3 Proof that**  $\lim_{h \rightarrow \infty} \pi(h) = 0$

<sup>410</sup> Consider a homogeneous spatial Poisson process with intensity  $\mu$ . Define  $A$  as the circle with center

<sup>411</sup>  $(\mathbf{s}_1 + \mathbf{s}_2)/2$  and radius  $h/2$ . Then  $\mathbf{s}_1$  and  $\mathbf{s}_2$  are in different partitions almost surely if two or more points are

<sup>412</sup> in  $A$ . Let  $N(A)$  be the number of points in  $A$ , and let

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

<sup>413</sup> Then

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

<sup>414</sup> which goes to one as  $h \rightarrow \infty$ .

415 **A.4 Skew- $t$  distribution**

416 **Univariate skew- $t$  distribution**

417 We say that  $Y$  follows a univariate extended skew- $t$  distribution with location  $\xi \in \mathcal{R}$ , scale  $\omega > 0$ , skew

418 parameter  $\alpha \in \mathcal{R}$ , and degrees of freedom  $\nu$  if has distribution function

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

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

420 **Multivariate skew- $t$  distribution**

421 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  $\boldsymbol{\omega} = \text{diag}(\omega_1, \dots, \omega_d)$ ,

422 then the density of  $Y$  at  $y$  is

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

423 where

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

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

424 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

425  $\eta$ ,  $Q(\mathbf{z}) = \mathbf{z}^T \bar{\Omega}^{-1} \mathbf{z}$  and  $T(\cdot; \eta)$  denotes the univariate Student's  $t$  distribution function with  $\eta$  degrees of

426 freedom (Azzalini and Capitanio, 2014).

427 **Extremal dependence**

428 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

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

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

430  $\alpha_j = \alpha_i \sqrt{1 - \varrho^2}$ ,  $\tau_j = \sqrt{\eta+1}(\alpha_j + \alpha_i \varrho)$ , 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

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

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

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

434 Then

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

435 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

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

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

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

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

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

441 4 have different Brier scores at  $q(0.90)$ , whereas there is not significant evidence to suggest that method 1

<sup>442</sup> and method 2 have different Brier scores at  $q(0.90)$ . In each table group A represents the group with the  
<sup>443</sup> 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

## <sup>444</sup> References

- <sup>445</sup> Azzalini, A. and Capitanio, A. (2014) *The Skew-Normal and Related Families*. Institute of Mathematical  
<sup>446</sup> Statistics Monographs. Cambridge University Press.
- <sup>447</sup> Coles, S., Heffernan, J. and Tawn, J. (1999) Dependence Measures for Extreme Value Analyses. *Extremes*,  
<sup>448</sup> **2**, 339–365.
- <sup>449</sup> Davison, A. C. and Smith, R. L. (1990) Models for exceedances over high thresholds (with Discussion).  
<sup>450</sup> *Journal of the Royal Statistical Society. Series B (Methodological)*, **52**, 393–442.

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

- 451 Engelke, S., Malinowski, A., Kabluchko, Z. and Schlather, M. (2014) Estimation of Hüsler-Reiss distributions  
 452 and Brown-Resnick processes. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 239–265.
- 454 Gneiting, T. and Raftery, A. E. (2007) Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, **102**, 359–378.
- 456 Huser, R. and Davison, A. C. (2014) Space-time modelling of extreme events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **76**, 439–461.
- 458 Kim, H.-M., Mallick, B. K. and Holmes, C. C. (2005) Analyzing Nonstationary Spatial Data Using Piecewise Gaussian Processes. *Journal of the American Statistical Association*, **100**, 653–668.
- 460 Ledford, A. W. and Tawn, J. a. (1996) Statistics for near independence in multivariate extreme values. *Biometrika*, **83**, 169–187.
- 462 Padoan, S. A. (2011) Multivariate extreme models based on underlying skew- and skew-normal distributions. *Journal of Multivariate Analysis*, **102**, 977–991.
- 464 Padoan, S. A., Ribatet, M. and Sisson, S. A. (2010) Likelihood-Based Inference for Max-Stable Processes. *Journal of the American Statistical Association*, **105**, 263–277.
- 466 Reich, B. J. and Shaby, B. A. (2012) A hierarchical max-stable spatial model for extreme precipitation. *The Annals of Applied Statistics*, **6**, 1430–1451.
- 468 Samet, J. M., Dominici, F., Zeger, S. L., Schwartz, J. and Dockery, D. W. (2000) The National Morbidity, Mortality and Air Pollution Study Part I : Methods and Methodologic Issues. *Tech. Rep. 94*.
- 470 Thibaud, E., Mutzner, R. and Davison, a. C. (2013) Threshold modeling of extreme spatial rainfall. *Water Resources Research*, **49**, 4633–4644.

- 472 Thibaud, E. and Opitz, T. (2013) Efficient inference and simulation for elliptical Pareto processes. 1–19.
- 473 Wadsworth, J. L. and Tawn, J. a. (2012) Dependence modelling for spatial extremes. *Biometrika*, **99**, 253–  
474 272.
- 475 — (2014) Efficient inference for spatial extreme value processes associated to log-Gaussian random func-  
476 tions. *Biometrika*, **101**, 1–15.