

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

2 June 10, 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. ? considered modeling threshold exceedances of
17 univariate time series by the generalized Pareto distribution. Bivariate threshold models for extreme value
18 distributions were considered by ? who introduced a censored approach that provides a way to deal with
19 different types of exceedances of a bivariate threshold in terms of only one or both components. These
20 threshold models were extended to spatial models for extremes by ? and ? who fit various models to
21 spatial extremes using a censored pairwise likelihood (?) based on the approach of ?. ? further extended
22 this to spate-time modeling. ?, ?, and ? introduced more efficient inference for threshold exceedances of

23 extremal spatial processes with full likelihood methods. The previous approaches to threshold modeling are
24 motivated by extreme value theory and assume the threshold is high enough to assume extremal models are
25 valid for the data, and for extrapolation beyond the range of observed values. Moreover, these approaches
26 are computationally intensive and limited to rather small datasets. Our application with ozone data does not
27 fit into this framework because we do not focus on exceedances of a very high threshold, but on exceedances
28 of a fixed threshold.

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

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

40 The paper is organized as follows. Section 2 is a brief review of the spatial skew-*t* process. In Section
41 3.3, we build upon the traditional skew-*t* by incorporating censoring to focus on tails, partitioning to remove
42 long-range asymptotic dependence, and extending the model to space-time data. The computing is described
43 in Section 4. In Section 5, we present a simulation study that examines the predictive capabilities of this
44 model compared with a naïve Gaussian method. We then compare our method to Gaussian and max-stable
45 methods with a data analysis of ozone measurements from throughout the US in section 6. The final section

⁴⁶ provides brief discussion and direction for future research.

⁴⁷ 2 Spatial skew processes

⁴⁸ Many types of data demonstrate some level of skewness and therefore should be modeled with distributions
⁴⁹ that allow for asymmetry. The skew-elliptical family of distributions provides models that are mathemati-
⁵⁰ cally tractable while introducing a slant parameter to account for asymmetric data (?). A brief review of the
⁵¹ additive process by which a skew-*t* process is created is given here.

⁵² 2.1 Skew-*t* process

⁵³ 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
⁵⁴ spatial skew-*t* process can be written

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

⁵⁵ 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
⁵⁶ 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
⁵⁷ a spatial Gaussian process with mean zero, variance one, and a positive definite correlation function.

⁵⁸ For a finite collection of locations $\mathbf{s}_1, \dots, \mathbf{s}_n$, consider observations $\mathbf{Y} = [Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n)]^T$. After
⁵⁹ marginalizing over both z and σ ,

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

⁶⁰ that is, \mathbf{Y} follows an n -dimensional skew-*t* distribution with location $\mathbf{X}\boldsymbol{\beta}$, correlation matrix $\boldsymbol{\Omega}$, slant param-
⁶¹ eters $\boldsymbol{\alpha}$ and degrees of freedom $2a$, where $\mathbf{X} = [\mathbf{X}(\mathbf{s}_1)^T, \dots, \mathbf{X}(\mathbf{s}_n)^T]$, $\boldsymbol{\Omega} = \boldsymbol{\omega} \bar{\boldsymbol{\Omega}} \boldsymbol{\omega}$, $\boldsymbol{\omega} = \text{diag} \left(\frac{1}{\sqrt{ab}}, \dots, \frac{1}{\sqrt{ab}} \right)$,

62 $\bar{\Omega} = (\Sigma + \lambda^2 \mathbf{1} \mathbf{1}^T)$, Σ is the positive definite correlation matrix of $[v(\mathbf{s}_1), \dots, v(\mathbf{s}_n)]$, $\alpha = \lambda(1 + \lambda^2 \mathbf{1}^T \Sigma^{-1} \mathbf{1})^{-1/2} \mathbf{1}^T \Sigma^{-1}$
 63 is a vector of slant parameters. Although Σ can be any positive definite correlation function, we choose to
 64 use the stationary isotropic Matérn correlation with

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

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

69 2.2 Extremal dependence

70 Our interest lies in spatial dependence in the tail of the skew- t process. One measure of extremal dependence
 71 is the χ statistic (?). For a stationary and isotropic spatial process, the χ statistic for two locations \mathbf{s} and \mathbf{t}
 72 separated by distance $h = \|\mathbf{s} - \mathbf{t}\|$ with identical marginal distributions is

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

73 where c^* is the upper limit of the support of Y . If $\chi(h) = 0$, then observations are asymptotically in-
 74 dependent at distance h . For Gaussian processes, $\chi(h) = 0$ regardless of the distance, so they are not
 75 suitable for modeling asymptotically dependent extremes. Unlike the Gaussian process, the skew- t process
 76 is asymptotically dependent (see Appendix A.4). However, one problem with the spatial skew- t process is
 77 that $\lim_{h \rightarrow \infty} \chi(h) > 0$. This occurs because all observations, both near and far, share the same z and σ
 78 terms. Therefore, this long-range dependence feature of the skew- t process is not ideal for spatial analysis

79 of large geographic regions where we expect only local spatial dependence. The explicit expression for $\chi(h)$
80 is given in Appendix A.4.

81 **3 Spatiotemporal skew- t model for extremes**

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

86 **3.1 Censoring to focus on the tails**

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

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

89 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
90 is a pre-specified threshold value. Then, assuming the uncensored data $Y(\mathbf{s})$ are observations from a skew- t
91 process, we update values censored below the threshold using standard Bayesian missing data methods as
92 described in Section 4.

93 **3.2 Partitioning to remove long-range asymptotic dependence**

94 The motivation for the partition is that for a large spatial domain, it may not be reasonable to assume sites that
95 are far apart demonstrate asymptotic dependence. Modeling different levels of asymptotic dependence was

96 discussed by ? with a hybrid spatial dependence model. ? also allow for asymptotic dependence across both
 97 space and time with a partition structure represented by random discs moving across the space for a random
 98 duration with a random velocity and random radius. We handle the problem of long-range asymptotic
 99 dependence with a random partition. As discussed in Section 2, the source of long-range dependence is
 100 the shared z and σ . Therefore, to alleviate this dependence, we allow z and σ to vary by site. The model
 101 becomes

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

102 Let $\mathbf{w} = (w_1, w_2)$ be the location of a spatial knot. To model spatial variation, consider a set of spatial knots
 103 $\mathbf{w}_1, \dots, \mathbf{w}_K$ from a homogeneous Poisson process with intensity μ over spatial domain $\mathcal{D} \in \mathbb{R}^2$. The knots
 104 define a random partition of \mathcal{D} by subregions P_1, \dots, P_K defined as

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

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

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

106 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
 107 of an inverse gamma random variable. So, within each partition, $Y(\mathbf{s})$ follows the spatial skew- t process
 108 defined in Section 2. Across partitions, the $Y(\mathbf{s})$ remain correlated via the correlation function for $v(\mathbf{s})$
 109 because $v(\mathbf{s})$ spans all partitions.

110 When incorporating the random partition, conditional on knots $\mathbf{w}_1, \dots, \mathbf{w}_K$, the χ statistic for two sites

111 \mathbf{s} and \mathbf{t} in partitions k_s and k_t respectively is

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

112 where $I(\cdot)$ is an indicator function, $\chi_{\text{skew-}t}(h)$ is the χ statistic for a skew- t process, $\chi_{\text{Gaus}}(h)$ is the χ statistic
113 for a Gaussian process, and $h = \|\mathbf{s} - \mathbf{t}\|$. Therefore, sites in different subregions are asymptotically inde-
114 pendent because $\chi_{\text{Gaus}}(h) = 0$ for all h . Marginally, over the knots $\mathbf{w}_1, \dots, \mathbf{w}_K$, $\chi(h) = \pi(h)\chi_{\text{skew-}t}(h)$,
115 where $\pi(h) = \Pr(k_s = k_t)$ is the probability that two sites separated by distance h are in the same partition.
116 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 of this is given in
117 Appendix A.3.

118 In Figure 1, we give $\chi(h)$ for $K = 1, 3, 5, 10$ partitions for a skew- t distribution with $\alpha = 10$, and
119 3 degrees of freedom. To estimate $\pi(h)$, we generate 500 sites uniformly over the unit-square. We then
120 randomly generate 400 different sets of partitions using $K = 3, 5$, and 10. For each set of knots, we
121 take $\pi(h)$ to be the proportion of sites in the same partition that are separated by distance h . This plot
122 demonstrates how partitioning helps to reduce extremal dependence as h increases.

123 3.3 Extension to space-time data

124 When using daily measurements, the assumption of temporal independence is often inappropriate. There are
125 several places where temporal dependence could be incorporated in the model, including the residual $v_t(\mathbf{s})$.
126 However, we choose to allow for temporal dependence in the \mathbf{w} , z , and σ terms because these terms dictate

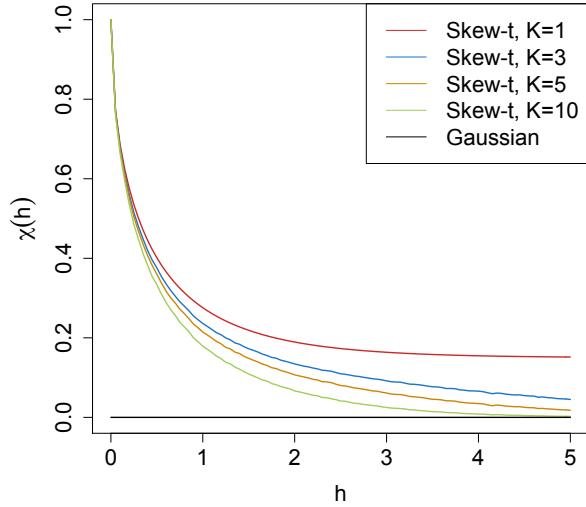


Figure 1: $\chi(h)$ for $K = 1, 3, 5$, and 10 knots as a function of distance.

127 the tail behavior which is our primary focus. In this section, we extend (6) to the spatiotemporal setting. Let

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

128 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
 129 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
 130 partition P_{t1}, \dots, P_{tK} , and for $\mathbf{s} \in P_{tk}$,

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

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

133 We use an AR(1) time series model for w_{tk} , z_{tk} , and σ_{tk} . The time series model must be specified after
 134 a transformation to preserve the skew- t process at each time point. For each time-varying parameter, we

135 transform to obtain a standard normal marginal distribution, place a Gaussian prior with autocorrelation on
 136 the transformed parameter, and then transform back to obtain the marginal distribution required to preserve
 137 the skew- t process. We first transform the spatial knots from \mathcal{D} to \mathcal{R}^2 as follows. Let

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

138 where Φ is a univariate standard normal density function, and $\mathbf{s}_i = [s_{1i}, \dots, s_{ni}]$. Then the transformed
 139 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
 140 gamma. Let

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

141 where IG is defined as before. We also use a copula on $z_t(\mathbf{s})$ to ensure that the marginal distributions of
 142 $z_t(\mathbf{s})$ are half-normal. Let

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

143 where HN is the distribution function of a half-normal random variable. The AR(1) process for each tail
 144 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

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

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

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

145 where $|\phi_w|, |\phi_z|, |\phi_\sigma| < 1$. These are stationary time series models with marginal distributions $\mathbf{w}_k^* \sim N_2(0, 1)$,
 146 $z_k^* \sim N(0, \sigma_k^2)$, and $\sigma_k^{2*} \sim N(0, 1)$. After transformation back to the original space, $\mathbf{w}_{tk} \sim \text{Unif}(\mathcal{D})$,
 147 $z_{tk} \sim HN(0, \sigma_{tk}^2)$ $\sigma_{tk}^2 \sim \text{IG}(a, b)$. For each day, the model is identical to the spatial-only model in (6)
 148 by construction.

149 4 Computation

150 First, we impute values below the threshold conditional on observations above the threshold. This is feasible
 151 for large datasets with our model because for a single day, conditional on the model parameters, we only
 152 need to draw from a truncated multivariate normal distribution. Then, we update model parameters, Θ ,
 153 using Metropolis Hastings or Gibbs sampling when appropriate. Finally, we make spatial predictions using
 154 conditional multivariate normal results and the fact that the distribution of $Y_t(\mathbf{s}) \mid \Theta, z(\mathbf{s})$ is the usual
 155 multivariate normal distribution with a Matérn spatial covariance structure.

156 We can use Gibbs sampling to update $Y_t(\mathbf{s})$ for censored observations that are below the threshold T .
 157 After conditioning on $\lambda, z_t(\mathbf{s})$ and non-censored observations, $Y_t(\mathbf{s})$ has truncated normal full conditionals.
 158 So we sample $Y_t(\mathbf{s}) \sim N_{(-\infty, T)}(\mathbf{X}_t^T(\mathbf{s})\beta + \lambda|z_t(\mathbf{s})|, \Sigma)$. After imputing the censored observations, we
 159 update the model parameters. To update the model parameters, we use standard Gibbs updates for parame-
 160 ters when possible. In the case Gibbs sampling is not possible, parameters are updated using a random-walk
 161 Metropolis Hastings algorithm. See Appendices A.1 and A.2 for details regarding the MCMC. The final step
 162 of the computation is to use Bayesian Kriging to generate a predictive distribution for $Y_t(\mathbf{s}^*)$ at prediction
 163 location \mathbf{s}^* . This step is similar to the imputation for censored observations except that the full conditionals
 164 are no longer truncated at T .

¹⁶⁵ **4.1 Hierarchical model**

¹⁶⁶ Conditioned on $z_{tk}(\mathbf{s})$, $\sigma_{tk}^2(\mathbf{s})$, and P_{tk} , the marginal distributions are Gaussian and the joint distribution
¹⁶⁷ multivariate Gaussian. However, we do not fix the partitions, they are treated as unknown and updated in the
¹⁶⁸ MCMC. We model this with a Bayesian hierarchical model as follows. Let $\mathbf{w}_{t1}, \dots, \mathbf{w}_{tK}$ be a set of daily
¹⁶⁹ spatial knots in a spatial domain of interest, \mathcal{D} , and P_{tk} as defined in (7). In practice, we fix K at many
¹⁷⁰ different levels, and assess the impact of fit as described in 5.2. Then

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

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

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

$$\lambda = \lambda_1 \lambda_2$$

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

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

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

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

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

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

¹⁷¹ where $\Theta = \{\rho, \nu, \gamma, \lambda, \beta\}$, and Σ is a Matérn covariance matrix as described in Section 2.1. We parameterize
¹⁷² $\lambda = \lambda_1 \lambda_2$ to help with convergence in the MCMC.

173 **5 Simulation study**

174 In this section, we conduct a simulation study to investigate how the number of partitions and the level of
175 thresholding impact the accuracy of predictions made by the model.

176 **5.1 Design**

177 For all simulation designs, we generate data from the model in Section 3.2 using $n_s = 144$ sites and
178 $n_t = 50$ independent days. The sites are generated Uniform($[0, 10] \times [0, 10]$). We generate data from 5
179 different simulation designs:

- 180 1. Gaussian marginal, $K = 1$ knot
- 181 2. Skew- t marginal, $K = 1$ knots
- 182 3. Skew- t marginal, $K = 5$ knots
- 183 4. Max-stable
- 184 5. Transformation below $T = q(0.80)$

185 In the first three designs, the $v_t(\mathbf{s})$ terms are generated using a Matérn covariance with smoothness parameter
186 $\nu = 0.5$ and spatial range $\rho = 1$. For the covariance matrices in designs 1 – 3, the proportion of the variance
187 accounted for by the spatial variation is $\gamma = 0.9$ while the proportion of the variance accounted for by the
188 nugget effect is 0.1. In the first design, $\sigma^2 = 2$ is used for all days which results in a Gaussian distribution.

189 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 designs 1,
190 we set $\lambda = 0$. For designs 2 and 3, $\lambda = 3$ was used as to simulate moderate skewness, and the z_t are
191 generated as described in (8). In the fourth design, we generate from a spatial max-stable distribution (?).

192 In this design, data have marginal distributions that follow a generalized extreme value distribution with
193 parameters $\mu = 1, \sigma = 1, \xi = 0.2$. In this model, a random effect is used to induce spatial dependence using
194 144 spatial knots on a regular lattice in the square $[1, 9] \times [1, 9]$. For this setting, we set $\gamma = 0.5$. In the final

195 design, we generate \tilde{y} using the setting from design two, and then consider the data

$$y = \begin{cases} \tilde{y}, & \tilde{y} > T \\ T \exp\{\tilde{y} - T\}, & \tilde{y} \leq T \end{cases} \quad (19)$$

196 where $T = q(0.80)$ is the 80th sample quantile of the data. In all five designs, the mean $\mathbf{X}^T \boldsymbol{\beta} = 10$ is
 197 assumed to be constant across space.

198 $M = 50$ data sets are generated for each design. For each data set we fit the data using five models

- 199 1. Gaussian marginal, $K = 1$ knots
- 200 2. Skew- t marginal, $K = 1$ knots, $T = -\infty$
- 201 3. Symmetric- t marginal, $K = 1$ knots, $T = q(0.80)$
- 202 4. Skew- t marginal, $K = 5$ knots, $T = -\infty$
- 203 5. Symmetric- t marginal, $K = 5$ knots, $T = q(0.80)$
- 204 6. A max-stable model based on ? thresholded at $T = q(0.80)$

205 where $q(0.80)$ is the 80th sample quantile of the data. The design matrix \mathbf{X} includes an intercept with a first-
 206 order spatial trend with priors of β_{int} , β_{lat} , β_{long} , $\stackrel{iid}{\sim} N(0, 10)$. The spatial covariance parameters have priors
 207 $\log(\nu) \sim N(-1.2, 1)$, $\gamma \sim \text{Unif}(0, 1)$, $\rho \sim \text{Unif}(15)$. The skewness parameter has prior $\lambda_2 \sim \text{IG}(0.1, 0.1)$.
 208 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
 209 tried also fitting the skew- t marginals for the thresholded models, but it is very challenging for the MCMC
 210 to properly identify the skewness parameter with only one tail worth of data. Each chain of the MCMC
 211 ran for 20000 iterations with a burn-in period of 10000 iterations. Parameters appear to converge properly;
 212 however, in the models with multiple partitions (i.e. models 4 and 5) it is hard to assess the convergence of
 213 \mathbf{w} , $z(\mathbf{s})$, and $\sigma^2(\mathbf{s})$ because of partition label switching throughout the MCMC.

214 **5.2 Cross validation**

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

222 **5.3 Results**

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

$$BS_{\text{rel}} = \frac{BS_{\text{method}}}{BS_{\text{Gaussian}}}.$$
 (20)

226 We analyzed the results for the simulation study using a Friedman test at $\alpha = 0.05$. If the Friedman test
227 came back with a significant results, we conducted a Wilcoxon-Nemenyi-McDonald-Thompson test to see
228 which methods had different results. The full results for the Wilcoxon-Nemenyi-McDonald-Thompson tests
229 are given in Appendix A.5.

230 Figure 2 shows that when the data come from a Gaussian process, our methods perform comparably to
231 the Gaussian method. For data settings with skew- t marginals (settings 2 – 3), we find significant improve-
232 ment over the Gaussian method. Furthermore in these data settings, we find the best performance occurs

233 when the number of knots used in the method matches the number of knots used for data generation. The
234 non-thresholded methods tend to outperform the thresholded methods, but this is not surprising given that
235 the data are generated directly from the model used in the method. For the max-stable data, we see that for
236 low-extreme quantiles, the Gaussian method performs better, for more extreme quantiles, the single-partition
237 method, both thresholded and non-thresholded, perform significantly better than the Gaussian. Finally, for
238 setting 5, although the thresholded version of the single-partition model tends to perform the best across all
239 of the extreme quantiles, the difference between the thresholded and non-thresholded methods is no longer
240 significant in the more extreme quantiles.

241 **6 Data analysis**

242 To illustrate this method, we consider the daily maximum 8-hour ozone measurements for July 1 - 31, 2005
243 at 1089 Air Quality System (AQS) monitoring sites in the United States as the response (see Figure 3). For
244 each site, we also have covariate information containing the estimated ozone from the Community Multi-
245 scale Air Quality (CMAQ) modeling system. Initially, we fit a linear regression assuming a mean function
246 of

$$\mathbf{X}_t^T(\mathbf{s})\boldsymbol{\beta} = \beta_0 + \beta_1 \cdot \text{CMAQ}_t(\mathbf{s}). \quad (21)$$

247 The data from July 10 are shown in Figure 3 along with a Q-Q plot of the residuals compared to a skew-*t*
248 distribution with 10 d.f. and $\alpha = 1$. Exploratory data analysis indicates that there is dependence in the high
249 quantile levels of the residuals beyond what we expect in the case of independence.

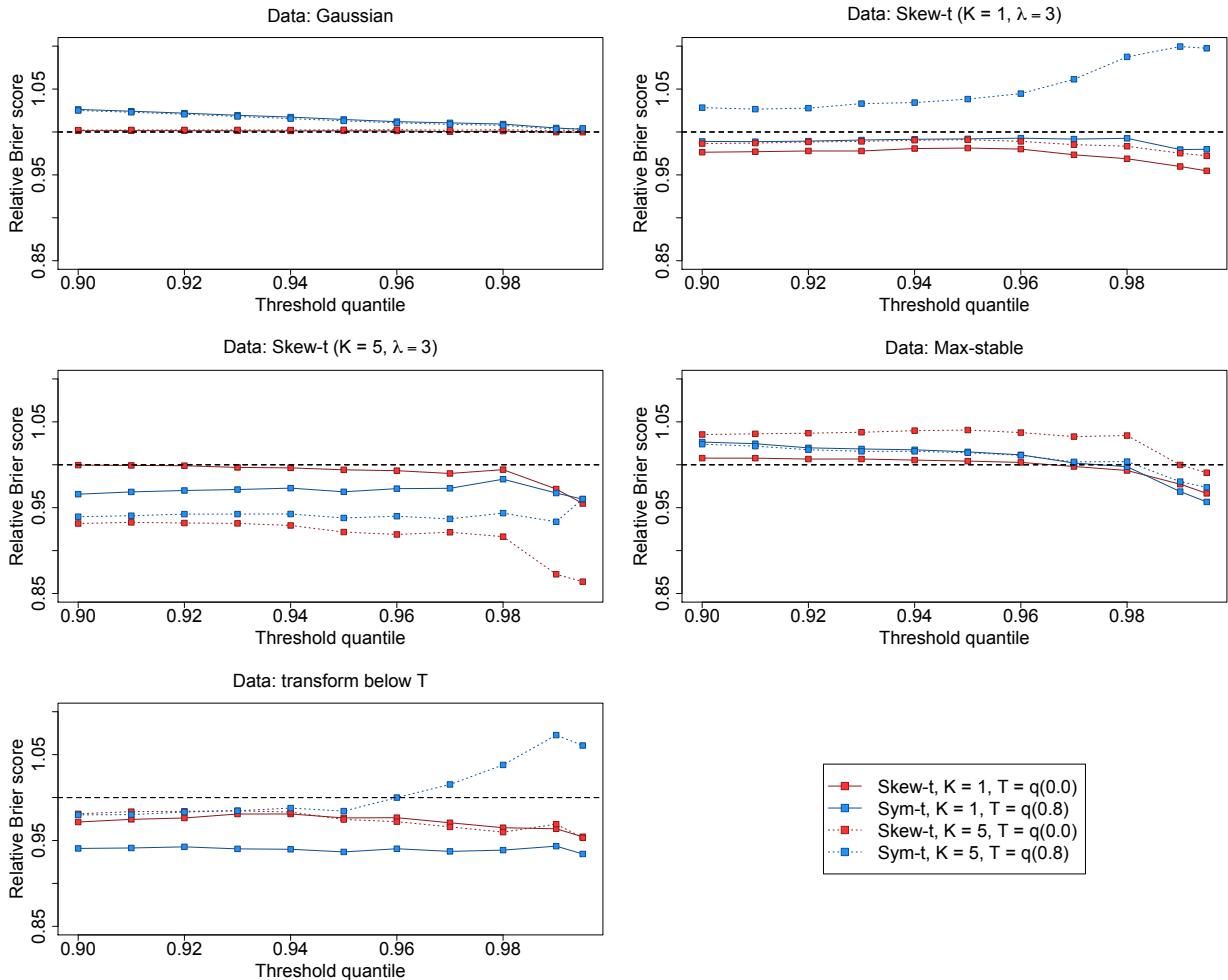


Figure 2: Brier scores relative to the Gaussian method for simulation study results. A ratio lower than 1 indicates that the method outperforms the Gaussian method.

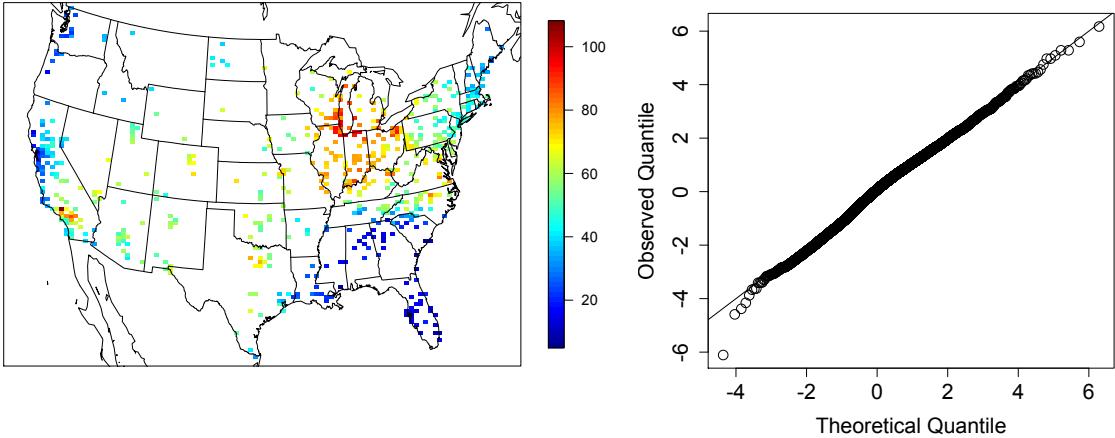


Figure 3: Ozone values on 10 July 2005 (left) Q-Q plot of the residuals (right)

250 6.1 Model comparisons

251 We fit the model using Gaussian and skew- t marginal distributions with $K = 1, 5, 6, 7, 8, 9, 10, 15$ partitions.
 252 We choose to censor $Y(\mathbf{s})$ at $T = 0, 50$ (0.42 sample quantile), and 75 (0.92 sample quantile) ppb in order
 253 to compare results from no, moderate, and high censoring. The upper threshold of 75 ppb was used because
 254 the current air quality standard is based on exceedance of 75 ppb. As with the simulation study, for models
 255 with a threshold of $T = 75$, we use a symmetric- t marginal distribution. We also compare models with no
 256 time series to models that include the time series. Finally, as a comparison to max-stable methods, we fit
 257 the model using the hierarchical max-stable model of ? with the data thresholded at $T = 75$. All methods
 258 assume the mean function given in (21). To ensure that the max-stable method runs in a reasonable amount
 259 of time, we take a stratified sample of the sites to get 800 sites and consider this our new dataset. We conduct
 260 two-fold cross validation using 400 training sites and 400 validation sites as described in Section 5.2
 261 Each chain of the MCMC ran for 30000 iterations with a burn-in period of 25000 iterations. Parameters
 262 appear to converge properly; however, as before, for models with multiple partitions it is hard to assess the
 263 convergence of \mathbf{w} , $z(\mathbf{s})$, and $\sigma^2(\mathbf{s})$ because of partition label switching throughout the MCMC. For each

264 model, Brier scores were averaged over all sites and days to obtain a single Brier score for each dataset. At
265 a particular threshold or quantile level, the model that fits the best is the one with the lowest score. We then
266 compute the relative (to Gaussian) Brier scores (see Section 5.3) to compare each model.

267 **6.2 Results**

268 The results suggest that the skew- t , thresholded, partitioned, and time series models all give an improvement
269 in predictions over the Gaussian model, whereas the max-stable method results in relative Brier scores
270 between 1.07 and 1.15 indicating poorer performance than the Gaussian model. The plots in Figure 4
271 show the relative Brier scores for time-series and non-time-series models, using $K = 1, 7$, and 15 knots at
272 thresholds $T = 0, 50$, and 75 ppb. Most of the models perform similarly across all the Brier scores; however,
273 for single-partition models without thresholding, performance tends to diminish in the extreme quantiles.
274 The results also suggest that thresholding improves performance for estimates in the extreme quantiles. Both
275 plots have similar features suggesting that most settings do reasonably well. In particular, for all extreme
276 quantiles, selecting a moderate number of knots (e.g. $K = 5, \dots, 10$) tends to give the best results. Table 1
277 shows the best two models for selected extreme quantiles.

278 We illustrate the predictive capability of our model in Figure 6 by plotting the 99th quantile of the
279 posterior predictive density for July in South Carolina and Georgia. We fit the model using four methods,
280 two reference and two that performed better. These four methods are

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

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

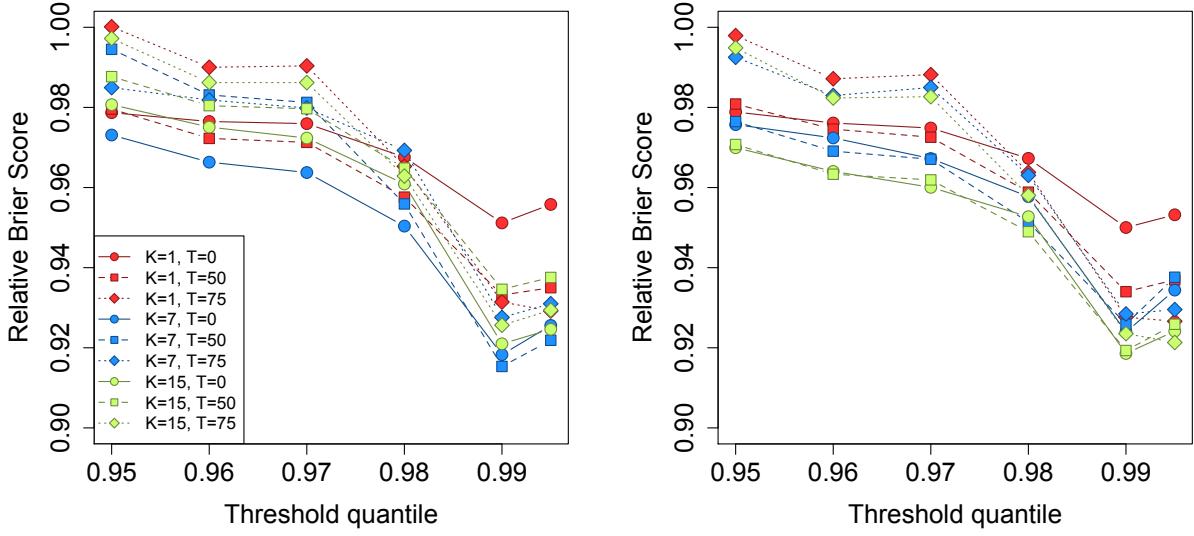


Figure 4: Relative Brier scores for time-series models (left) and non-time-series models (right). Relative brier score for the max-stable model is between 1.07 and 1.15

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

286 differences between the reference methods and the comparison methods is that the comparison methods tend
 287 to give higher estimates of the 99th quantile along the I-85 corridor between Charlotte and Atlanta.

288 NEED TO ADD STUFF HERE ACKNOWLEDGING THAT MARGINALS ARE DIFFERENT Also,
 289 add an explanation about the purpose of the plot in 5

290 7 Discussion

291 In this paper we propose a new threshold exceedance approach for spatiotemporal modeling based on the
 292 skew- t process. The proposed model gives flexible tail behavior, demonstrates asymptotic dependence for

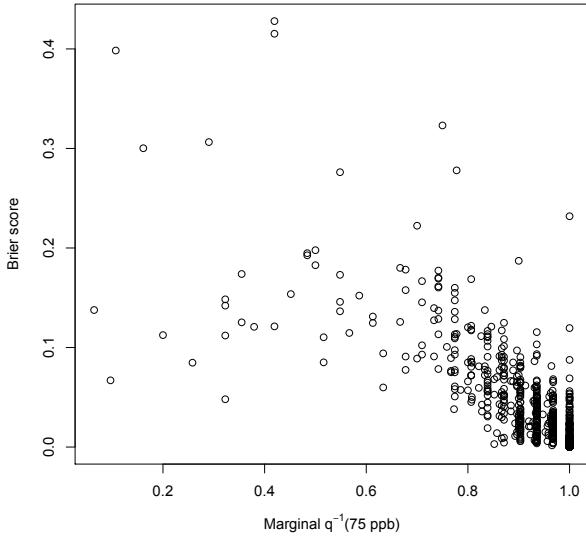


Figure 5: Comparison of Brier score performance by marginal quantile representing 75ppb

293 observations at sites that are near to one another, and has computation on the order of Gaussian models
 294 for large space-time datasets. In the simulation study, we demonstrate that this model shows statistically
 295 significant improvements over a naïve Gaussian approach. In both the simulation study, and the application
 296 to ozone data, we find that incorporating a partition in the model improves extreme prediction. Furthermore
 297 the results from the data analysis suggest that thresholding can improve performance when predicting in the
 298 extreme tails of the data.

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

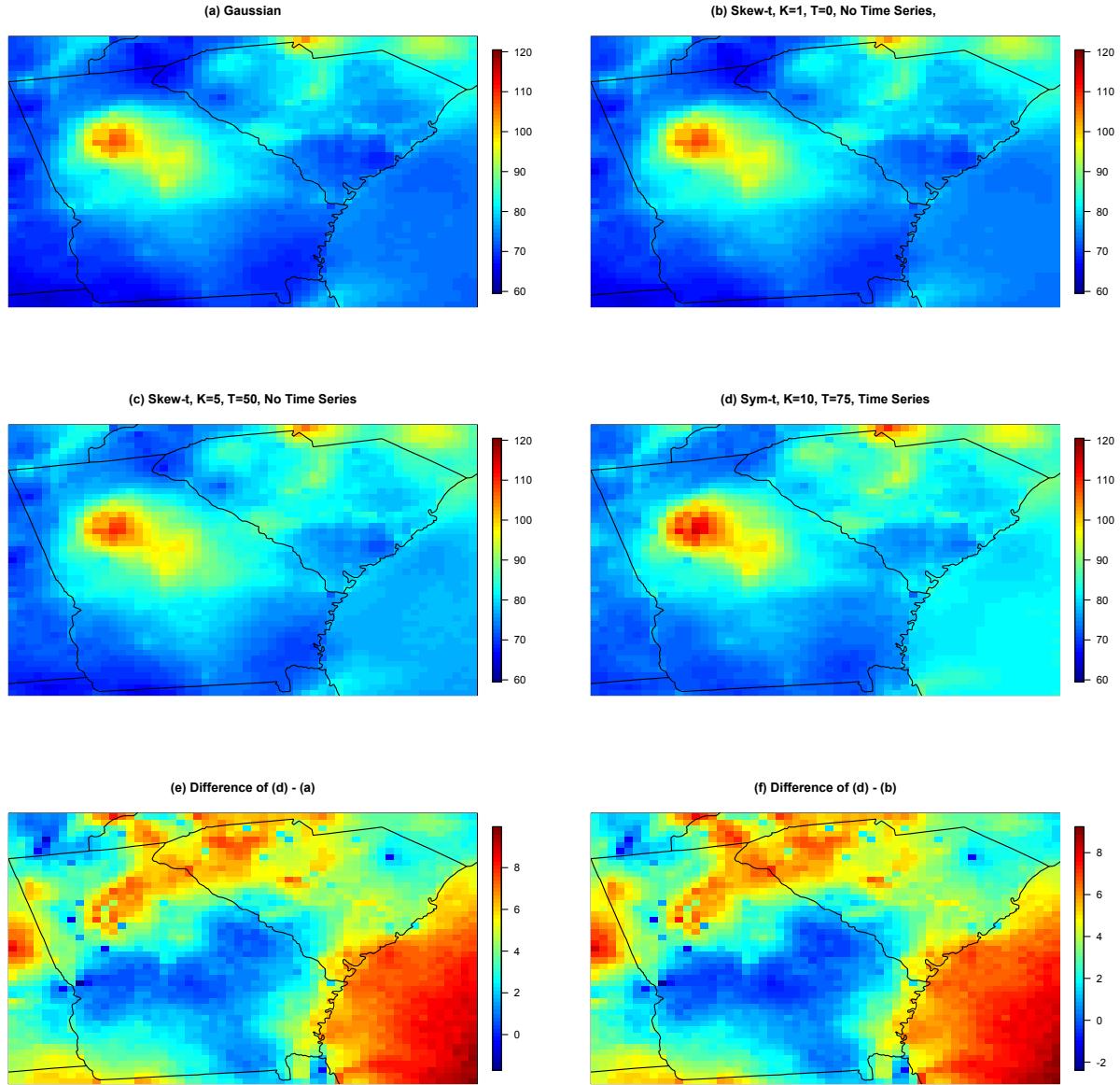


Figure 6: (a) – (d) give the posterior predictive $\hat{q}(0.99)$ for the month of July under four different models, (e) gives the difference between $\hat{q}(0.99)$ in plots (d) and (a), (f) gives the difference between $\hat{q}(0.99)$ in plots (d) and (b).

306 number of knots, we may be underestimating the uncertainty in the model. This could be incorporated by
 307 treating the number of knots as a model parameter instead of fixing it to be a specific value.

308 Acknowledgments

309 A Appendices

310 A.1 MCMC details

311 The MCMC sampling for the model 4.1 is done using R (<http://www.r-project.org>). Whenever possible,
 312 we select conjugate priors (see Appendix A.2); however, for some of the parameters, no conjugate prior
 313 distributions exist. When no conjugate prior distribution exists, we use a random walk Metropolis Hastings
 314 update step. In each Metropolis Hastings update, we tune the algorithm to give acceptance rates near 0.40.

315 Spatial knot locations

316 For each day, we update the spatial knot locations, $\mathbf{w}_1, \dots, \mathbf{w}_K$, using a Metropolis Hastings block up-
 317 date. Because the spatial domain is bounded, we generate candidate knots using the transformed knots
 318 $\mathbf{w}_1^*, \dots, \mathbf{w}_K^*$ (see section 3.3) and a random walk bivariate Gaussian candidate distribution

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

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

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

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

323 **Spatial random effects**

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

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

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

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

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

331 **Variance terms**

332 When there is more than one site in a partition, then we update σ_{tk}^2 using a Metropolis Hastings update.
 333 First, we generate a candidate for σ_{tk}^2 using an $IG(a^*/s, b^*/s)$ candidate distribution in an independence
 334 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
 335 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

³³⁶ day t . The acceptance ratio is

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

³³⁷ where $p[\cdot]$ is the prior either taken from the time series given in Section 3.3 or assumed to be $\text{IG}(a, b)$, and

³³⁸ $c[\cdot]$ is the candidate distribution. The candidate is accepted with probability $\min\{R, 1\}$.

³³⁹ Spatial covariance parameters

³⁴⁰ We update the three spatial covariance parameters, $\log(\rho)$, $\log(\nu)$, γ , using a Metropolis Hastings block

³⁴¹ update step. First, we generate a candidate using a random walk Gaussian candidate distribution

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

³⁴² where $\log(\rho)^{(r-1)}$ is the value at MCMC iteration $r - 1$, and s is a tuning parameter. Candidates are

³⁴³ generated for $\log(\nu)$ and γ in a similar fashion. The acceptance ratio is

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

³⁴⁴ All three candidates are accepted with probability $\min\{R, 1\}$.

³⁴⁵ **A.2 Posterior distributions**

³⁴⁶ **Conditional posterior of $z_{tk} | \dots$**

³⁴⁷ If knots are independent over days, then the conditional posterior distribution of $|z_{tk}|$ is conjugate. For
³⁴⁸ simplicity, drop the subscript t , let $\tilde{z}_{tk} = |z_{tk}|$, and define

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

³⁴⁹ Let

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

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

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

³⁵⁰ Then

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

³⁵¹ where

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

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

³⁵² Then $\tilde{Z}_l | \dots \sim N_{(0,\infty)}(\Lambda_l^{-1} \mu_l, \Lambda_l^{-1})$

³⁵³ **Conditional posterior of β | ...**

³⁵⁴ Let $\beta \sim N_p(0, \Lambda_0)$ where Λ_0 is a precision matrix. Then

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

³⁵⁵ where

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

³⁵⁶ **Conditional posterior of $\sigma^2 | \dots$**

³⁵⁷ 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

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

³⁵⁹ where

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

³⁶⁰ In the case that $L > 1$, a random walk Metropolis Hastings step will be used to update σ_{lt}^2 .

³⁶¹ **Conditional posterior of $\lambda | \dots$**

³⁶² For convergence purposes we model $\lambda = \lambda_1 \lambda_2$ where

$$\lambda_1 = \begin{cases} +1 & \text{w.p.0.5} \\ -1 & \text{w.p.0.5} \end{cases} \quad (22)$$

$$\lambda_2^2 \sim IG(\alpha_\lambda, \beta_\lambda). \quad (23)$$

$$(24)$$

³⁶³ Then

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

³⁶⁴ Then $\lambda_2 | \dots \sim IG(\alpha_\lambda + kt, \beta_\lambda + \frac{z^2}{2\sigma_{tk}^2})$

³⁶⁵ **A.3 Proof that** $\lim_{h \rightarrow \infty} \pi(h) = 0$

³⁶⁶ Let $N(A)$ be the number of knots in A , the area between sites s_1 and s_2 . Consider a spatial Poisson process

³⁶⁷ with intensity $\mu(A)$. So,

$$P[N(A) = k] = \frac{\mu(A)^k \exp\{-\mu(A)\}}{k!}.$$

³⁶⁸ Then for any finite k , $\lim_{h \rightarrow \infty} P[N(A) = k] = 0$ because $\lim_{h \rightarrow \infty} \mu(A) = \infty$. With each additional knot

³⁶⁹ in A , the chance that s_1 and s_2 will be in the same partition will decrease, because partition membership

³⁷⁰ is defined by the closest knot to a site. Therefore, $\lim_{h \rightarrow \infty} \pi(h) = 0$.

³⁷¹ **A.4 Skew-t distribution**

³⁷² **Univariate extended skew-t distribution**

³⁷³ We say that Y follow a univariate extended skew-t distribution with location $\xi \in \mathcal{R}$, scale $\omega > 0$, skew

³⁷⁴ parameter $\alpha \in \mathcal{R}$, extended parameter $\tau \in \mathcal{R}$, and degrees of freedom ν if has distribution function

$$f_{EST}(y) = \omega^{-1} \frac{f_T(z; \nu)}{F_T(\tau/\sqrt{1+\alpha^2}; \nu)} F_T \left[(\alpha z + \tau) \sqrt{\frac{\nu+1}{\nu+z^2}}; 0, 1, \nu+1 \right] \quad (25)$$

³⁷⁵ where $f_T(t; \nu)$ is a univariate Student's t with ν degrees of freedom, $F_T(t; \nu) = P(T < t)$, and $z = (y - \xi)/\omega$.

³⁷⁶ In the case that $\tau = 0$, then Y follows a univariate skew- t distribution.

³⁷⁷ Multivariate skew- t distribution

³⁷⁸ 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 $\boldsymbol{\omega} = \text{diag}(\omega_1, \dots, \omega_d)$,

³⁷⁹ then the density of Y at y is

$$f_y(\mathbf{y}) = \det(\boldsymbol{\omega})^{-1} f_z(\mathbf{z}) \quad (26)$$

³⁸⁰ where

$$f_z(\mathbf{z}) = 2t_d(\mathbf{z}; \bar{\Omega}, \eta) T \left[\boldsymbol{\alpha}^T \mathbf{z} \sqrt{\frac{\eta + d}{\nu + Q(\mathbf{z})}}; \eta + d \right] \quad (27)$$

$$\mathbf{z} = \boldsymbol{\omega}^{-1}(\mathbf{y} - \xi) \quad (28)$$

³⁸¹ 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

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

³⁸³ freedom (?).

³⁸⁴ Extremal dependence

³⁸⁵ For a bivariate skew- t random variable $\mathbf{Y} = [Y(\mathbf{s}), Y(\mathbf{t})]^T$, the $\chi(h)$ statistic (?) is given by

$$\chi(h) = \bar{F}_{\text{EST}} \left\{ \frac{[x_1^{1/\eta} - \varrho(h)]\sqrt{\eta+1}}{\sqrt{1-\varrho(h)^2}}; 0, 1, \alpha_1, \tau_1, \eta + 1 \right\} + \bar{F}_{\text{EST}} \left\{ \frac{[x_2^{1/\eta} - \varrho(h)]\sqrt{\eta+1}}{\sqrt{1-\varrho(h)^2}}; 0, 1, \alpha_2, \tau_2, \eta + 1 \right\}, \quad (29)$$

386 where \bar{F}_{EST} is the univariate survival extended skew- t function with zero location and unit scale, $\varrho(h) = \text{cor}(y_1, y_2)$,
 387 $\alpha_j = \alpha_i \sqrt{1 - \varrho^2}$, $\tau_j = \sqrt{\eta + 1}(\alpha_j + \alpha_i \varrho)$, 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
 388 $j = 1, 2$ and $i = 2, 1$ and where $\bar{\alpha}_j = (\alpha_j + \alpha_i \varrho) / \sqrt{1 + \alpha_i^2[1 - \varrho(h)^2]}$.

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

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

391 Then

$$\lim_{h \rightarrow \infty} \chi(h) = \bar{F}_{\text{EST}} \left\{ \sqrt{\eta + 1}; 0, 1, \alpha_1, \tau_1, \eta + 1 \right\} + \bar{F}_{\text{EST}} \left\{ \sqrt{\eta + 1}; 0, 1, \alpha_2, \tau_2, \eta + 1 \right\}. \quad (30)$$

392 Because the extended skew- t distribution is not bounded above, for all $\bar{F}_{\text{EST}}(x) = 1 - F_{\text{EST}} > 0$ for all
 393 $x < \infty$. Therefore, for a skew- t distribution, $\lim_{h \rightarrow \infty} \chi(h) > 0$.

394 A.5 Simulation study pairwise difference results

395 The following tables show the methods that have significantly different Brier scores when using a Wilcoxon-
 396 Nemenyi-McDonald-Thompson test. In each column, different letters signify that the methods have signifi-
 397 cantly different Brier scores. For example, there is significant evidence to suggest that method 1 and method
 398 4 have different Brier scores at $q(0.90)$, whereas there is not significant evidence to suggest that method 1
 399 and method 2 have different Brier scores at $q(0.90)$. In each table group A represents the group with the
 400 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
Method 2	A B	A B	A	A
Method 3	C	C	C	B
Method 4	B	B	B	B
Method 5	C	C	C	B

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

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

Table 5: Setting 4 – Max-stable

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

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	A B	A B
Method 3	A	A	A	A
Method 4	B C	B	B	B C
Method 5	B C	B	C	C