

1 **A new spatial model for points above a threshold**

2 March 30, 2015

3 **1 Introduction**

4 In most climatological applications, researchers are interested in learning about the average behavior of
5 different climate variables (e.g. ozone, temperature, rainfall). However, averages do not help regulators
6 prepare for the unusual events that only happen once every 100 years. For example, it is important to have
7 an idea of how much rain will come in a 100-year floor in order to construct strong enough river levees to
8 protect lands from flooding.

9 Unlike multivariate normal distributions, it is challenging to model multivariate extreme value distri-
10 butions (e.g. generalized extreme value and generalized Pareto distribution) because few closed-form ex-
11 pressions exist for the density in more than two-dimensions (Coles and Tawn, 1991). Given this limitation,
12 pairwise composite likelihoods have been used when modeling dependent extremes (Padoan et al., 2010;
13 Blanchet and Davison, 2011; Huser, 2013).

14 One way around the multi-dimensional limitation of multivariate extreme value distributions is to use
15 skew elliptical distributions to model dependent extreme values (Genton, 2004; Zhang and El-Shaarawi,
16 2010; Padoan, 2011). Due to their flexibility, the skew-normal and skew- t distribution offer a flexible way
17 to handle non-symmetric data within a framework of multivariate normal and multivariate t-distributions. As
18 with the spatial Gaussian process, the skew-normal distribution is also asymptotically independent; however,
19 the skew- t does demonstrate asymptotic dependence (Padoan, 2011). Although asymptotic dependence is
20 desirable between sites that are near one another, one drawback to the skew- t is that sites remain asymptot-
21 ically dependent even at far distances.

22 In this paper, we present a model that has marginal distributions with flexible tails, demonstrates asymp-

23 totic dependence for observations at sites that are near to one another, and has computation on the order of
24 Gaussian models for large space-time datasets. Specifically, our contribution is to incorporate thresholding
25 and random spatial partitions using a multivariate skew- t distribution. The advantage of using a thresholded
26 model as opposed to a non-thresholded model is that it allows for the tails of the distribution to inform the
27 predictions in the tails (DuMouchel, 1983). The random spatial partition alleviates the long-range spatial
28 dependence seen by the skew- t .

29 The paper is organized as follows. Section 2.1 is a brief review of the spatial skew- t process. In section
30 3.3, we build upon the traditional skew- t by incorporating censoring to focus on tails, partitioning to remove
31 long-range asymptotic dependence, and extending the model to space-time data. The computing is described
32 in section 4. In section 5, we present a simulation study that examines the predictive capabilities of this
33 model compared with a naïve Gaussian method. We then compare our method to Gaussian and max-stable
34 methods with a data analysis of ozone measurements from the eastern US in section 6. The final section
35 provides brief discussion and direction for future research.

36 **2 Spatial skew processes**

37 Many types of data demonstrate some level of skewness and therefore should be modeled with distributions
38 that allow for asymmetry. The skew-elliptical family of distributions provides models that are mathemati-
39 cally tractable while introducing a slant parameter to account for asymmetric data (Genton, 2004). A brief
40 review of the additive process by which a skew- t process is created is given here.

⁴¹ **2.1 Skew-*t* process**

⁴² Let $Y(\mathbf{s})$ be the observation at spatial location $\mathbf{s} = (s_1, s_2)$. The spatial skew-*t* process can be written

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

⁴³ where $\mathbf{X}(\mathbf{s})$ is a set of spatial covariates at site \mathbf{s} , $\boldsymbol{\beta}$ is a set of regression parameters, $\lambda \in \mathcal{R}$ is a 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 a spatial Gaussian process with mean zero and variance one. A common spatial correlation function is the ⁴⁵ Matérn with

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

⁴⁷ where ρ is the spatial range, ν is the smoothness, γ is the proportion of variance accounted for by the spatial ⁴⁸ variation, K_ν is a modified Bessel function of the second kind, and $h = \|\mathbf{s} - \mathbf{t}\|$.

⁴⁹ Let $\mathbf{Y} = [Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n)]^T$ be a set of observations at a finite collection of locations $\mathbf{s}_1, \dots, \mathbf{s}_n$. After ⁵⁰ marginalizing over both z and σ ,

$$\mathbf{Y} \sim \text{ST}_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, 2a), \quad (3)$$

⁵¹ that is, \mathbf{Y} follows an n -dimensional skew-*t* process with location $\mathbf{X}\boldsymbol{\beta}$, correlation matrix $\boldsymbol{\Omega}$, slant parameters ⁵² $\boldsymbol{\alpha}$ and degrees of freedom $2a$, where $\mathbf{X} = [\mathbf{X}(\mathbf{s}_1)^T, \dots, \mathbf{X}(\mathbf{s}_n)^T]$, $\boldsymbol{\Omega} = \boldsymbol{\omega} \bar{\boldsymbol{\Omega}} \boldsymbol{\omega}$, $\boldsymbol{\omega} = \text{diag} \left(\frac{1}{\sqrt{ab}}, \dots, \frac{1}{\sqrt{ab}} \right)$, ⁵³ $\bar{\boldsymbol{\Omega}} = (\boldsymbol{\Sigma} + \lambda^2 \mathbf{1} \mathbf{1}^T)$, $\boldsymbol{\Sigma}$ is a positive definite correlation matrix, $\boldsymbol{\alpha} = \lambda(1 + \lambda^2 \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1})^{-1/2} \mathbf{1}^T \boldsymbol{\Sigma}^{-1}$ is ⁵⁴ a vector of slant parameters. This process is desirable because of its flexible tail that is controlled by the ⁵⁵ skewness parameter λ and degrees of freedom $2a$. Furthermore, the marginal distributions at each location

56 also follow a univariate skew-*t* distribution (Azzalini and Capitanio, 2013).

57 **2.2 Extremal dependence**

58 One measure of extremal dependence is the χ statistic (Padoan, 2011). The χ statistic for the upper tail is
59 given by $\lim_{c \rightarrow \infty} \Pr(Y(\mathbf{s}_1) > c | Y(\mathbf{s}_2) > c)$. For a stationary spatial process, we can write the χ coefficient
60 as

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

61 If $\chi(h) = 0$, then observations are asymptotically independent at distance h . For Gaussian processes,
62 $\chi(h) = 0$ regardless of the distance, so they are not suitable for modeling spatially-dependent extremes.
63 Unlike the Gaussian process, the skew-*t* process is asymptotically dependent. However, one problem with
64 the spatial skew-*t* process is that $\lim_{h \rightarrow \infty} \chi(h) > 0$. This occurs because all observations, both near and
65 far, share the same z and σ terms. Therefore, the skew-*t* process is not ideal for spatial analysis of large geo-
66 graphic regions where we expect only local spatial dependence. The explicit expression for $\chi(h)$ (Padoan,
67 2011) and a proof of this are given in Appendix A.4.

68 **3 Spatiotemporal skew-*t* model for extremes**

69 In this section, we propose extensions to the skew-*t* process to model spatial extremes over a large geo-
70 graphic region by introducing censoring to focus on tail behavior and a random partition similar to Kim
71 et al. (2005) to remove long-range asymptotic dependence.

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

⁷³ To avoid bias in estimating tail parameters, we model censored data. Let

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

⁷⁴ 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
⁷⁵ is a pre-specified threshold value. Then, assuming the uncensored data $Y(\mathbf{s})$ are observations from a skew- t
⁷⁶ process, we update values censored below the threshold using standard Bayesian missing data methods as
⁷⁷ described in Section 4.

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

⁷⁹ We handle the problem of long-range asymptotic dependence with a random partition model. As discussed in
⁸⁰ Section 2, the source of long-range dependence is the shared z and σ . Therefore, to alleviate this dependence,
⁸¹ we allow z and σ to vary by site. Then, the model becomes

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

⁸² Let $\mathbf{w} = (w_1, w_2)$ be the location of a spatial knot. To model spatial variation, consider a set of spatial knots
⁸³ $\mathbf{w}_1, \dots, \mathbf{w}_K$ from a homogeneous Poisson process with intensity μ over spatial domain $\mathcal{D} \in \mathbb{R}^2$. The knots
⁸⁴ define a random daily partition of \mathcal{D} by subregions P_1, \dots, P_K defined as

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

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

$$z(\mathbf{s}) = z_k \quad (8)$$

$$\sigma(\mathbf{s}) = \sigma_k, \quad (9)$$

86 and the z_k and σ_k^2 are distributed as $z_k \stackrel{iid}{\sim} N(0, 1)$ and $\sigma^2 \stackrel{iid}{\sim} \text{IG}(a, b)$. So, within each partition, $Y(\mathbf{s})$
87 follows the spatial skew- t process defined in Section 2. Across partitions, the $Y(\mathbf{s})$ remain correlated via the
88 correlation function for $v(\mathbf{s})$ because it spans the partitions.

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

$$\begin{aligned} \chi(h) &= I(k_s = k_t)\chi_{\text{skew-}t}(h) + I(k_s \neq k_t)\chi_{\text{Gaus}}(h) \\ &= I(k_s = k_t)\chi_{\text{skew-}t}(h) \end{aligned} \quad (10)$$

91 where $\chi_{\text{skew-}t}(h)$ is the χ statistic for a skew- t process, $\chi_{\text{Gaus}}(h)$ is the χ statistic for a Gaussian process,
92 and $h = ||\mathbf{s} - \mathbf{t}||$. Therefore, sites in different subregions are asymptotically independent. Marginally, over
93 the knots $\mathbf{w}_1, \dots, \mathbf{w}_K$, $\chi(h) = \pi(h)\chi_{\text{skew-}t}(h)$, where $\pi(h) = \Pr(k_s = k_t)$ is the probability that two sites
94 separated by distance h are in the same partition. So, to show that $\lim_{h \rightarrow \infty} \chi(h) = 0$, we need only know
95 that $\lim_{h \rightarrow \infty} \pi(h) = 0$. A proof of this is given in Appendix A.3. Figure 1 shows how partitioning reduces
96 the extremal dependence.

97 3.3 Extension to space-time data

98 When using daily measurements, the assumption of temporal independence is inappropriate. There are
99 several places where temporal dependence could be incorporated in the model, including the residual $v_t(\mathbf{s})$.

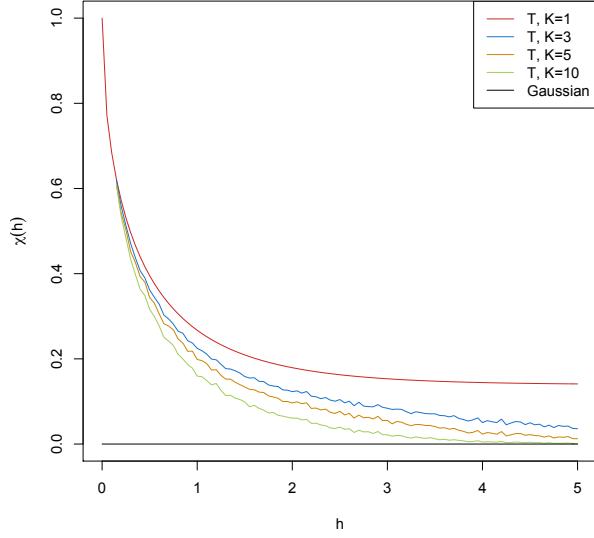


Figure 1: $\chi(h)$ for $K = 1, 3, 5$, and 10 partitions as a function of distance.

100 However, we choose to allow for temporal dependence in the \mathbf{w} , z , and σ terms because these terms dictate
 101 the tail behavior which is our primary focus. In this section, we extend (6) to the spatiotemporal setting. Let

$$Y_t(\mathbf{s}) = \mathbf{X}_t(\mathbf{s})^T \boldsymbol{\beta} + \lambda \sigma_t(\mathbf{s}) |z_t(\mathbf{s})| + \sigma_t(\mathbf{s}) v_t(\mathbf{s}), \quad (11)$$

102 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
 103 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
 104 partition P_{t1}, \dots, P_{tK} , and for $\mathbf{s} \in P_{tk}$,

$$z_t(\mathbf{s}) = z_{tk} \quad (12)$$

$$\sigma_t(\mathbf{s}) = \sigma_{tk}. \quad (13)$$

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

¹⁰⁶ a transformation to preserve the skew- t process at each time point. We first transform the spatial knots from
¹⁰⁷ \mathcal{D} to \mathcal{R}^2 as follows. Let

$$w_{tki}^* = \Phi^{-1} \left[\frac{(w_{tki} - \min(\mathbf{s}_i))}{\text{range}(\mathbf{s}_i)} \right], \quad i = 1, 2 \quad (14)$$

¹⁰⁸ where Φ is a univariate standard normal density function, and $\mathbf{s}_i = [s_{1i}, \dots, s_{ni}]$. Then $\mathbf{w}_{tk}^* \in \mathcal{R}^2$. We use
¹⁰⁹ a copula on $\sigma_t^2(\mathbf{s})$ to ensure that the marginal distributions of $\sigma_t^2(\mathbf{s})$ are inverse gamma. Let

$$\sigma_t^{2*}(\mathbf{s}) = \Phi^{-1} \{ \text{IG}[\sigma_t^2(\mathbf{s})] \} \quad (15)$$

¹¹⁰ where IG is the distribution function for an $\text{IG}(a, b)$ random variable. We also use a copula on $z_t(\mathbf{s})$ to
¹¹¹ ensure that the marginal distributions of $z_t(\mathbf{s})$ are half-normal. Let

$$z_t^*(\mathbf{s}) = \Phi^{-1} \{ \text{HN}[\sigma_t^2(\mathbf{s})] \} \quad (16)$$

¹¹² The AR(1) process for each tail 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
¹¹³ $t > 1$ the time series is modeled as

$$\mathbf{w}_{tk}^* | \mathbf{w}_{t-1,k}^* \sim N_2 [\phi_w \mathbf{w}_{t-1,k}^*, (1 - \phi_w^2)] \quad (17)$$

$$z_{tk}^* | z_{t-1,k}^* \sim N [\phi_z z_{t-1,k}^*, \sigma_{tk}^2 (1 - \phi_z^2)] \quad (18)$$

$$\sigma_{tk}^{2*} | \sigma_{t-1,k}^{2*} \sim N [\phi_\sigma \sigma_{t-1,k}^{2*}, (1 - \phi_\sigma^2)] \quad (19)$$

¹¹⁴ where $|\phi_w|, |\phi_z|, |\phi_\sigma| < 1$. These are stationary time series models with marginal distributions $\mathbf{w}_k^* \sim$
¹¹⁵ $N_2(0, 1)$, $z_k^* \sim N(0, \sigma_k^2)$, $\sigma_k^{2*} \sim N(0, 1)$. After transformation back to the original space, $\mathbf{w}_{tk} \sim \text{Unif}(\mathcal{D})$,

¹¹⁶ $z_{tk} \sim HN(0, \sigma_{tk}^2)$ $\sigma_{tk}^2 \sim IG(a, b)$. For each day, the model is identical to the spatial-only model in (6) by
¹¹⁷ construction.

¹¹⁸ **3.4 Hierarchical model**

¹¹⁹ Conditioned on $z_{tk}(\mathbf{s})$, $\sigma_{tk}^2(\mathbf{s})$, and P_{tk} , the marginal distributions are Gaussian and the joint distribution
¹²⁰ multivariate Gaussian. However, we do not fix the partitions, they are treated as unknown and updated in the
¹²¹ MCMC. We model this with a Bayesian hierarchical model as follows. Let $\mathbf{w}_{t1}, \dots, \mathbf{w}_{tK}$ be a set of daily
¹²² spatial knots in a spatial domain of interest, \mathcal{D} , and P_{tk} as defined in (7).

¹²³ Then

$$Y_t(\mathbf{s}) \mid z_t(\mathbf{s}), \sigma^2(\mathbf{s}), P_{tk}, \alpha, \beta, \Theta = \mathbf{X}_t(\mathbf{s})^T \beta + \lambda |z_t(\mathbf{s})| + \sigma_t(\mathbf{s}) v_t(\mathbf{s}) \quad (20)$$

$$z_t(\mathbf{s}) = z_{tk} \text{ if } \mathbf{s} \in P_{tk} \quad (21)$$

$$\sigma_t^2(\mathbf{s}) = \sigma_{tk}^2 \text{ if } \mathbf{s} \in P_{tk} \quad (22)$$

$$\lambda = \lambda_1 \lambda_2 \quad (23)$$

$$\lambda_1 = \begin{cases} +1 & \text{w.p.0.5} \\ -1 & \text{w.p.0.5} \end{cases} \quad (24)$$

$$\lambda_2^2 \sim IG(a, b) \quad (25)$$

$$v_t(\mathbf{s}) \mid \Theta \sim \text{Matérn}(0, \Sigma) \quad (26)$$

$$z_{tk}^* \mid z_{t-1,k}^*, \sigma_{tk}^2 \sim N(\phi_z z_{t-1,k}^*, \sigma_{tk}^2(1 - \phi_z^2)) \quad (27)$$

$$\sigma_{tk}^{2*} \mid \sigma_{t-1,k}^{2*} \sim N(\phi_\sigma \sigma_{t-1,k}^{2*}, (1 - \phi_\sigma^2)) \quad (28)$$

$$\mathbf{w}_{tk}^* \mid \mathbf{w}_{t-1,k}^* \sim N_2(\phi_w \mathbf{w}_{t-1,k}^*, (1 - \phi_w^2)) \quad (29)$$

¹²⁴ where $\Theta = \{\rho, \nu, \gamma\}$, and Σ is a Matérn covariance matrix as described in Section 2.1. We parameterize

125 $\lambda = \lambda_1 \lambda_2$ to help with convergence in the MCMC.

126 **4 Computation**

127 First, we impute values below the threshold. Then, we update Θ using Metropolis Hastings or Gibbs sam-
128 pling when appropriate. Finally, we make spatial predictions using conditional multivariate normal results
129 and the fact that the distribution of $Y_t(\mathbf{s}) \mid \Theta, z(\mathbf{s})$ is the usual multivariate normal distribution with a
130 Matérn spatial covariance structure.

131 We can use Gibbs sampling to update $Y_t(\mathbf{s})$ for censored observations that are below the threshold T .

132 After conditioning on $\lambda, z_t(\mathbf{s})$ and non-censored observations, $Y_t(\mathbf{s})$ has truncated normal full conditionals.

133 So we sample $Y_t(\mathbf{s}) \sim N_{(-\infty, T)}(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$. After imputing the censored observations, we update the model
134 parameters. To update the model parameters, we use standard Gibbs updates for parameters when possible.

135 In the case Gibbs sampling is not possible, parameters are updated using a random-walk Metropolis Hastings
136 algorithm. See Appendices A.1 and A.2 for details regarding the MCMC. The final step of the computation
137 is to use Bayesian Kriging to generate a predictive distribution for $Y_t(\mathbf{s}^*)$ at prediction location \mathbf{s}^* . This
138 step is similar to the imputation for censored observations except that the full conditionals are no longer
139 truncated at T .

140 **5 Simulation study**

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

¹⁴³ **5.1 Design**

¹⁴⁴ For all simulation designs, we generate data from the model in Section 3.2 using $n_s = 144$ sites and
¹⁴⁵ $n_t = 50$ independent days. The sites are generated Uniform($[0, 10] \times [0, 10]$). We generate data from 7
¹⁴⁶ different simulation designs:

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

¹⁴⁸ 2. T marginal, $K = 1$ knot

¹⁴⁹ 3. T marginal, $K = 5$ knots

¹⁵⁰ 4. Skew- t marginal, $K = 1$ knots

¹⁵¹ 5. Skew- t marginal, $K = 5$ knots

¹⁵² 6. Max-stable

¹⁵³ 7. Transformation below $T = q(0.80)$

¹⁵⁴ In the first five designs, the $v_t(\mathbf{s})$ terms are generated using a Matérn covariance with smoothness parameter

¹⁵⁵ $\nu = 0.5$ and spatial range $\rho = 1$. For the covariance matrices in designs 1 – 5, the proportion of the variance

¹⁵⁶ accounted for by the spatial variation is $\gamma = 0.9$ while the proportion of the variance accounted for by the

¹⁵⁷ nugget effect is 0.1. In the first design, $\sigma^2 = 2$ is used for all days. For designs 2 – 4, $\sigma_{tk}^2 \stackrel{iid}{\sim} \text{IG}(3, 8)$

¹⁵⁸ For designs 1 – 3, we set $\lambda = 0$. For designs four and five, $\lambda = 3$ was used, and the z_t are generated as

¹⁵⁹ described in (8). In the sixth design, we generate from a spatial max-stable distribution (Reich and Shaby,

¹⁶⁰ 2012). In this design, data have marginal distributions that follow a generalized extreme value distribution

¹⁶¹ with parameters $\mu = 1, \sigma = 1, \xi = 0.2$. In this model, a random effect is used to induce spatial dependence

¹⁶² using 144 spatial knots on a regular lattice in the square $[1, 9] \times [1, 9]$. For this setting, we set $\gamma = 0.5$. In

163 the final design, we generate \tilde{y} using the setting from design 4, and then consider the data

$$y = \begin{cases} \tilde{y}, & \tilde{y} > T \\ T \exp\{\tilde{y} - T\}, & \tilde{y} \leq T \end{cases} \quad (30)$$

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

166 $M = 50$ data sets are generated for each design. For each data set we fit the data using

- 167 1. Gaussian marginal, $K = 1$ knots
- 168 2. Skew- t marginal, $K = 1$ knots, $T = -\infty$
- 169 3. Symmetric- t marginal, $K = 1$ knots, $T = q(0.80)$
- 170 4. Skew- t marginal, $K = 5$ knots, $T = -\infty$
- 171 5. Symmetric- t marginal, $K = 5$ knots, $T = q(0.80)$

172 where $q(0.80)$ is the 80th sample quantile of the data. The design matrix \mathbf{X} includes an the intercept with a
173 prior of $\beta \sim N(0, 10)$. The spatial covariance parameters have priors $\log(\nu) \sim N(-1.2, 1)$, $\gamma \sim Unif(0, 1)$,
174 $\rho \sim Unif(15)$. The skewness parameter has prior $\lambda \sim N(0, 2)$. The residual variance terms have priors
175 $\sigma_t^2(s) \sim IG(0.1, 0.1)$. The knots have priors $\mathbf{w} \sim Unif(\mathcal{D})$. We do not fit the data using the max-stable
176 methods from Reich and Shaby (2012) because of the time it takes.

177 5.2 Cross validation

178 Models were compared using cross validation with 100 sites used as training sites and 44 sites withheld for
179 testing. The model was fit using the training set, and predictions were generated at the testing site locations.
180 Because one of the primary goals of this model is to predict extreme events, we use Brier scores to select
181 the model that best fits the data (Gneiting and Raftery, 2007). The Brier score for predicting exceedance of

182 a threshold c is given by $[e(c) - P(c)]^2$ where $e(c) = I[y > c]$ is an indicator function indicating that a test
183 set value, y , has exceeded the threshold, c , and $P(c)$ is the predicted probability of exceeding c . We average
184 the Brier scores over all test sites and days. For the Brier score, a lower score indicates a better fit.

185 **5.3 Results**

186 We compared the Brier scores for exceeding 11 different thresholds for each dataset. The thresholds used
187 for the Brier scores are extreme quantiles from the simulated data. Figure 2 gives the Brier score relative to
188 the Brier score for the Gaussian method calculated as

$$BS_{\text{rel}} = \frac{BS_{\text{method}}}{BS_{\text{Gaussian}}}. \quad (31)$$

189 We analyzed the results for the simulation study using two-sided Wilcoxon signed rank tests at a signifi-
190 cance level of $\alpha = 0.001$. When the data come from a Gaussian distribution, the Gaussian method performs
191 significantly better than the other methods for quantiles below $q(0.95)$. However, as we move further out in
192 the tails, the difference in Brier scores is no longer significant beyond $q(0.95)$ when compared to method
193 two and beyond $q(0.99)$ for methods three through five. When the data are generated using method two, our
194 methods perform significantly better than the Gaussian method in all cases except for $q(0.90)$ with method
195 five. For data setting three, method four performs significantly better at all quantile levels, and method
196 five performs significantly better for quantiles more extreme than $q(0.93)$. In this setting, methods two and
197 three do not show significant improvement over the Gaussian method until $q(0.995)$. In setting four, method
198 two shows a significant improvement over the Gaussian method at all quantile levels. Methods three and
199 four show improvement over the Gaussian method at quantile levels below $q(0.94)$. The Gaussian method
200 performs significantly better than method five for all quantile levels in data setting four. For setting five,
201 methods three and four perform significantly better than the Gaussian method at all quantile levels. Method

202 one performs better than the Gaussian method at quantiles higher than $q(0.99)$, and method two performs
 203 better than the Gaussian method at quantiles below $q(0.98)$. For setting six, the Gaussian method performs
 204 significantly better than all methods for quantiles $q(0.90) - q(0.94)$. However, for quantiles $q(0.99)$ and
 205 $q(0.995)$, methods two and three outperform the Gaussian method. Finally, in setting seven, methods three
 206 demonstrate a significant improvement over the Gaussian method for all quantiles, and method two demon-
 207 strates a significant improvement of the Gaussian method for all quantiles except $q(0.94)$. Method four
 208 performs significantly better than the Gaussian method for $q(0.90)$, $q(0.92)$, and $q(0.95) - q(0.99)$. Finally,
 209 method five for setting seven shows no significant difference in Brier scores from the Gaussian method.

210 **6 Data analysis**

211 To illustrate this method, we consider the daily maximum 8-hour ozone measurements for July 2005 at 735
 212 Air Quality System (AQS) monitoring sites in the eastern United States as the response (see Figure 3). For
 213 each site, we also have covariate information containing the estimated ozone from the Community Multi-
 214 scale Air Quality (CMAQ) modeling system. Initially, we fit a linear regression assuming a mean function
 215 of

$$\mathbf{X}\boldsymbol{\beta} = \beta_0 + \beta_1 \cdot \text{CMAQ}_t(\mathbf{s}). \quad (32)$$

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

218 We explore spatial and temporal extremal dependence by considering $\chi_c = \Pr[Y(\mathbf{s}) > c | Y(\mathbf{t}) > c]$. To
 219 examine spatial dependence in high quantiles, we consider observations at all pairs of sites \mathbf{s} and \mathbf{t} that are
 220 distance h apart where h is separated into bins of size 0.25 km. Then conditioned on $Y(\mathbf{t}) > c$, we take

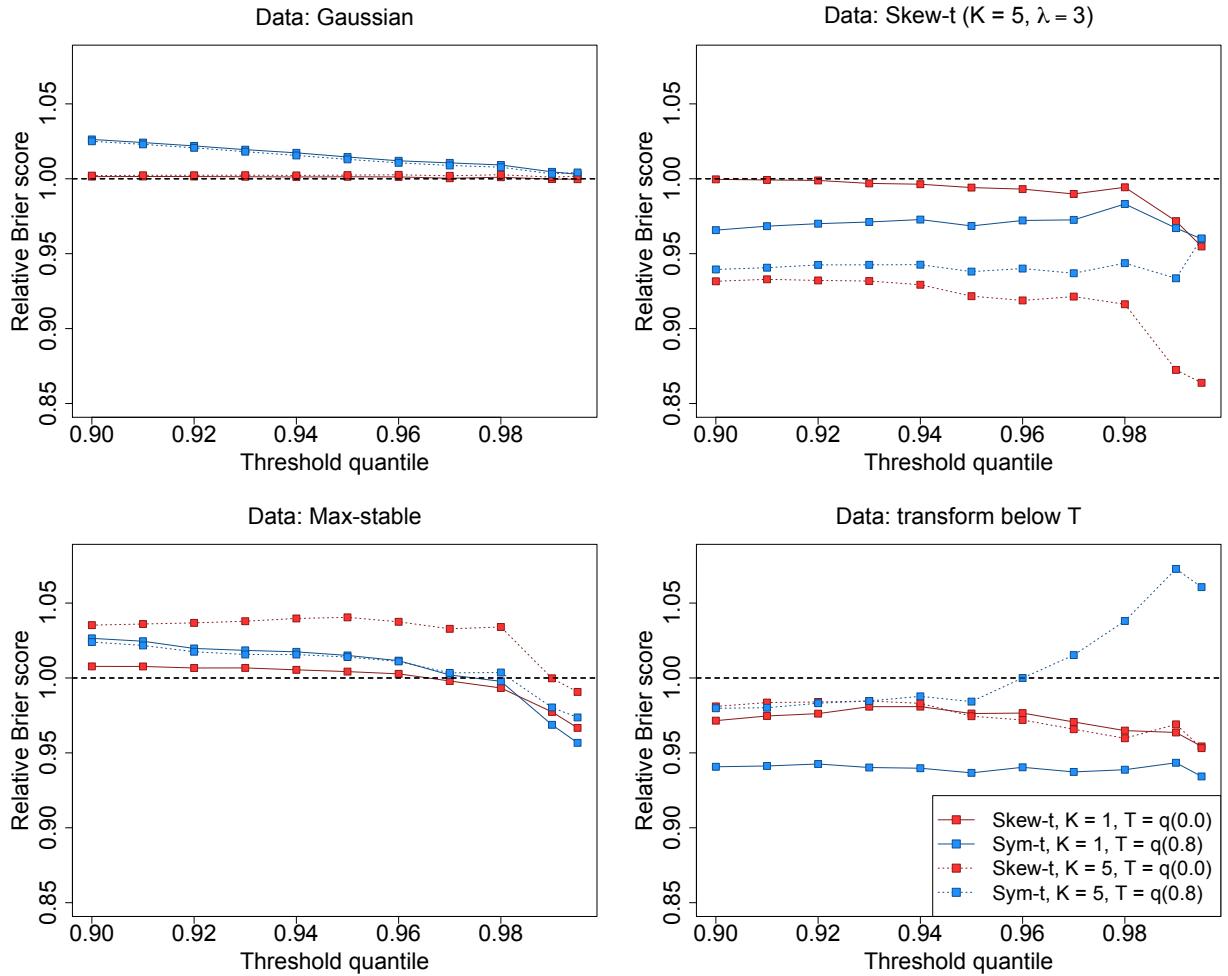


Figure 2: Brier scores relative to the Gaussian method for simulation study results. A ratio lower than 1 indicates that the method outperforms relative to the Gaussian method.

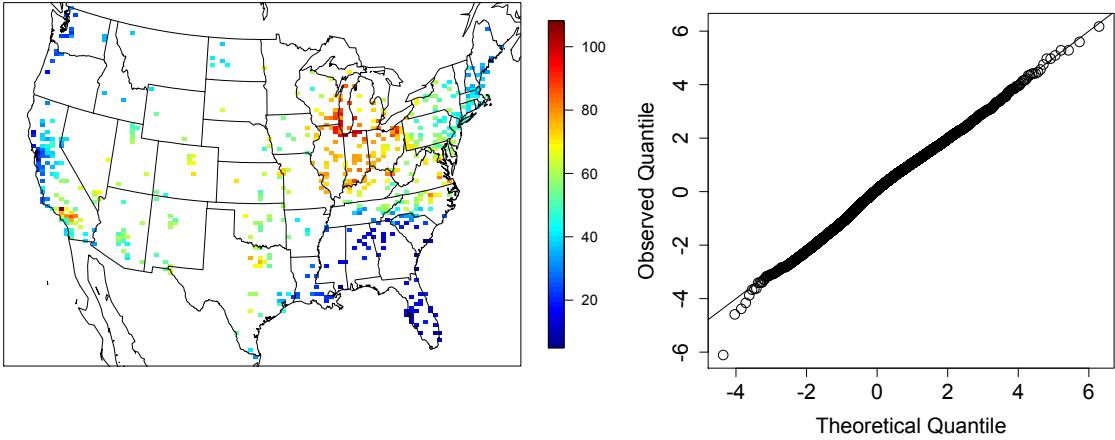


Figure 3: Ozone values on 10 July 2005 (left) Q-Q plot of the residuals (right)

the sample proportion of $Y(\mathbf{s}) > c$. Finally, $\hat{\chi}_c(h)$ is averaged over all days at each of the three threshold quantiles. To examine temporal dependence in high quantiles, we consider observations at a single site that are taken lag- t days apart. Then conditioned on $Y_n(\mathbf{s}) > c$, we take the sample proportion of $Y_{n+t}(\mathbf{s}) > c$. Finally, $\hat{\chi}(t)$ is averaged over all sites at each of the three threshold quantiles. The $\hat{\chi}_c(h)$ and $\hat{\chi}_c(t)$ plots in Figure 4 show the estimated spatial and temporal dependence of the residuals for the ozone data at three quantile levels $q(0.90)$, $q(0.95)$, and $q(0.99)$. Both plots indicate that there is dependence in the high quantile levels beyond what we expect if the residuals were independent.

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 $T = 0$, $\hat{q}(0.42) = 50$, $\hat{q}(0.92) = 75$ ppb in order to compare results from no, moderate, and high censoring. We also compare models with no time series to models that include the time series. Finally, as a comparison to max-stable methods, we fit the model using the hierarchical max-stable model of Reich and Shaby (2012). All methods assume the mean function given in (32). For each model,

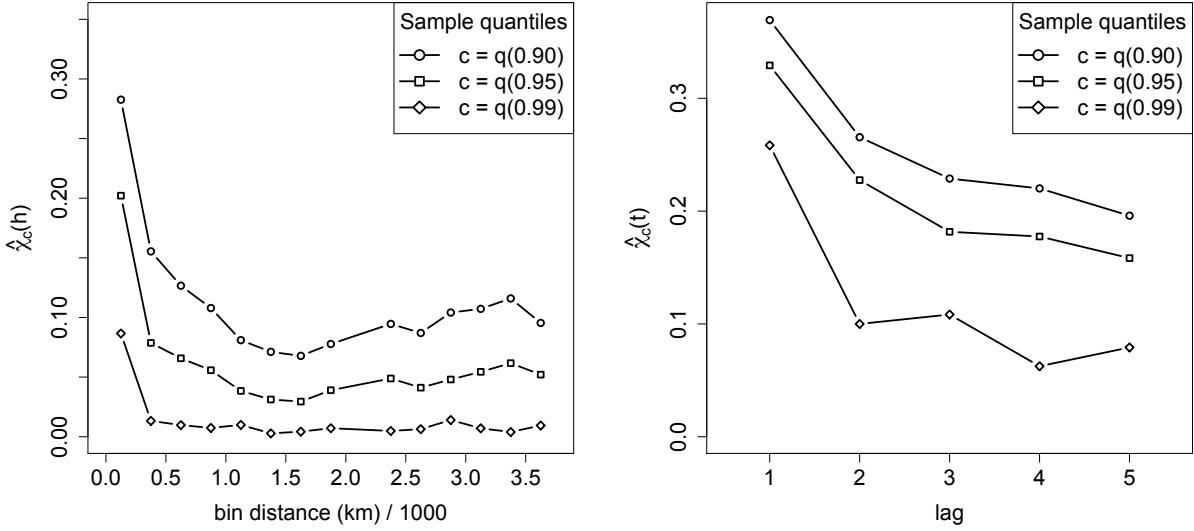


Figure 4: $\hat{\chi}_c(h)$ plot for the residuals (left). $\hat{\chi}_c(t)$ plot for the residuals (right).

234 Brier scores were averaged over all sites and days to obtain a single Brier score for each dataset. At a
 235 particular threshold or quantile level, the model that fits the best is the one with the lowest score. We then
 236 compute the relative Brier scores to compare each model.

237 **6.2 Results**

238 The plots in Figure 5 shows the relative Brier scores (see Section 5.3) for time-series and non-time-series
 239 models, using $K = 1, 7$, and 15 knots at thresholds $T = 0, 50$, and 75 ppb. Both plots have similar features
 240 suggesting that most settings do reasonably well. In particular, for all extreme quantiles, selecting a moder-
 241 ate number of knots (e.g. $K = 5, \dots, 10$) tends to give the best results. Table 1 shows the best two models
 242 for selected extreme quantiles. The results demonstrate the importance of accounting for the temporal de-
 243 pendence when making extreme predictions. They also show the importance of thresholding for predictions
 244 further out in the tails of the data. This is further illustrated by the plots in Figure 6.

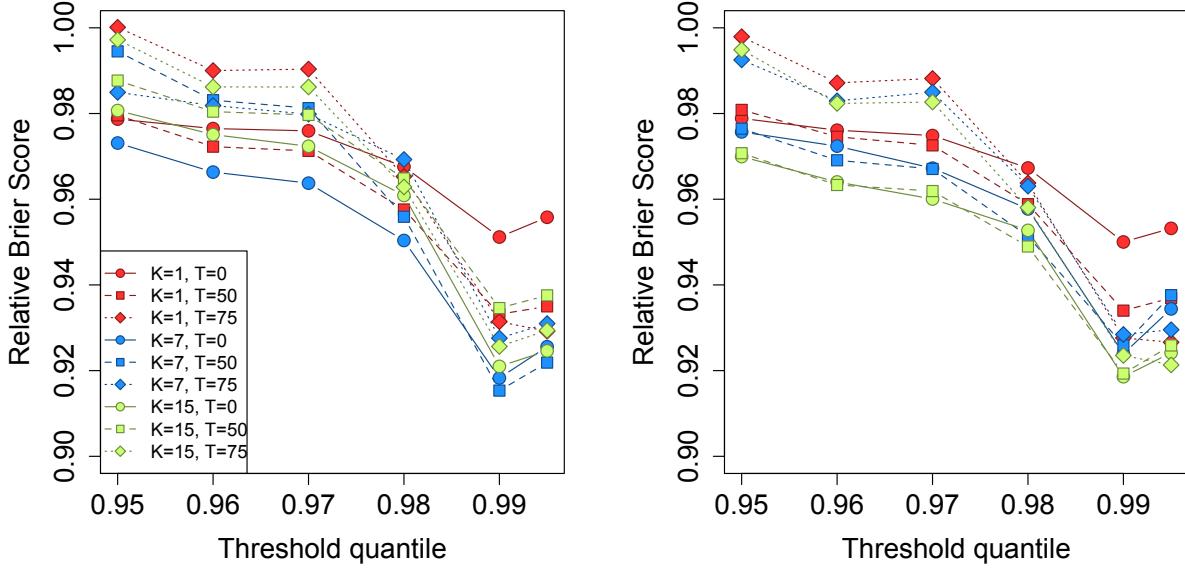


Figure 5: Relative Brier scores for time-series models (left) and non-time-series models (right). Relative brier score for the max-stable model is between 1.07 and 1.15

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

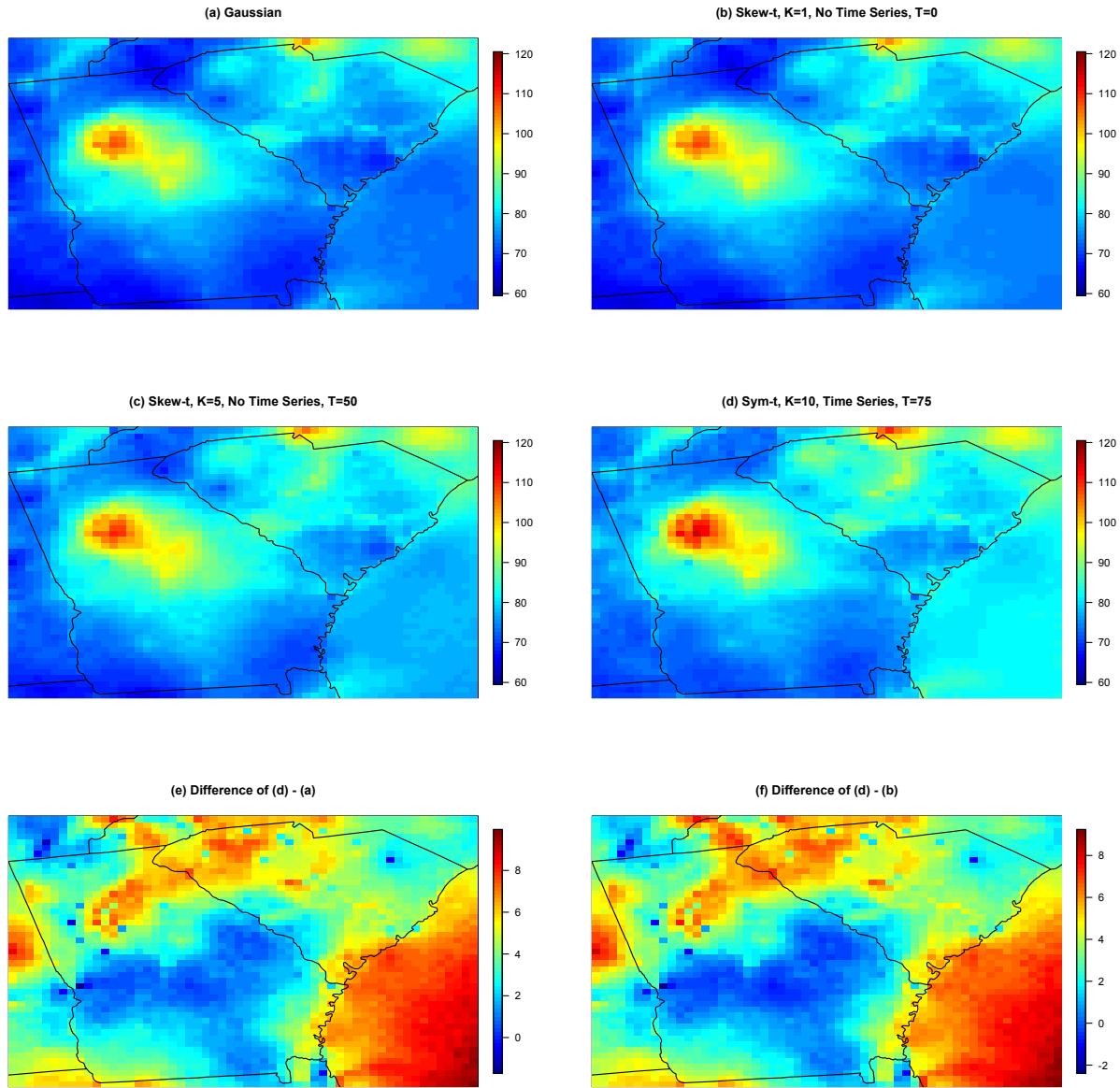


Figure 6: (a) – (d) give the posterior predictive $\hat{q}(0.99)$ for the month of July under four different models, (e) gives the difference between $\hat{q}(0.99)$ in plots (d) and (a), (f) gives the difference between $\hat{q}(0.99)$ in plots (d) and (b).

245 **7 Discussion**

246 In this paper we propose a new approach for spatiotemporal modeling of extreme values. The proposed model
247 gives flexible tail behavior, demonstrates asymptotic dependence for observations at sites that are near to one
248 another, and has computation on the order of Gaussian models for large space-time datasets – Do I need to
249 add anything about the computation in the paper?. In the simulation study, we demonstrate that this model
250 shows statistically significant improvements over a naïve Gaussian approach. In both the simulation study,
251 and the application to ozone data, we find that incorporating a partition in the model improves extreme
252 prediction. Furthermore the results suggest that thresholding can improve performance when predicting in
253 the extreme tails of the data.

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

261 **Acknowledgments**

262 **A Appendices**

263 **A.1 MCMC details**

264 The MCMC sampling for the model 3.4 is done using R (<http://www.r-project.org>). Whenever possible,
265 we select conjugate priors (see Appendix A.2); however, for some of the parameters, no conjugate prior
266 distributions exist. When no conjugate prior distribution exists, we use a random walk Metropolis Hastings
267 update step. In each Metropolis Hastings update, we tune the algorithm to give acceptance rates near 0.40.

268 **Spatial knot locations**

269 For each day, we update the spatial knot locations, $\mathbf{w}_1, \dots, \mathbf{w}_K$, using a Metropolis Hastings block up-
270 date. Because the spatial domain is bounded, we generate candidate knots using the transformed knots
271 $\mathbf{w}_1^*, \dots, \mathbf{w}_K^*$ (see section 3.3) and a random walk bivariate Gaussian candidate distribution

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

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

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

274 where l is the likelihood given in (20), and $p(\cdot)$ is the prior either taken from the time series given in (3.3)
275 or assumed to be uniform over \mathcal{D} . The candidate knots are accepted with probability $\min\{R, 1\}$.

276 **Spatial random effects**

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

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

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

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

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

284 **Variance terms**

285 When there is more than one site in a partition, then we update σ_{tk}^2 using a Metropolis Hastings update.
 286 First, we generate a candidate for σ_{tk}^2 using an $IG(a^*/s, b^*/s)$ candidate distribution in an independence
 287 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
 288 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
 289 day t . The acceptance ratio is

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

290 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
 291 $c[\cdot]$ is the candidate distribution. The candidate is accepted with probability $\min\{R, 1\}$.

292 **Spatial covariance parameters**

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

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

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

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

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

298 **A.2 Posterior distributions**

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

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

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

302 Let

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

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

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

303 Then

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

304 where

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

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

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

³⁰⁶ **Conditional posterior of β | ...**

³⁰⁷ Let $\beta \sim N_p(0, \Lambda_0)$ where Λ_0 is a precision matrix. Then

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

³⁰⁸ where

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

³⁰⁹ **Conditional posterior of σ^2 | ...**

³¹⁰ In the case where $L = 1$ and temporal dependence is negligible, then σ^2 has a conjugate posterior distribution.

³¹¹ Let $\sigma_t^2 \stackrel{iid}{\sim} IG(\alpha_0, \beta_0)$. For simplicity, drop the subscript t . Then

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

³¹² where

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

³¹³ In the case that $L > 1$, a random walk Metropolis Hastings step will be used to update σ_{lt}^2 .

³¹⁴ **Conditional posterior of $\lambda | \dots$**

³¹⁵ For convergence purposes we model $\lambda = \lambda_1 \lambda_2$ where

$$\lambda_1 = \begin{cases} +1 & \text{w.p.0.5} \\ -1 & \text{w.p.0.5} \end{cases} \quad (33)$$

$$\lambda_2^2 \sim IG(\alpha_\lambda, \beta_\lambda). \quad (34)$$

$$(35)$$

³¹⁶ Then

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

³¹⁷ Then $\lambda_2 | \dots \sim IG(\alpha_\lambda + kt, \beta_\lambda + \frac{z^2}{2\sigma_{tk}^2})$

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

319 Let $N(A)$ be the number of knots in A , the area between sites s_1 and s_2 . Consider a spatial Poisson process
320 with intensity $\mu(A)$. So,

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

321 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
322 in A , the chance that s_1 and s_2 will be in the same partition will decrease, because partition membership
323 is defined by the closest knot to a site. Therefore, $\lim_{h \rightarrow \infty} \pi(h) = 0$.

324 **A.4 Skew-t distribution**

325 **Univariate extended skew-t distribution**

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

$$f_{EST}(y) = \omega^{-1} \frac{f_T(z; \nu)}{F_T(\tau/\sqrt{1+\alpha^2}; \nu)} F_T \left[(\alpha z + \tau) \sqrt{\frac{\nu+1}{\nu+z^2}}; 0, 1, \nu+1 \right] \quad (36)$$

328 where $f_T(t; \nu)$ is a univariate Student's t with ν degrees of freedom, $F_T(t; \nu) = P(T < t)$, and $z = (y - \xi)/\omega$.
329 In the case that $\tau = 0$, then Y follows a univariate skew-t distribution.

330 **Multivariate skew- t distribution**

331 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 $\boldsymbol{\omega} = \text{diag}(\omega_1, \dots, \omega_d)$,
 332 then the density of Y at y is

$$f_y(\mathbf{y}) = \det(\boldsymbol{\omega})^{-1} f_z(\mathbf{z}) \quad (37)$$

333 where

$$f_z(\mathbf{z}) = 2t_d(\mathbf{z}; \bar{\Omega}, \eta) T \left[\boldsymbol{\alpha}^T \mathbf{z} \sqrt{\frac{\eta+d}{\nu+Q(\mathbf{z})}}; \eta+d \right] \quad (38)$$

$$\mathbf{z} = \boldsymbol{\omega}^{-1}(\mathbf{y} - \xi) \quad (39)$$

334 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
 335 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
 336 freedom (Azzalini and Capitanio, 2013).

337 **Extremal dependence**

338 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

$$\chi(h) = \bar{F}_{\text{EST}} \left\{ \frac{[x_1^{1/\eta} - \varrho(h)]\sqrt{\eta+1}}{\sqrt{1-\varrho(h)^2}}; 0, 1, \alpha_1, \tau_1, \eta+1 \right\} + \bar{F}_{\text{EST}} \left\{ \frac{[x_2^{1/\eta} - \varrho(h)]\sqrt{\eta+1}}{\sqrt{1-\varrho(h)^2}}; 0, 1, \alpha_2, \tau_2, \eta+1 \right\}, \quad (40)$$

339 where \bar{F}_{EST} is the univariate survival extended skew- t function with zero location and unit scale, $\varrho(h) = \text{cor}(y_1, y_2)$,
 340 $\alpha_j = \alpha_i \sqrt{1-\varrho^2}$, $\tau_j = \sqrt{\eta+1}(\alpha_j + \alpha_i \varrho)$, 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
 341 $j = 1, 2$ and $i = 2, 1$ and where $\bar{\alpha}_j = (\alpha_j + \alpha_i \varrho)/\sqrt{1+\alpha_i^2[1-\varrho(h)^2]}$.

³⁴² **Proof that** $\lim_{h \rightarrow \infty} \chi(h) > 0$

³⁴³ Consider the bivariate distribution of $\mathbf{Y} = [Y(\mathbf{s}), Y(\mathbf{t})]^T$, with $\varrho(h)$ given by (2). So, $\lim_{h \rightarrow \infty} \varrho(h) = 0$.

³⁴⁴ Then

$$\lim_{h \rightarrow \infty} \chi(h) = \bar{F}_{\text{EST}} \left\{ \sqrt{\eta + 1}; 0, 1, \alpha_1, \tau_1, \eta + 1 \right\} + \bar{F}_{\text{EST}} \left\{ \sqrt{\eta + 1}; 0, 1, \alpha_2, \tau_2, \eta + 1 \right\}. \quad (41)$$

³⁴⁵ Because the extended skew-*t* distribution is not bounded above, for all $\bar{F}_{\text{EST}}(x) = 1 - F_{\text{EST}} > 0$ for all

³⁴⁶ $x < \infty$. Therefore, for a skew-*t* distribution, $\lim_{h \rightarrow \infty} \chi(h) > 0$.

³⁴⁷ References

³⁴⁸ Azzalini, A. and Capitanio, A. (2013) *The Skew-Normal and Related Families*. Institute of Mathematical
³⁴⁹ Statistics Monographs. Cambridge University Press.

³⁵⁰ Blanchet, J. and Davison, A. C. (2011) Spatial modeling of extreme snow depth. *The Annals of Applied
351 Statistics*, **5**, 1699–1725.

³⁵² Coles, S. G. and Tawn, J. A. (1991) Modelling Extreme Multivariate Events. *Journal of the Royal Statistical
353 Society: Series B (Methodological)*, **53**, 377–392.

³⁵⁴ DuMouchel, W. H. (1983) Estimating the stable index α in order to measure tail thickness: a critique. *The
355 Annals of Statistics*, **11**, 1019–1031.

³⁵⁶ Genton, M. G. (2004) *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*.
³⁵⁷ Statistics (Chapman & Hall/CRC). Taylor & Francis.

³⁵⁸ Gneiting, T. and Raftery, A. E. (2007) Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of
359 the American Statistical Association*, **102**, 359–378.

- 360 Huser, R. (2013) *Statistical Modeling and Inference for Spatio-Temporal Extremes*. Ph.D. thesis.
- 361 Kim, H.-M., Mallick, B. K. and Holmes, C. C. (2005) Analyzing Nonstationary Spatial Data Using Piece-
362 wise Gaussian Processes. *Journal of the American Statistical Association*, **100**, 653–668.
- 363 Padoan, S. A. (2011) Multivariate extreme models based on underlying skew- and skew-normal distribu-
364 tions. *Journal of Multivariate Analysis*, **102**, 977–991.
- 365 Padoan, S. A., Ribatet, M. and Sisson, S. A. (2010) Likelihood-Based Inference for Max-Stable Processes.
366 *Journal of the American Statistical Association*, **105**, 263–277.
- 367 Reich, B. J. and Shaby, B. A. (2012) A hierarchical max-stable spatial model for extreme precipitation. *The
368 Annals of Applied Statistics*, **6**, 1430–1451.
- 369 Zhang, H. and El-Shaarawi, A. (2010) On spatial skewGaussian processes and applications. *Environmetrics*,
370 **21**, 33–47.