A new spatial model for points above a threshold

November 14, 2014

1 Introduction

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

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

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

In this paper, we present a model that has marginal distributions with flexible tails, demonstrates asymptotic dependence for observations at sites that are near to one another, and has computation on the order of Gaussian models for large space-time datasets. Specifically, our contribution is to incorporate thresholding and random spatial partitions using a multivariate skew-t distribution. The advantage of using a thresholded model as opposed to a non-thresholded model is that is allows for the tails of the distribution to inform the predictions in the tails (DuMouchel, 1983). The random spatial partition alleviates the long-range spatial dependence seen by the skew-t.

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

2 Spatial skew processes

Many types of data demonstrate some level of skewness and therefore should be modeled with distributions that allow for non-symmetry. The skew-elliptical family of distributions provides models that are mathematically tractable while introducing a slant parameter α to account for non-symmetric data (Genton, 2004).

2.1 Skew-normal process

Let $U(\mathbf{s})$ be a normalized skew-normal process with slant α. We write $U(\mathbf{s}) \sim \text{SN}_d(\Omega, \alpha)$, and the density is given as

$$2\phi_d(U(\mathbf{s}); \Omega)\Phi(\alpha \mathbf{1}^T U(\mathbf{s})) \tag{1}$$

where ϕ_d is the density function for a $N_d(0, \Sigma)$ random variable, and Φ is a standard normal cumulative distribution function. There is an additive representation (Zhang and El-Shaarawi, 2010) to construct a standardized skew-normal process given by

$$U(\mathbf{s}) = \alpha|z| + v(\mathbf{s}) \tag{2}$$

where $\alpha \in \mathcal{R}$ is a parameter controlling skew, $z \sim N(0, 1)$, and $v(\mathbf{s})$ is a Gaussian process with mean zero, variance one, Matérn spatial correlation with spatial range ρ, smoothness ν, and γ is the proportion of variance accounted for by the spatial variation. So,

$$\text{cor}(U(\mathbf{s}), U(\mathbf{s})) = \gamma I(s = t) + (1 - \gamma)\frac{1}{\Gamma(\nu)2^{\nu-1}}\left(\sqrt{2\nu}\frac{||\mathbf{s} - \mathbf{t}||}{\rho}\right)^{\nu} K_{\nu}\left(\sqrt{2\nu}\frac{||\mathbf{s} - \mathbf{t}||}{\rho}\right). \tag{3}$$

Let $Y(\mathbf{s}) = \mathbf{X}(\mathbf{s})^T + \sigma U(\mathbf{s})$. Then, after marginalizing over the z term, we find that

$$Y(\mathbf{s}) \sim \text{SN}_d(\mathbf{X}^T(\mathbf{s})\boldsymbol{\beta}, \Omega, \alpha) \tag{4}$$

where $\Omega = \sigma^2(\Sigma + \alpha^2 \mathbf{1}\mathbf{1}^T)$.

2.2 Skew-t process

If $\sigma^2 \sim \text{IG}(a, b)$, after over the σ^2 terms, we have a skew-t process with distribution function

$$\mathbf{Y} \sim \text{skew-}t(\mu, \boldsymbol{\Sigma}^*, \alpha, \text{df} = 2a)$$

where μ is the location, $\boldsymbol{\Sigma}^* = \frac{b}{a}\boldsymbol{\Sigma}$, $\boldsymbol{\Sigma}$ is a Matérn covariance matrix, and $\alpha \in \mathcal{R}$ controls the skewness. The skew-t process is desirable because of its flexible tail that is controlled by the skewness parameter, α, and the degrees of freedom, $2a$. Furthermore, if $\mathbf{Y}$ follows a multivariate skew-t distribution, the marginal distributions also follow a skew-t distribution (Azzalini and Capitanio, 2003).

One member of this family, is the skew-t distribution (Azzalini and Capitanio, 2003, see Appendix A.5). Zhang and El-Shaarawi (2010) show that the skew-t process can be expressed as a hierarchical model

$$Y(\mathbf{s}) = \mathbf{X}(\mathbf{s})^T\boldsymbol{\beta} + \alpha|z| + \sigma v(\mathbf{s}) \tag{5}$$

where $\alpha \in \mathcal{R}$ controls the skewness; $z \sim N(0, \sigma^2)$ is a normal random effect; $v(\mathbf{s})$ is a Gaussian process with mean zero, variance one, Matérn spatial correlation with spatial range ρ, smoothness ν, and γ is the proportion of variance accounted for by the spatial variation; and $\sigma^2 \sim \text{IG}(a, b)$. So,

$$\text{cor}(Y(\mathbf{s}), Y(\mathbf{s})) = \gamma I(s = t) + (1 - \gamma)\frac{1}{\Gamma(\nu)2^{\nu-1}}\left(\sqrt{2\nu}\frac{||\mathbf{s} - \mathbf{t}||}{\rho}\right)^{\nu} K_{\nu}\left(\sqrt{2\nu}\frac{||\mathbf{s} - \mathbf{t}||}{\rho}\right) \tag{6}$$

2.3 Extremal dependence

One measure of extremal dependence is the χ statistic (Padoan, 2011). The χ statistic for the upper tail is given by

$$\chi = \lim_{c \to \infty} \Pr(Y(\mathbf{s}_1) > c | Y(\mathbf{s}_2) > c).$$

In a stationary spatial process, we can write the χ coefficient as

$$\chi(h) = \lim_{c \to \infty} \Pr(Y(0) > c | Y(h) > c)$$

where $h = ||\mathbf{s}_1 - \mathbf{s}_2||$. If $\chi(h) = 0$, then observations are asymptotically independent at distance h. For Gaussian processes, $\chi(h) = 0$ regardless of the the distance, so they are not suitable for modeling spatially dependent extremes. However, for the skew-t process described in 2, $\chi(h) > 0$ (Padoan, 2011, they give an expression that is the sum of two survival functions from the skew-t distribution. The expression isn't terribly informative, but I can include it if you think it would be helpful.).

3 Spatiotemporal skew-t model for extremes

In this section, we describe the

3.1 Censoring to focus on the tails

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

$$\tilde{Y}(\mathbf{s}) = \begin{cases} Y(\mathbf{s}) & \delta(\mathbf{s}) = 1 \\ T & \delta(\mathbf{s}) = 0 \end{cases} \tag{7}$$

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

3.2 Partitioning to remove long-range asymptotic dependence

One problem with the spatial skew-t process is that all sites are asymptotically dependent regardless of their spatial separation. This occurs because all observations, both near and far, share the same z and σ^2 terms. Even more problematic is that $\lim_{h \to \infty} \chi(h) \neq 0$. We handle this problem with a random partition similar to Kim et al. (2005) that allows z and σ^2 to vary by site.

Consider a set of spatial knots $\mathbf{w}_k \sim \text{Uniform}(\mathcal{D})$ that define a random daily partition $P_1, \dots, P_K$ of the spatial domain of interest $\mathcal{D} \subset \mathcal{R}^2$ such that

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

Then, the observation at each site becomes

$$Y(\mathbf{s}) = X(\mathbf{s})\boldsymbol{\beta} + \alpha|z(\mathbf{s})| + \sigma v(\mathbf{s}), \tag{8}$$

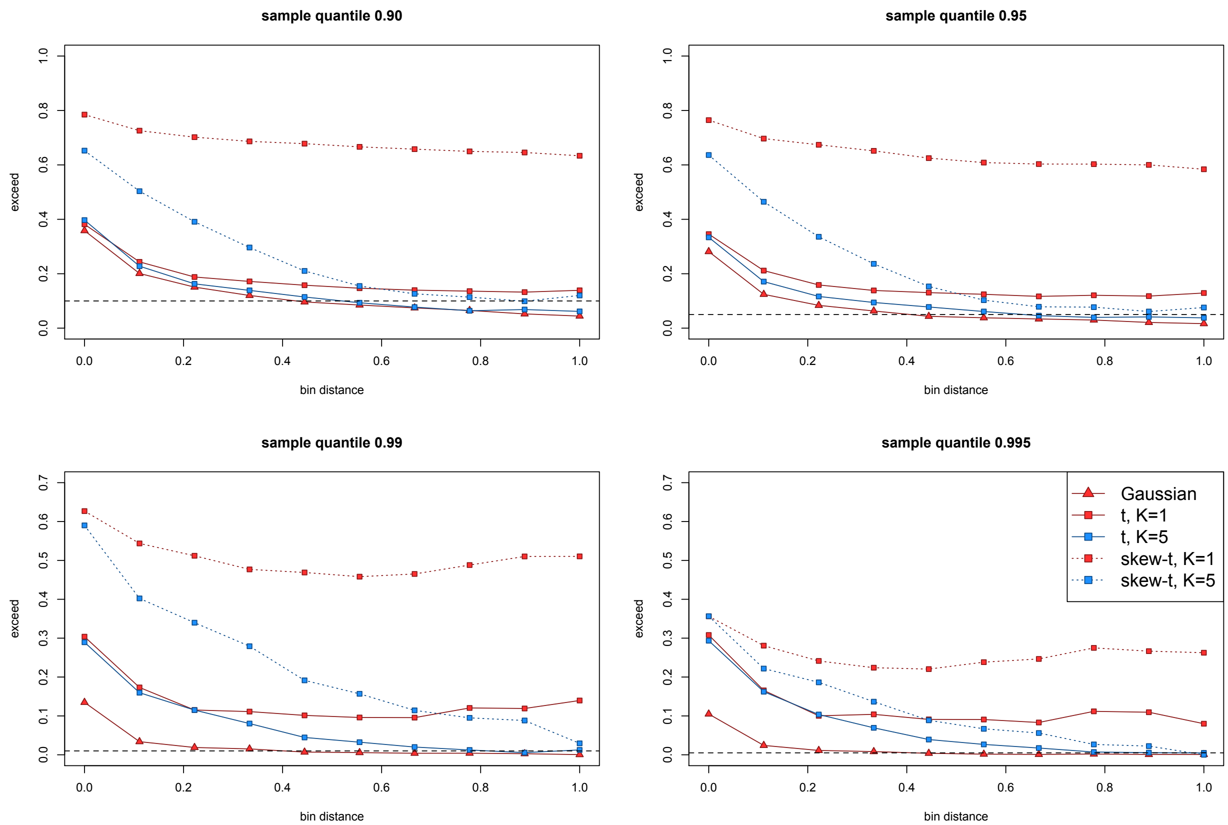

Figure 1: $\chi(h)$ for simulated datasets.

and, for $\mathbf{s} \in P_k$,

$$z(\mathbf{s}) = z_{tk} \tag{9}$$
$$\sigma(\mathbf{s}) = \sigma_k. \tag{10}$$

So, within each partition, $\mathbf{Y}$ follows the distribution given in (5).

When incorporating the random daily partition, for two sites in the same partition, the χ statistic is $\chi(h) = \chi_{\text{skew-}t}(h)$. However, for two sites in different partitions, the χ statistic should be the same as the Gaussian $\chi(h)$, because the only correlation between the two sites, is the correlation given by $\mathbf{\Sigma}$. So, to show that $\lim_{h\to\infty} \chi(h) = 0$, we need only know that $\lim_{h\to\infty} \pi(h) = 0$ where $\pi(h)$ is the probability that two sites separated by distance h are in the same partition. A proof of this is given in Appendix A.3.

Figure 1 shows $\chi(h)$ for different sample quantiles from simulated datasets where $\mathbf{s} \in [0, 1] \times [0, 1]$. These plots show how the partitioning allows for short range extremal dependence while reducing extremal dependence as the distance between sites increases.

3.3 Extension to space-time data

In a standard block-maxima approach, the temporal dependence is assumed to be negligible because blocks (e.g. yearly maxima) are taken to be large enough to assume independence. However, when using daily meausrements, the assumption of temporal independence is no longer appropriate. There are several places where temporal dependence could be included, including the residual $v(\mathbf{s})$. However, we choose to allow for temporal dependence in the $\mathbf{w}$, $z(\mathbf{s})$, and $\sigma(\mathbf{s})$ terms because these terms derive the tail behavior. Specifically, we incorporate an AR(1) time series. We first, transform the spatial knots from $\mathcal{D}$ to $\mathcal{R}^2$ as follows. Let

$$\mathbf{w}_k^* = \Phi_2^{-1}\left[\frac{(\mathbf{w}_k - \min(\mathbf{s}))}{\text{range}(\mathbf{s})}\right].$$

What's the best way to write the $\mathbf{w}$ transformation as 1-D? w_x and w_y or w_1 and w_2? where Φ_2 represents a bivariate standard normal density function. Then $\mathbf{w}_k^* \in \mathcal{R}^2$. We use a copula on the $\sigma_t^2(\mathbf{s})$ to ensure that the marginal distributions of $\sigma_t^2(\mathbf{s})$ are inverse gamma. Let

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

where Φ is a univariate standard normal density function, and IG is the distribution function for an IG(a, b) random variable. Then, when $t > 1$ the time series is modeled as

$$\begin{aligned}
\mathbf{w}_{1,k}^* &\sim N_2(0, 1)\\
z_{1k} &\sim N(0, \sigma_{1k}^2)\\
\sigma_{1k}^{2*} &\sim N(0, 1)\\
\mathbf{w}_{tk}^* | \mathbf{w}_{t-1,k}^* &\sim N_2\left[\phi_w \mathbf{w}_{t-1,k}^*, (1 - \phi_w^2)\right]\\
z_{tk} | z_{t-1,k} &\sim N\left[\phi_z z_{t-1,k}, \sigma_{tk}^2(1 - \phi_z^2)\right]\\
\sigma_{tk}^{2*} | \sigma_{t-1,k}^{2*} &\sim N\left[\phi_\sigma \sigma_{t-1,k}^{2*}, (1 - \phi_\sigma^2)\right]
\end{aligned}$$

where $|\phi_w| < 1$, $0 < \phi_z < 1$, and $|\phi_\sigma| < 1$. These are stationary time series models with marginal distributions

$$\begin{aligned}
\mathbf{w}_k^* &\sim N_2(0, 1)\\
z_k &\sim N(0, \sigma_k^2)\\
\sigma_k^{2*} &\sim N(0, 1).
\end{aligned}$$

For each day, the model is identical to the spatial-only model in Seciton 3.2.

3.4 Hierarchical model

Conditioned on $z_{tk}(\mathbf{s})$, $\sigma_{t,k}^2(\mathbf{s})$, and $P_{t,k}$, the marginal distributions are Gaussian and the joint distribution multivariate Gaussian. However, we do not fix the partitions, they are treated as unknown and updated in the MCMC. We model this with a Bayesian hierarchical model as follows. Let $\mathbf{w}_{t,1}, \dots, \mathbf{w}_{t,K}$ be a set of daily spatial knots in a spatial domain of interest, $\mathcal{D}$, and P_{tk} as defined in 3.2.

Then

$$\begin{aligned}
Y_t(\mathbf{s}) \mid z_t(\mathbf{s}), \sigma^2(\mathbf{s}), P_{tk}, \alpha, \beta, \Theta &= X_t(\mathbf{s})\beta + \alpha|z_t(\mathbf{s})| + \sigma_t(\mathbf{s})v_t(\mathbf{s}) && (11)\\
z_t(\mathbf{s}) &= z_{tk} \text{ if } \mathbf{s} \in P_{tk} && (12)\\
\sigma_t^2(\mathbf{s}) &= \sigma_{tk}^2 \text{ if } \mathbf{s} \in P_{tk} && (13)\\
\alpha &\sim N(0, 10) && (14)\\
v_t(\mathbf{s}) \mid \Theta &\sim \text{Matérn } (0, \Sigma) && (15)\\
z_{tk} \mid z_{t-1,k}, \sigma_{tk}^2 &\sim N(\phi_z z_{t-1,k}, \sigma_{tk}^2(1 - \phi_z^2)) && (16)\\
\sigma_{tk}^{2*} \mid \sigma_{t-1,k}^{2*} &\sim N(\phi_\sigma \sigma_{t-1,k}^{2*}, (1 - \phi_\sigma^2)) && (17)\\
\mathbf{w}_{tk}^* \mid \mathbf{w}_{t-1,k}^* &\sim N_2(\phi_w \mathbf{w}_{t-1,k}^*, (1 - \phi_w^2)) && (18)
\end{aligned}$$

where $\Theta = \{\rho, \nu, \gamma\}$; $k = \arg\min_\ell ||\mathbf{s} - \mathbf{w}_\ell||$; and Σ is a Matérn covariance matrix with variance one, spatial range ρ, smoothness ν, and the proportion of variance accounted for by the spatial variation is γ.

4 Computation

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

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

5 Simulation study

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

5.1 Design

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

1. Gaussian marginal, $K = 1$ knot
2. T marginal, $K = 1$ knot
3. T marginal, $K = 5$ knots
4. Skew-t marginal, $K = 1$ knots
5. Skew-t marginal, $K = 5$ knots
6. Max-stable.

In the first five designs, the $v_t(\mathbf{s})$ terms are generated using a Matérn covariance with smoothness parameter $\nu = 0.5$ and spatial range $\rho = 0.1$. For the covariance matrices in designs 1 – 5, the proportion of the variance accounted for by the spatial variation is $\gamma = 0.9$ while the proportion of the variance accounted for by the nugget effect is 0.1. In the first design, $\sigma^2 = 2$ is used for all days. For designs 2 – 4, $\sigma^2_{tk} \stackrel{iid}{\sim}$ IG$(3, 8)$ For designs 1 – 3, we set $\alpha = 0$. For designs four and five, $\alpha = 3$ was used, and the z_t are generated as described in (9). In the sixth design, we generate from a spatial max-stable distribution (Reich and Shaby, 2012) with parameters $\mu = 1, \sigma = 1, \xi = 0.2$ and 144 spatial knots on a regular lattice in the square $[1, 9] \times [1, 9]$. I know I need more here about the max-stable method In all six designs, the mean $\mu(\mathbf{s}) = 10$ is assumed to be constant across space.

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

1. Gaussian marginal, $K = 1$ knots
2. Skew-t marginal, $K = 1$ knots, $T = -\infty$
3. T marginal, $K = 1$ knots, $T = q(0.80)$
4. Skew-t marginal, $K = 5$ knots, $T = -\infty$
5. T marginal, $K = 5$ knots, $T = q(0.80)$

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

5.2 Cross validation

Models were compared using cross validation with 100 sites used as training sites and 30 sites witheld for testing. The model was fit using the training set, and predictions were generated at the testing site locations. Because one of the primary goals of this model is to predict extreme events, we use Brier scores to select the model that best fits the data (Gneiting and Raftery, 2007). The Brier score for predicting exceedance of 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 set value, y, has exceeded the threshold, c, and $P(c)$ is the predicted probability of exceeding c. We average the Brier scores over all test sites and days. For the Brier score, a lower score indicates a better fit.

5.3 Results

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

$$\text{BS}_{\text{rel}} = \frac{\text{BS}_{\text{method}}}{\text{BS}_{\text{Gaussian}}}.$$

For data that are relatively symmetric, the standard Gaussian method outperforms the other methods. However, when the data come from a skewed distribution (settings 3 – 6), the partitioned and thresholded methods outperform the Gaussian method. Additionally, for both of the skew-t settings ($K = 1$, $K = 5$), the skew-t method with the appropriate number of partitions performs the best. Finally, the simulation results suggest that for data from a skew-t distribution, thresholded methods provide some improvement over non-thresholded models for the extreme thresholds. With the exception of the Brier score for $c = q(0.995)$ in setting 6, the improvement over the Gaussian method is statistically significant using a Wilcoxon signed rank test ($\alpha = 0.05$).

6 Data analysis

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

$$\mu_t(\mathbf{s}) = \beta_0 + \beta_1 \cdot \text{CMAQ}_t(\mathbf{s}). \tag{19}$$

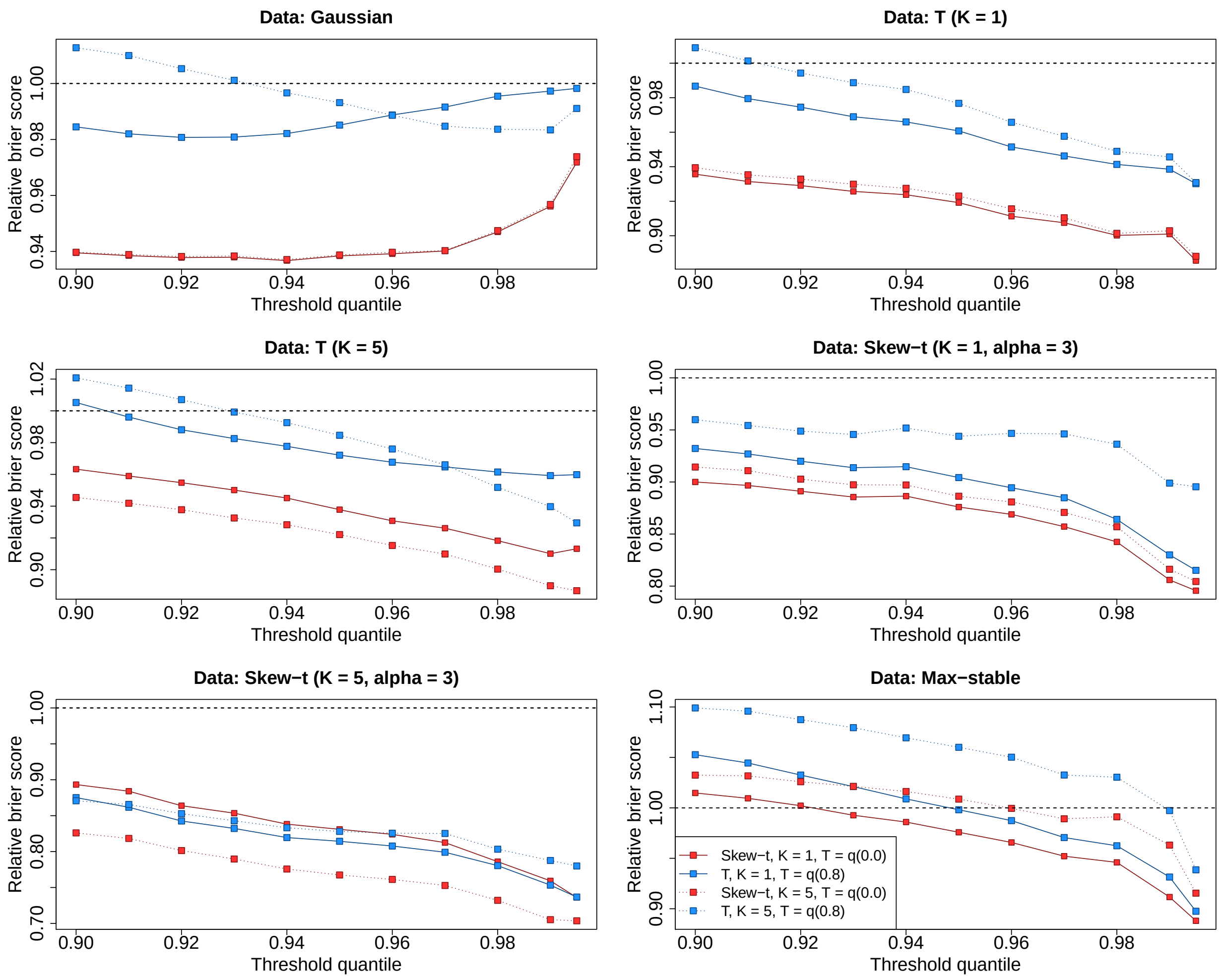

Figure 2: Brier scores relative to the Gaussian method for simulation study results. A ratio lower than 1 indicates that the method outperforms relative to the Gaussian method.

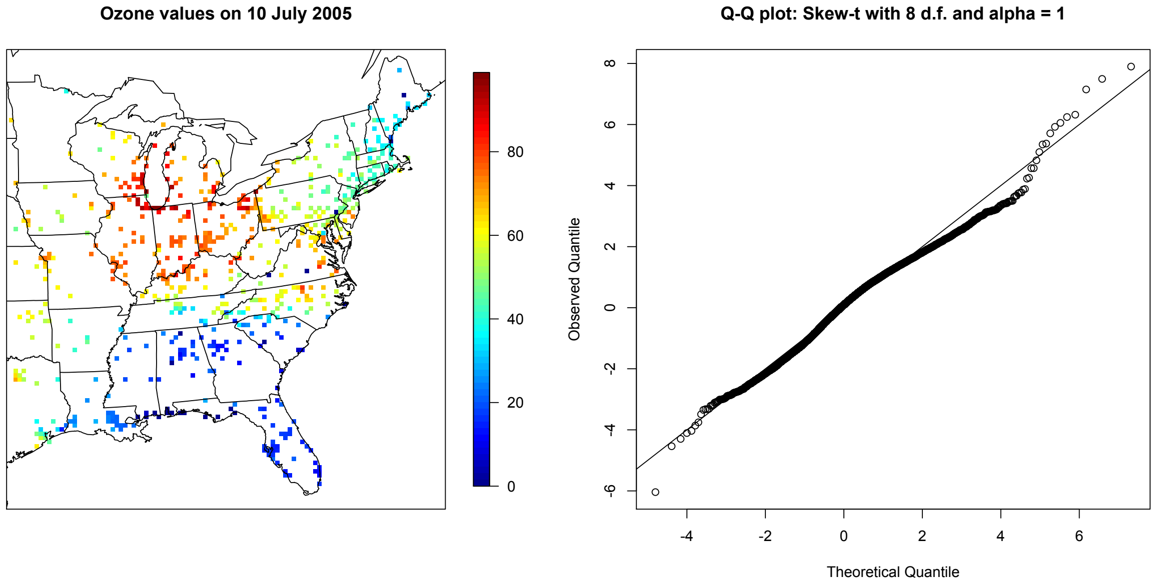

Figure 3: Ozone values on 10 July 2005 (left) Q-Q plot of the residuals (right)

The data from July 10 are shown in Figure 3 along with a Q-Q plot of the residuals compared to a skew-t distribution with 8 d.f. and $\alpha = 1$.

The $\widehat{\chi}(h)$ and $\widehat{\chi}(t)$ plots estimate the spatial and temporal dependence using sample data. For $\widehat{\chi}(h)$, we take the sample proportion of all pairs of sites within distance h of one another that are above a threshold T. I know I need another sentence or two here about $\widehat{\chi}(t)$ indicate that the residuals still demonstrate spatial and temporal dependence (see Figure 4)

6.1 Model comparisons

We fit the model using Gaussian and skew-t marginal distributions, $K = 1, 5, 10, 15$ partitions, with $Y(\mathbf{s})$ censored at $T = 0, 50, 75, 90$ ppb as described in Section **??**. We also include a max-stable analysis using the method by ???? All methods assume the mean function given in (19). For each model, Brier scores and quantile scores were were averaged over all sites and days to obtain a single quantile score for each dataset. At a particular threshold or quantile level, the model that fits the best is the one with the lowest score. Looking at the heatplot in Figure 5, we can see that the methods using no thresholding tend to outperform those methods that do use thresholding.

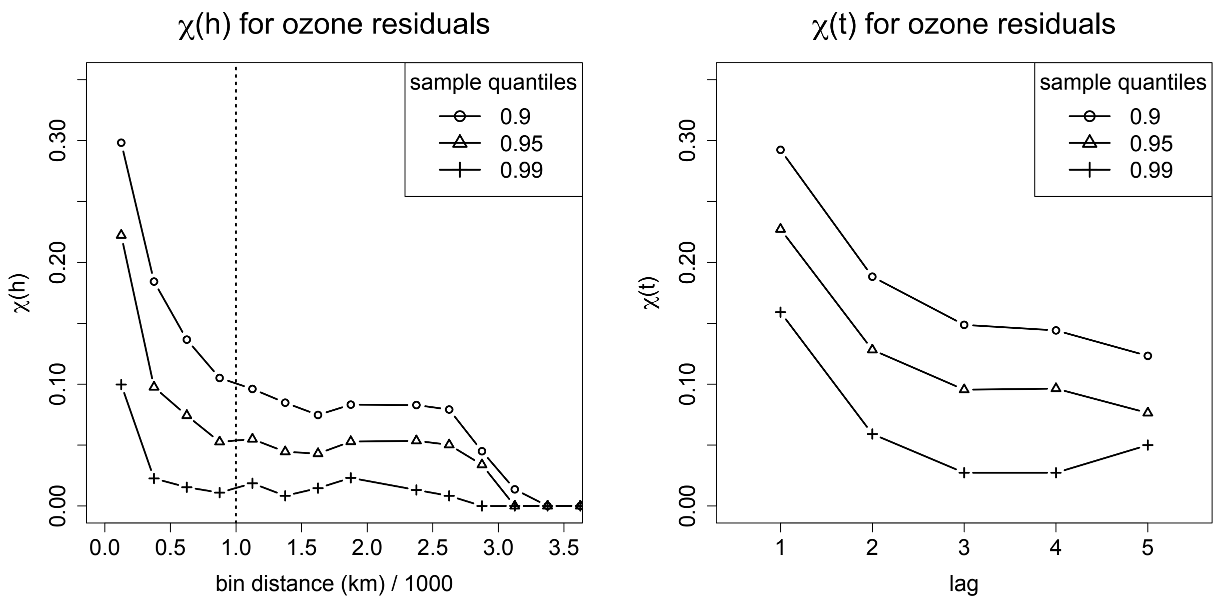

Figure 4: $\widehat{\chi}(h)$ plot for the residuals. The vertical line indicates the distance after which observations no longer demonstrate dependence (left) $\widehat{\chi}(t)$ plot for the residuals (right)

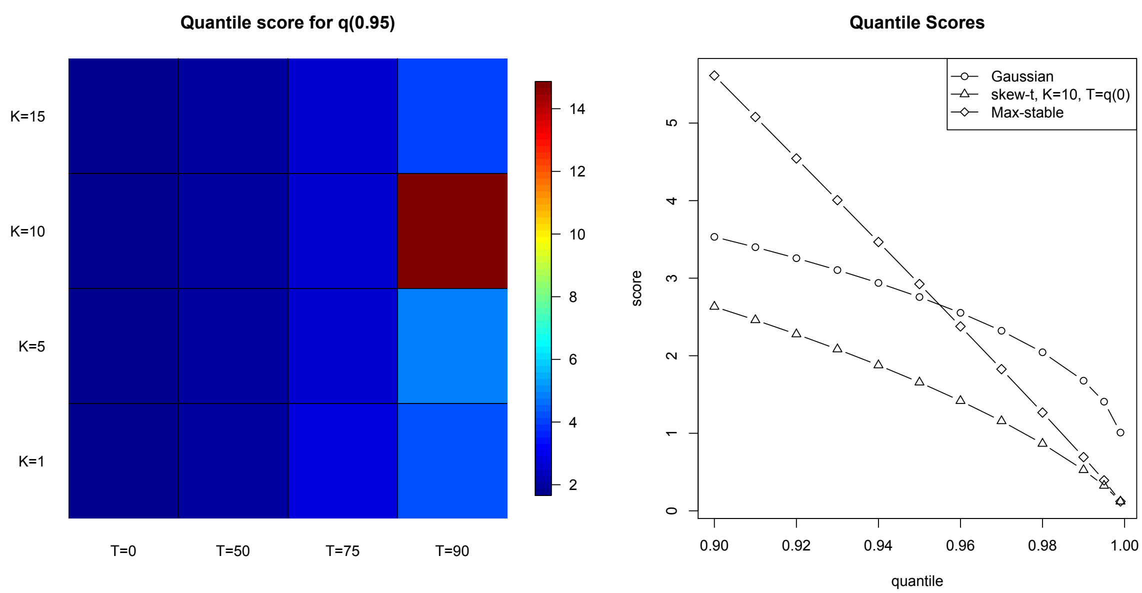

Figure 5: Heatplot of quantile scores for the 99th quantile (left) Quantile scores of Gaussian, skew-t (K=10, T=q(0)), and Max-stable (right)

6.2 Results

7 Conclusions

Acknowledgments

Appendix A.1: MCMC details

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

Spatial knot locations

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

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

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

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

where l is the likelihood given in (11), and $p(\cdot)$ is the prior either taken from the time series given in (3.3) or assumed to be uniform over $\mathcal{D}$. The candidate knots are accepted with probability $\min\{R, 1\}$.

Spatial random effects

If there is temporal dependence amongst the observations, then we update z_{tk} using a Metropolis Hastings update. First, we generate a candidate using a random walk Gaussian candidate distribution

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

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

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

The candidate is accepted with probability $\min\{R, 1\}$.

Variance terms

When there is more than one site in a partition, then we update σ_{tk}^2 using a Metropolis Hastings update. First, we generate from one of two candidate distributions. In the case of temporal dependence, we generate a candidate for ${\sigma_{tk}^*}^2$ using a random walk Gaussian candidate distribution

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

where ${\sigma_{tk}^*}^{2(r-1)}$ is the value at MCMC iteration $r-1$, and s is a tuning parameter. In the case where there is no temporal dependence, we generate a candidate for σ_{tk}^2 using an $\mathrm{IG}(a^*/s, b^*/s)$ candidate distribution in an independence 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 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 day t. The acceptance ratio is

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

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

Spatial covariance parameters

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

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

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

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

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

Appendix A.2: Posterior distributions

Conditional posterior of $z_{tk} \mid \ldots$

If knots are independent over days, then the conditional posterior distribution of $|z_{tk}|$ is conjugate. For simplicity, drop the subscript t, let $z_{tk}^* = |z_{tk}|$, and define

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

Let

$$\begin{aligned}R_1 &= \text{the vector of } R(\mathbf{s}) \text{ for } s \in P_l\\ R_2 &= \text{the vector of } R(\mathbf{s}) \text{ for } s \notin P_l\\ \Omega &= \Sigma^{-1}.\end{aligned}$$

Then

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

where

$$\begin{aligned}\mu_l &= \alpha(R_1^T \Omega_{11} + R_2^T \Omega_{21})\mathbf{1}\\ \Lambda_l &= \alpha^2 \mathbf{1}^T \Omega_{11} \mathbf{1} + \frac{1}{\sigma_l^2}.\end{aligned}$$

Then $Z_l^*|\ldots \sim N_{(0,\infty)}(\Lambda_l^{-1}\mu_l, \Lambda_l^{-1})$

Conditional posterior of $\beta \mid \ldots$

Let $\beta \sim \mathrm{N}_p(0, \Lambda_0)$ where Λ_0 is a precision matrix. Then

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

where

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

Conditional posterior of $\sigma^2 \mid \ldots$

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

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

where

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

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

Conditional posterior of $\alpha \mid \ldots$

Let $\alpha \sim N(0, \tau_\alpha)$ where τ_α is a precision term. Then

$$\begin{aligned}
\pi(\alpha \mid \ldots) &\propto \exp\left\{ -\frac{1}{2}\tau_\alpha \alpha^2 + \sum_{t=1}^{T} \frac{1}{2}[\mathbf{Y}_t - X_t\beta - \alpha|z_t|]^T \Omega [\mathbf{Y}_t - X_t\beta - \alpha|z_t|] \right\} \\
&\propto \exp\left\{ -\frac{1}{2}[\alpha^2(\tau_\alpha + \sum_{t=1}^{T} |z_t|^T \Omega |z_t|^T) - 2\alpha \sum_{t=1}^{T} [|z_t|^T \Omega(\mathbf{Y}_t - X_t\beta) \right\} \\
&\propto \exp\left\{ -\frac{1}{2}[\tau^*_\alpha \alpha^2 - 2\mu_\alpha] \right\}
\end{aligned}$$

where

$$\mu_\alpha = \sum_{t=1}^{T} |z_t|^T \Omega(\mathbf{Y}_t - X_t\beta)$$
$$\tau^*_\alpha = t_\alpha + \sum_{t=1}^{T} |z_t|^T \Omega |z_t|.$$

Then $\alpha \mid \ldots \sim N(\tau^{*-1}_\alpha \mu_\alpha, \tau^{*-1}_\alpha)$

Appendix A.3: Proof that $\lim_{h\to\infty} \pi(h) = 0$

Let $N(A)$ be the number of knots in A, the area between sites $\mathbf{s}_1$ and $\mathbf{s}_2$. Consider a spatial Poisson process with intensity $\mu(A)$. So,

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

Then for any finite k, $\lim_{h\to\infty} P[N(A) = k] = 0$ because $\lim_{h\to\infty} \mu(A) = \infty$. With each additional knot in A, the chance that $\mathbf{s}_1$ and $\mathbf{s}_2$ will be be in the same partition will decrease, because partition membership is defined by the closest knot to a site. Therefore, $\lim_{h\to\infty} \pi(h) = 0$.

Appendix A.4: Half-normal distribution

Let $u = |z|$ where $Z \sim N(\mu, \sigma^2)$. Specifically, we consider the case where $\mu = 0$. Then U follows a half-normal distribution which we denote as $U \sim HN(0, 1)$, and the density is given by

$$f_U(u) = \frac{\sqrt{2}}{\sqrt{\pi\sigma^2}} \exp\left(-\frac{u^2}{2\sigma^2}\right) I(u > 0) \tag{20}$$

When $\mu = 0$, the half-normal distribution is also equivalent to a $N_{(0,\infty)}(0, \sigma^2)$ where $N_{(a,b)}(\mu, \sigma^2)$ represents a normal distribution with mean μ and standard deviation σ that has been truncated below at a and above at b.

Appendix A.5: Multivariate skew-t distribution

Let $Z_{0t} \sim \text{SN}_d(0, \bar{\Omega}, \alpha)$ be a d-dimensional skew-normal random variable. Note, the relationship between $\bar{\Omega}$ and the Matérn covariance matrix given for (5) is given as follows

$$\begin{aligned} \alpha &= \frac{\delta}{\sqrt{1-\delta^2}} \\ D_\delta &= (diag(1-\delta))^{1/2} \\ \bar{\Omega} &= D_\delta(\Sigma + \lambda\lambda^T)D_\delta \\ \lambda &= D_\delta^{-1}\delta \end{aligned}$$

Let $\sigma_{tk}^2 \sim \chi_\nu^2/\nu$ where ν is the degrees of freedom. Let $Z_{tk} = \sigma_{tk}^{-1} Z_{0,tk}$. After marginalizing out the σ_{tk}^2 terms, Z follows a multivariate skew-t distribution, and the density is given by

$$f_z(z) = 2t_d(z; \bar{\Omega}, \nu)T\left(\alpha^T z\sqrt{\frac{\nu+d}{\nu+Q(z)}}; \nu+d\right)$$

where $Q(z) = z^T\bar{\Omega}^{-1}z$ and $T(\cdot;\rho)$ denotes the univariate Student's t distribution function on ρ degrees of freedom (**?**).

References

Azzalini, A. and Capitanio, A. (2003) Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **65**, 367–389.

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

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

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

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

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

Huser, R. (2013) *Statistical Modeling and Inference for Spatio-Temporal Extremes*. Ph.D. thesis.

Kim, H.-M., Mallick, B. K. and Holmes, C. C. (2005) Analyzing Nonstationary Spatial Data Using Piecewise Gaussian Processes. *Journal of the American Statistical Association*, **100**, 653–668.

Padoan, S. A. (2011) Multivariate extreme models based on underlying skew- and skew-normal distributions. *Journal of Multivariate Analysis*, **102**, 977–991.

Padoan, S. A., Ribatet, M. and Sisson, S. A. (2010) Likelihood-Based Inference for Max-Stable Processes. *Journal of the American Statistical Association*, **105**, 263–277.

Reich, B. J. and Shaby, B. A. (2012) A hierarchical max-stable spatial model for extreme precipitation. *The Annals of Applied Statistics*, **6**, 1430–1451.

Zhang, H. and El-Shaarawi, A. (2010) On spatial skewGaussian processes and applications. *Environmetrics*, **21**, 33–47.