

A new spatial model for points above a threshold

September 25, 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 it 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. In section 2, we describe the skew- t process with partitioning and thresholding. The computing is described in section 3. In section 4, we present a simulation study that examines the predictive capabilities of this model compared with a naive 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 5. The final section provides brief discussion and direction for future research.

2 Statistical model

Let $Y_t(\mathbf{s}) \in \mathcal{R}$ be the observed value at location \mathbf{s} and timepoint t . To avoid bias in estimating tail parameters, we model censored data

$$\tilde{Y}_t(\mathbf{s}) = \begin{cases} Y_t(\mathbf{s}) & Y_t(\mathbf{s}) > T \\ T & Y_t(\mathbf{s}) \leq T \end{cases} \quad (1)$$

where T is a pre-specified threshold. Then, assuming the full data are observations from a skew- t process, we update values censored below the threshold using standard Bayesian missing data methods as described in Section 3.

2.1 Skew- t process

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 additional parameters to account for non-symmetric data (Genton, 2004). One member of this family, is the skew- t distribution (Azzalini and Capitanio, 2003). We assume the data can be modeled using a skew- t distribution. Zhang and El-Shaarawi (2010) show that the skew- t can be expressed as a hierarchical model

$$Y_t(\mathbf{s}) = X_t(\mathbf{s})\beta + \alpha z_t + \sigma_t v_t(\mathbf{s}) \quad (2)$$

where $\alpha \in \mathcal{R}$ controls the skewness, $z_t \stackrel{\text{ind}}{\sim} N_{(0,\infty)}(0, \sigma_t^2)$ are a random effect from a half-normal distribution (see appendix A.4), $v_t(\mathbf{s})$ is a Gaussian process with mean zero, variance one, Matérn correlation, and $\sigma_t^2 \stackrel{\text{iid}}{\sim} \text{IG}(a, b)$. When marginalizing over the z_t and σ_t^2 terms,

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

where μ is the location, $\Sigma^* = \frac{b}{a}\Sigma$, Σ 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).

2.2 Extremal dependence

One common measure of extremal spatial dependence is the extremal coefficient which describes the pairwise dependence between spatial locations (Smith, 1990). Consider a spatial process $Y(\mathbf{s}) \in \mathcal{R}^n$ observed at locations $s \in \mathcal{D} \subset \mathcal{R}^2$. Then the bivariate extremal coefficient, $\theta(\mathbf{s}_i, \mathbf{s}_j) \in [1, 2]$, is defined as

$$\Pr(Y(\mathbf{s}_i) < c, Y(\mathbf{s}_j) < c) = \Pr(Y(\mathbf{s}_i) < c)^{\theta(\mathbf{s}_i, \mathbf{s}_j)}. \quad (3)$$

One way to characterize the dependence over the entire set of spatial locations is to calculate all of the pairwise extremal coefficients. Although this method provides information regarding the spatial structure of the observations, it does not fully characterize the joint spatial dependence.

Another measure of extremal dependence is the χ statistic. The χ statistic for the upper tail is given by

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

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

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

where $\mathbf{h} = \|\mathbf{s}_1 - \mathbf{s}_2\|$. If $\chi(\mathbf{h}) = 0$, then observations are asymptotically independent at distance \mathbf{h} . For Gaussian processes, $\chi(\mathbf{h}) = 0$ regardless of the distance, so they are not suitable for modeling spatially dependent extremes. However, for the process described in Section 2.1, $\chi(\mathbf{h}) > 0$ (Padoan, 2011).

2.3 Random daily partition

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_t and σ_t^2 terms. We handle this problem with a daily random partition similar to Huser and Davison (2014) that allows z_t and σ_t^2 to vary by site. The model then becomes

$$Y_t(\mathbf{s}) = X_t(\mathbf{s})\beta + \alpha z_t(\mathbf{s}) + \sigma_t(\mathbf{s})v_t(\mathbf{s}). \quad (4)$$

In this extension of (2), $z_t(\mathbf{s})$ and $\sigma_t(\mathbf{s})$ are allowed to vary by site. Their spatial variation is determined by the random partition model defined below. Consider a set of daily spatial knots $\mathbf{w}_{tk} \sim \text{Uniform}(\mathcal{D})$ that define a random daily partition P_{t1}, \dots, P_{tK} of the spatial domain of interest $\mathcal{D} \subset \mathcal{R}^2$ such that

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

So, for $\mathbf{s} \in P_{tk}$, let

$$z_t(\mathbf{s}) = z_{tk} \quad (5)$$

$$\sigma_t^2(\mathbf{s}) = \sigma_{tk}^2. \quad (6)$$

Then within each partition, $Y_t(\mathbf{s})$ follows the distribution given in (2). When incorporating the random daily partition, the χ statistic becomes

$$\chi(\mathbf{h}) = \lim_{c \rightarrow \infty} \pi(\mathbf{h})\chi(\mathbf{h}) \quad (7)$$

where $\pi(h)$ is the probability that two sites separated by distance h are in the same partition. Asymptotic dependence is eliminated as h increases, because

$$\lim_{h \rightarrow \infty} \chi(\mathbf{h}) = \lim_{h \rightarrow \infty} \pi(\mathbf{h})\chi(\mathbf{h}) = 0. \quad (8)$$

A proof of this can be found in Appendix A.3.

2.4 Accounting for temporal dependence

We account for temporal dependence through a AR(1) time series on the spatial knots and $z_t(\mathbf{s})$ terms. These are modeled as

$$\mathbf{w}_{t,k} | \mathbf{w}_{t-1,k} \sim N_2(\phi_w \mathbf{w}_{t-1,k}, (1 - \phi_w^2)) \quad (9)$$

$$z_{t,k} | z_{t-1,k} \sim N(\phi_z z_{t-1,k}, (1 - \phi_z^2)) \quad (10)$$

where $|\phi_w|, |\phi_z| < 1$.

2.5 Hierarchical model

Conditioned on $z_{tk}(\mathbf{s}) \stackrel{iid}{\sim} N(0, \sigma_{tk}^2)$, $\sigma_{tk}^2(\mathbf{s}) \sim IG(a, b)$, and P_{tk} , the marginal distributions are skew- t and the joint distribution within a partition is multivariate skew- t . 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 $\mathbf{w}_{t1}, \dots, \mathbf{w}_{tK}$ be a set of daily spatial knots in a spatial domain of interest, \mathcal{D} , so that

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

89 Because the spatial domain is bounded, we do a transformation on the knot locations to \mathcal{R}^2 . Let Φ_2 represent
 90 a bivariate standard normal density function, and let $\mathbf{w}_k = \min(\mathbf{s}) + \text{range}(\mathbf{s}) * \Phi_2^{-1}(\mathbf{w}_k^*)$. Then $\mathbf{w}_k^* \in \mathcal{R}^2$.
 91 Then

$$Y_t(\mathbf{s}) \mid z_t(\mathbf{s}), \sigma^2(\mathbf{s}), P_{tk}, \alpha, \beta, \Theta = X_t(\mathbf{s})\beta + \alpha|z_t(\mathbf{s})| + \sigma_t(\mathbf{s})v_t(\mathbf{s}) \quad (11)$$

$$z_t(\mathbf{s}) = z_{tk} \text{ if } \mathbf{s} \in P_{tk} \quad (12)$$

$$\sigma_t^2(\mathbf{s}) = \sigma_{tk}^2 \text{ if } \mathbf{s} \in P_{tk} \quad (13)$$

$$\alpha \sim N(0, 10) \quad (14)$$

$$v_t(\mathbf{s}) \mid \Theta \sim \text{Matérn}(0, \Sigma) \quad (15)$$

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

$$\sigma_{tk}^2 \stackrel{iid}{\sim} IG(\alpha, \beta) \quad (17)$$

$$\mathbf{w}_{tk}^* \mid \mathbf{w}_{t-1,k}^* \sim N_2(\phi_w \mathbf{w}_{t-1,k}^*, (1 - \phi_z^2)) \quad (18)$$

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

94 3 Computation

95 The MCMC for this model is fairly straightforward. First, we impute values below the threshold. Then, we
 96 update Θ using random walk MH or Gibbs sampling when appropriate. Finally, we make spatial predictions
 97 using conditional multivariate normal results and the fact that the distribution of $Y_t(\mathbf{s}) \mid \Theta, z_{tl}$ is the usual
 98 multivariate normal distribution with a Matérn spatial covariance structure.

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

108 4 Simulation study

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

111 4.1 Design

112 For all simulation designs, we generate data from the model presented in Section 2.3 using $n_s = 130$ sites
 113 and $n_t = 50$ independent days. The sites are generated Uniform($[0, 10] \times [0, 10]$). We generate data from 6
 114 different simulation designs:

- 115 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, $\sigma_{tk}^2 \stackrel{iid}{\sim} \text{IG}(3, 8)$. 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 (5). 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 $[1, 9] \times [1, 9]$. In all six designs, the mean ($\mu(\mathbf{s})$) 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 $\sigma_t^2(\mathbf{s}) \sim \text{IG}(0.1, 0.1)$.

4.2 Cross validation

Models were compared using cross validation with 100 sites used as training sites and 30 sites withheld 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 quantify use Brier scores and quantile 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 . The quantile score for the τ th quantile is $2\{I[y < \hat{q}(\tau)] - \tau\}(\hat{q}(\tau) - y)$ where y is a AQS test set value and $\hat{q}(\tau)$ is the estimated τ th quantile. For both the Brier score and the quantile score, a lower score indicates a better fit. These scores were averaged over all sites and days to obtain a single quantile score for each dataset.

4.3 Results

When comparing quantile scores, in almost all of the data settings, the biggest gains are seen when using thresholded methods. In all data settings except for the max-stable setting, method 3 performs the best. However, in most of the data settings, method 5 also performs reasonably well. In fact, for setting 5, paired t -tests suggest that the difference in performance is not significant for any quantile score. Furthermore, paired t -tests results indicate that the difference in quantile scores between methods 3 and 5 is only statistically significant after $q(0.97)$ for setting 4 and after $q(0.93)$ for setting 1.

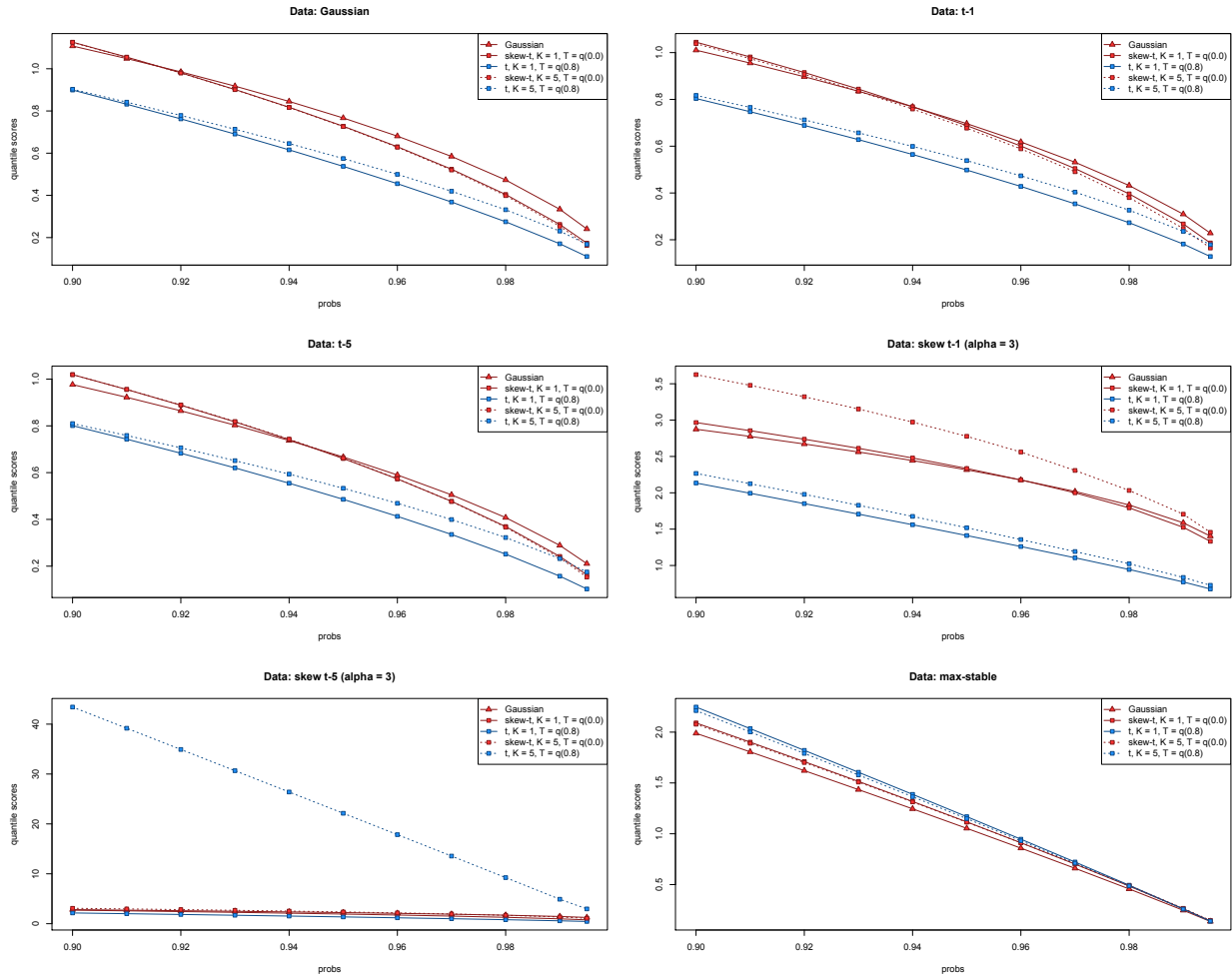


Figure 1: Quantile score plots for simulation study results

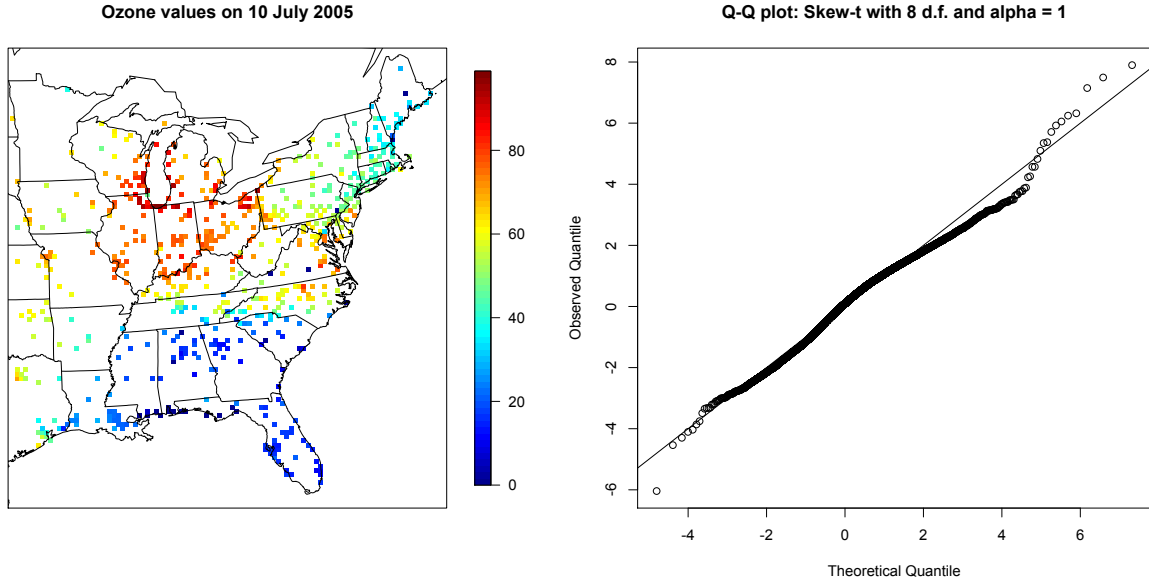


Figure 2: Ozone values on 10 July 2005 (left) Q-Q plot of the residuals (right)

5 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. For each site, we also have covariate information containing the estimated ozone from the Community Multi-scale Air Quality (CMAQ) modeling system. Initially, we fit a linear regression assuming that the mean function can be expressed as

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

The data from 10 July are shown in Figure 2 along with a Q-Q plot of the residuals compared to a skew- t distribution with 8 d.f. and $\alpha = 1$.

The $\chi(h)$ and $\chi(t)$ plots indicate that the residuals still demonstrate spatial and temporal dependence (see Figure 3)

5.1 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 2.1. We also include a max-stable analysis using the method by [1]. All methods assume the mean function given in (19). For each model, Brier scores and quantile scores 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 4, we can see that the methods using no thresholding tend to outperform those methods that do use thresholding.

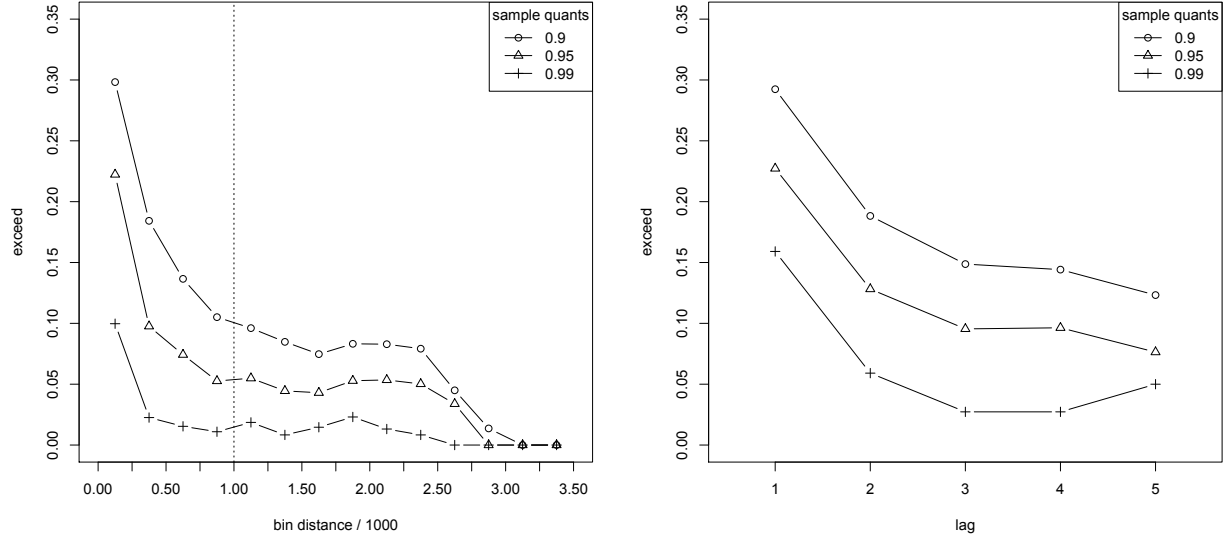


Figure 3: $\chi(h)$ plot for the residuals. The vertical line indicates distance after which observations no longer demonstrate dependence (left) $\chi(t)$ plot for the residuals (right)

5.2 Results

6 Conclusions

Acknowledgments

Appendix A.1: MCMC details

The MCMC sampling for the model 2.5 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, $\mathbf{w}_1, \dots, \mathbf{w}_K$, using a Metropolis Hastings block update. Because the spatial domain is bounded, we generate candidate knots using the transformed knots $\mathbf{w}_1^*, \dots, \mathbf{w}_K^*$ (see section 2.5) and a random walk bivariate Gaussian candidate distribution

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

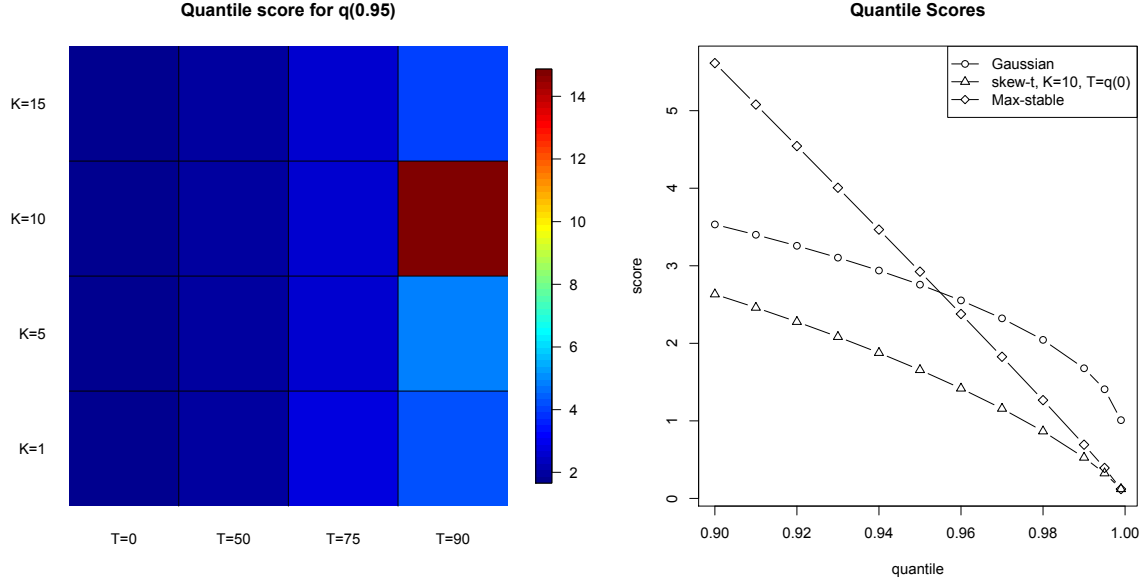


Figure 4: Heatplot of quantile scores for the 99th quantile (left) Quantile scores of Gaussian, skew- t ($K=10$, $T=q(0)$), and Max-stable (right)

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

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

191 where l is the likelihood given in (11). The candidate knots are accepted with probability $\min\{R, 1\}$.

192 Spatial random effects

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

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

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

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

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

197 Spatial covariance parameters

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

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

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

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

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

203 **Appendix A.2: Posterior distributions**

204 **Conditional posterior of $z_{tk} \mid \dots$**

205 If knots are independent over days, then the conditional posterior distribution of z_{tk} is conjugate. For
 206 simplicity, drop the subscript t and define

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

207 Let

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

208 Then

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

209 where

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

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

211 **Conditional posterior of $\beta \mid \dots$**

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

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

213 where

$$\mu_\beta = \sum_{t=1}^T [X_t^T \Omega(\mathbf{Y}_t - \alpha z_t)]$$

$$\Lambda_\beta = \Lambda_0 + \sum_{t=1}^T X_t^T \Omega X_t.$$

214 **Conditional posterior of $\sigma^2 \mid \dots$**

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

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

217 where

$$\alpha^* = \alpha_0 + \frac{1}{2} + \frac{n}{2}$$

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

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

219 **Conditional posterior of $\alpha \mid \dots$**

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

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

221 where

$$\mu_\alpha = \sum_{t=1}^T z_t^T \Omega (\mathbf{Y}_t - X_t \beta)$$

$$\tau_\alpha^* = t_\alpha + \sum_{t=1}^T z_t^T \Omega z_t.$$

222 Then $\alpha \mid \dots \sim N(\tau_\alpha^{*-1} \mu_\alpha, \tau_\alpha^{*-1})$

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

Consider two spatial locations, $\mathbf{s}_1, \mathbf{s}_2 \in \mathcal{R}^2$, and any two knots, $w_1, w_2 \in \mathcal{R}^2$. Let \bar{S} be the line segment connecting \mathbf{s}_1 and \mathbf{s}_2 . Let $h = \|\mathbf{s}_1 - \mathbf{s}_2\|$ be the distance between sites \mathbf{s}_1 and \mathbf{s}_2 . Let \bar{W} be the perpendicular bisector of the line connecting w_1 and w_2 . Then \bar{W} defines a partition dividing \mathcal{R}^2 into two half-planes. Let $\lim_{h \rightarrow \infty}$ be a limit in the sense that both \mathbf{s}_1 and \mathbf{s}_2 are moving at some rate away from their midpoint. Provided that \bar{W} and \bar{S} are not parallel to one another, we are guaranteed that as $h \rightarrow \infty$, \mathbf{s}_1 and \mathbf{s}_2 will be in different partitions.

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

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

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 .

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, École Polytechnique Fédérale de Lausanne.
- Huser, R. and Davison, A. C. (2014) Space-time modelling of extreme events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **76**, 439–461.

- 254 Padoan, S. A. (2011) Multivariate extreme models based on underlying skew- and skew-normal distribu-
255 tions. *Journal of Multivariate Analysis*, **102**, 977–991.
- 256 Padoan, S. A., Ribatet, M. and Sisson, S. A. (2010) Likelihood-Based Inference for Max-Stable Processes.
257 *Journal of the American Statistical Association*, **105**, 263–277.
- 258 Reich, B. J. and Shaby, B. A. (2012) A hierarchical max-stable spatial model for extreme precipitation. *The*
259 *Annals of Applied Statistics*, **6**, 1430–1451.
- 260 Smith, R. L. (1990) Max-stable processes and spatial extremes.
- 261 Zhang, H. and El-Shaarawi, A. (2010) On spatial skewGaussian processes and applications. *Environmetrics*,
262 33–47.