BIOMETRICS 000, 000–000 DOI: 000
000 0000

Web-based Supplementary Materials for A Space-time Skew-t Model for Threshold Exceedances by Morris, Reich, Thibaud, and Cooley

Samuel A Morris[1,*], Brian J Reich[1], Emeric Thibaud[2], and Daniel Cooley[2]

[1]Department of Statistics, North Carolina State University, Raleigh, North Carolina, U.S.A.

[2]Department of Statistics, Colorado State University, Fort Collins, Colorado, U.S.A.

**email:* samorris@ncsu.edu

This paper has been submitted for consideration for publication in *Biometrics*

Web Appendix A. MCMC details

The MCMC sampling for the model ~~??~~ in Section 4 is done using R (http://www.r-project.org). Whenever possible, we select conjugate priors (see ~~Appendix ??~~**??**); however, for some of the parameters, no conjugate prior distributions exist. For these parameters, we use a random walk Metropolis-Hastings update step. In each Metropolis-Hastings update, we tune the algorithm during the burn-in period 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 ??~~Section 3.3) and a random walk bivariate Gaussian candidate distribution

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

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. Let $\mathbf{Y}_t = [Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n)]$ be the vector of observed responses at each site for day t. After candidates have been generated for all K knots, the acceptance ratio is

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

where l is the likelihood given in ~~(??)~~(17), and $p(\cdot)$ is the prior either taken from the time series ~~given in (??~~(see Section 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 no temporal dependence amongst the observations, we use a Gibbs update for z_{tk}, and the posterior distribution is given in ~~??~~**??**. If there is temporal dependence amongst the observations, then we update z_{tk} using a Metropolis-Hastings update. Because this model uses $|z_{tk}|$, we generate candidate random effects using the z^*_{tk} (see ~~Section ??~~Section 3.3) and a random walk Gaussian candidate distribution

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

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

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

where $p[\cdot]$ is the prior taken from the time series given in ~~Section ??~~Section 3.3. 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^2_{tk} using a Metropolis-Hastings update. First, we generate a candidate for σ^2_{tk} using an IG$(a^*/s, b^*/s)$ candidate distribution in an independence Metropolis-Hastings update where $a^* = (n_{tk}+1)/2+a$, $\cancel{b^* = [Y_{tk}^T\Sigma_{tk}^{-1}Y_{tk} + z^2_{tk}]/2 + b}$$b^* = [\mathbf{Y}_{tk}^\top\Sigma_{tk}^{-1}\mathbf{Y}$ n_{tk} is the number of sites in partition k on day t, and $\cancel{Y_{tk}}$ $\mathbf{Y}_{tk}$ and Σ^{-1}_{tk} are the observations and precision matrix for partition k on day t. The acceptance ratio is

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

where $p[\cdot]$ is the prior either taken from the time series given in ~~Section ??~~ Section 3.3 or assumed to be 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

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

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

Web Appendix B. 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 ~~$\tilde{z}_{tk} = |z_{tk}|$,~~ $\tilde{z}_k = |z_k|$, $\tilde{\mathbf{z}}_{k^c}$ be the vector of $[|z(\mathbf{s}_1)|, \ldots, |z(\mathbf{s}_n)|]$ for $\mathbf{s} \notin P_k$, $\mathbf{X} = [\mathbf{X}(\mathbf{s}_1), \ldots, \mathbf{X}(\mathbf{s}_n)]^\top$, let $\mathbf{Y}_k$ and $\mathbf{X}_k$ be the observations and covariate measurements for $\mathbf{s} \in P_k$, and let $\mathbf{Y}_{k^c}$ and ~~define~~ $\mathbf{X}_{k^c}$ be the observations and covariate measurements for $\mathbf{s} \notin P_k$ and define

$$\underline{R()}\mathbf{R} = \begin{cases} \mathbf{Y}_k - \mathbf{X}_k \boldsymbol{\beta} & \mathbf{s} \in P_k \\ \mathbf{Y}_{k^c} - \mathbf{X}_{k^c} \boldsymbol{\beta} - \lambda \tilde{\mathbf{z}}_{k^c} & \mathbf{s} \notin P_k \end{cases}$$

Let

$$\underline{R}\mathbf{R}_1 = \text{the vector of } \underline{R()}\mathbf{R} \text{ for } \underline{s}\mathbf{s} \in P_{\underline{l}k}$$

$$\underline{R}\mathbf{R}_2 = \text{the vector of } \underline{R()}\mathbf{R} \text{ for } \underline{s}\mathbf{s} \notin P_{\underline{l}k}$$

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

Then

$$\pi(z_{lk}|\ldots) \propto \exp\left\{-\frac{1}{2}\left[\begin{pmatrix}\mathbf{R}_1 - \lambda\tilde{z}_k\mathbf{1}\\ \mathbf{R}_2\end{pmatrix}^{T\top}\begin{pmatrix}\Omega_{11} & \Omega_{12}\\ \Omega_{21} & \Omega_{22}\end{pmatrix}\begin{pmatrix}\mathbf{R}_1 - \lambda\tilde{z}_k\mathbf{1}\\ \mathbf{R}_2\end{pmatrix} + \frac{\tilde{z}_l^2}{\sigma_l^2}\frac{\tilde{z}_k^2}{\sigma_k^2}\right]\right\} I(z_{lk} > 0)$$

$$\propto \exp\left\{-\frac{1}{2}\left[\Lambda_{lk}\tilde{z}_{lk}^{\;2} - 2\mu_{lk}\tilde{z}_{lk}\right]\right\}$$

where

$$\mu_{lk} = \lambda(R\mathbf{R}_1{}^{T\top}\Omega_{11} + R\mathbf{R}_2{}^{T\top}\Omega_{21})\mathbf{1}$$

$$\Lambda_{lk} = \lambda^2\mathbf{1}^{T\top}\Omega_{11}\mathbf{1} + \frac{1}{\sigma_l^2}\frac{1}{\sigma_k^2}.$$

Then ~~$\tilde{Z}_l|\ldots \sim N_{(0,\infty)}(\Lambda_l^{-1}\mu_l, \Lambda_l^{-1})$~~ $\tilde{z}_k|\ldots \sim N_{(0,\infty)}(\Lambda_k^{-1}\mu_k, \Lambda_k^{-1})$

Conditional posterior of ~~$\beta \mid \ldots$~~ $\boldsymbol{\beta}, \lambda \mid \ldots$

~~Let $\beta \sim \mathrm{N}_p(0, \Lambda_0)$~~ For models that do not include a skewness parameter, we update $\boldsymbol{\beta}$ as follows. Let $\boldsymbol{\beta} \sim \mathrm{N}_p(0, \Lambda_0)$ where Λ_0 is a precision matrix. Then

$$\pi(\beta\boldsymbol{\beta} \mid \ldots) \propto \exp\left\{-\frac{1}{2}\beta^T\boldsymbol{\beta}^\top\Lambda_0\beta\boldsymbol{\beta} - \frac{1}{2}\sum{}_{t=1}^{T}{}_{t=1}^{n_t}[\mathbf{Y}_t - X\mathbf{X}_t\beta - \lambda|z_t|\boldsymbol{\beta}]^{T\top}\Omega[\mathbf{Y}_t - X\mathbf{X}_t\beta - \lambda|z_t|\boldsymbol{\beta}]\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left[\beta^T\boldsymbol{\beta}^\top\Lambda_\beta\beta\boldsymbol{\beta} - 2\sum{}_{t=1}^{T}\beta^T X_{t=1}^{n_t}(\boldsymbol{\beta}^\top\mathbf{X}_t{}^{T\top}\Omega(\mathbf{Y}_t - \lambda|z_t|)\right]\right\}$$

$$\propto \mathrm{N}(\Lambda_\beta^{-1}\mu_\beta, \Lambda_\beta^{-1})$$

where

$$\mu_\beta = \sum{}_{t=1}^{T}X_{t=1}^{n_t}\mathbf{X}_t{}^{T\top}\Omega(\mathbf{Y}_t - \lambda|z_t|)$$

$$\Lambda_\beta = \Lambda_0 + \sum{}_{t=1}^{T}X_{t=1}^{n_t}\mathbf{X}_t{}^{T\top}\Omega X\mathbf{X}_t.$$

For models that do include a skewness parameter, a simple augmentation of the covariate matrix $\mathbf{X}$ and parameter vector $\boldsymbol{\beta}$ allows for a block update of both $\boldsymbol{\beta}$ and λ. Let $\mathbf{X}_t^* = [\mathbf{X}_t, |\mathbf{z}_t|]$ where $\mathbf{z}_t = [z(\mathbf{s}_1), \ldots, z(\mathbf{s}_n)]^\top$ and let $\boldsymbol{\beta}^* = (\beta_1, \ldots, \beta_p, \lambda)^\top$. So to incorporate the $N(0, \sigma_\lambda^2)$ prior on λ, let $\boldsymbol{\beta}^* \sim \mathrm{N}_{p+1}(0, \Lambda_0^*)$ where

$$\Lambda_0^* = \begin{pmatrix} \Lambda_0 & 0 \\ 0 & \sigma_\lambda^{-2} \end{pmatrix}.$$

Then the update for both $\boldsymbol{\beta}$ and λ is done using the conjugate prior given above with $\mathbf{X}_t = \mathbf{X}_t^*$ and $\boldsymbol{\beta} = \boldsymbol{\beta}^*$

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

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

where

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

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

~~*Conditional posterior of* $\lambda \mid \ldots$~~

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

$$\begin{aligned}
\underline{\lambda_1} &= \underline{\begin{cases} +1 & \text{w.p.}0.5 \\ -1 & \text{w.p.}0.5 \end{cases}} \\
\underline{\lambda_2^2} &\ \underline{\sim IG(\alpha_\lambda, \beta_\lambda).}
\end{aligned}$$

~~Then~~

$$\pi(\lambda_2 \mid \ldots) \propto \lambda_2^{2(-\alpha_\lambda - 1)} \exp\left\{-\frac{\beta_\lambda}{\lambda_2^2}\right\} \prod_{t=1}^{T} \prod_{k=1}^{K}$$

$$\lambda_2 \exp\left\{-\frac{z_{tk}^2}{2\lambda_2^2\sigma_{tk})^2}\right\}$$

~~Then~~ $\cancel{\lambda_2 \mid \ldots \sim IG\left(\alpha_\lambda + kt, \beta_\lambda + \frac{z^2}{2\sigma_{tk}^2}\right)}$

Web Appendix C. Proof that $\lim_{h\to\infty} \pi(h) = 0$

Let c be the midpoint of $\mathbf{s}_1$ and $\mathbf{s}_2$. Define A as the circle centered at c with radius $h/2$ where $h = ||\mathbf{s}_1 - \mathbf{s}_2||$ is the distance between sites $\mathbf{s}_1$ and $\mathbf{s}_2$. Consider a homogeneous spatial Poisson process ~~with intensity~~ $\cancel{\mu}$~~. Define~~ over A with intensity given by

$$\mu_{PP}(A) = \lambda_{PP}|A| = \lambda_{PP}\pi\left(\frac{h}{2}\right)^2 = \lambda_{PPA}^* h^2.$$

Consider a partition of A ~~as the circle with center and radius~~ $\cancel{h/2}$~~. Then~~ into four regions, B_1, B_2, R_1, R_2 as seen in Web Figure **??**.

[Figure 1 about here.]

Let N_i be the number of knots in B_i and $L_i = l$ if $\mathbf{s}_i \in P_l$ for $i = 1, 2$. Then

$$P(L_1 \neq L_2) \geqslant P(N_1 > 0, N_2 > 0) \tag{1}$$

since knots in both B_1 and B_2 is sufficient, but not necessary, to ensure that $\mathbf{s}_1$ and $\mathbf{s}_2$ are in different ~~partitions almost surely if two or more points are in~~ $\cancel{A}$~~. Let~~ $\cancel{N(A)}$ ~~be the number of points in~~ $\cancel{A}$~~, and let~~

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

~~Then~~

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

partition sets. By definition of a Poission process, N_1 and N_2 are independent and thus $P(N_1 > 0, N_2 > 0) = P(N$

and the intensity measure over B_1 is given by

$$
\begin{aligned}
\mu_{PP}(B_1) &= \lambda_{PP}|B_1| = \lambda_{PP} \frac{h^2}{4} \left(\frac{2\pi}{3} - \frac{\sqrt{3}}{2} \right) \\
&= \lambda^*_{PPB1} h^2. \qquad (2)
\end{aligned}
$$

So,

$$
P(L_1 \neq L_2) >= P(N_1 > 0)^2 = [1 - P(N_1 = 0)]^2 = [1 - \exp\left(-\lambda^*_{PPB1} h^2\right)]^2 \qquad (3)
$$

which goes to ~~one as~~ $\cancel{h \to \infty}$ 1 as h goes to infinity.

Web Appendix D. Skew-t distribution

Univariate skew-t distribution

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

$$
f_{\text{EST}}(y) = 2 f_T(z; \nu) F_T \left[\alpha z \sqrt{\frac{\nu + 1}{\nu + z^2}}; \nu + 1 \right] \qquad (4)
$$

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

Multivariate skew-t distribution

If $\mathbf{Z} \sim \mathrm{ST}_d(0, \bar{\boldsymbol{\Omega}}, \boldsymbol{\alpha}, \eta)$ is a d-dimensional skew-t distribution, and $\mathbf{Y} = \xi + \boldsymbol{\omega}\mathbf{Z}$, where $\boldsymbol{\omega} = \mathrm{diag}(\omega_1, \ldots, \omega_d)$, then the density of Y at y is

$$f_y(\mathbf{y}) = det(\boldsymbol{\omega})^{-1} f_z(\mathbf{z}) \tag{5}$$

where

$$f_z(\mathbf{z}) = 2t_d(\mathbf{z}; \bar{\boldsymbol{\Omega}}, \eta) T \left[\boldsymbol{\alpha}^{\cancel{T}\top} \mathbf{z} \sqrt{\frac{\eta + d}{\nu + Q(\mathbf{z})}}; \eta + d \right] \tag{6}$$

$$\mathbf{z} = \boldsymbol{\omega}^{-1}(\mathbf{y} - \xi) \tag{7}$$

where $t_d(\mathbf{z}; \bar{\boldsymbol{\Omega}}, \eta)$ is a d-dimensional Student's t-distribution with scale matrix $\bar{\boldsymbol{\Omega}}$ and degrees of freedom η, ~~$Q(z) = \mathbf{z}^T \bar{\Omega}^{-1} \mathbf{z}$~~ $Q(z) = \mathbf{z}^\top \bar{\Omega}^{-1} \mathbf{z}$ and $T(\cdot; \eta)$ denotes the univariate Student's t distribution function with η degrees of freedom (**?**).

Extremal dependence

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

$$\chi(h) = \bar{F}_{\mathrm{EST}} \left\{ \frac{[x_1^{1/\eta} - \varrho(h)]\sqrt{\eta + 1}}{\sqrt{1 - \varrho(h)^2}}; 0, 1, \alpha_1, \tau_1, \eta + 1 \right\} + \bar{F}_{\mathrm{EST}} \left\{ \frac{[x_2^{1/\eta} - \varrho(h)]\sqrt{\eta + 1}}{\sqrt{1 - \varrho(h)^2}}; 0, 1, \alpha_2, \tau_2, \eta + 1 \right\}, \tag{8}$$

where $\bar{F}_{\mathrm{EST}}$ is the univariate survival extended skew-t function with zero location and unit scale, $\varrho(h) = \mathrm{cor}[y(\mathbf{s}), y(\mathbf{t})]$, $\alpha_j = \alpha_i\sqrt{1 - \varrho^2}$, $\tau_j = \sqrt{\eta + 1}(\alpha_j + \alpha_i\varrho)$, and $x_j = F_T(\bar{\alpha}_i\sqrt{\eta + 1}; 0, 1, \eta)/F_T(\bar{\alpha}_j\sqrt{\eta + 1};$ with $j = 1, 2$ and $i = 2, 1$ and where $\bar{\alpha}_j = (\alpha_j + \alpha_i\varrho)/\sqrt{1 + \alpha_i^2[1 - \varrho(h)^2]}$.

Proof that $\lim_{h\to\infty} \chi(h) > 0$

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

$$\lim_{h\to\infty} \chi(h) = \bar{F}_{\mathrm{EST}} \left\{ \sqrt{\eta + 1}; 0, 1, \alpha_1, \tau_1, \eta + 1 \right\} + \bar{F}_{\mathrm{EST}} \left\{ \sqrt{\eta + 1}; 0, 1, \alpha_2, \tau_2, \eta + 1 \right\}. \tag{9}$$

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

Web Appendix E. Comparisons with other parameterizations

Various forms of multivariate skew-normal and skew-t distributions have been proposed in the literature. In this section, we make a connection between our parameterization in (1) of the main text and another popular version. **?** and **?** define a skew-normal process as

$$\tilde{X}(\mathbf{s}) = \tilde{\lambda}|z| + (1 - \tilde{\lambda}^2)^{1/2} v(\mathbf{s}) \tag{10}$$

where $\tilde{\lambda} \in (-1, 1)$, $z \sim N(0, 1)$, and $v(\mathbf{s})$ is a Gaussian process with mean zero, variance one, and spatial correlation function ρ. To extend this to the skew-t distribution, **?** take $\tilde{Y}(\mathbf{s}) = W\tilde{X}(\mathbf{s})$ where $W^{-2} \sim \text{Gamma}(a/2, a/2)$. Returning to the proposed parameterization (with $\boldsymbol{\beta} = 0$), let $W^{-2} = \frac{b}{a}\sigma^{-2} \sim \text{Gamma}(a/2, a/2)$ so that (1) in the manuscript becomes

$$Y(\mathbf{s}) = W\left[\lambda\left(\frac{b}{a}\right)^{1/2}|z| + \left(\frac{b}{a}\right)^{1/2} v(\mathbf{s})\right]. \tag{11}$$

Clearly setting $b = a(1 - \tilde{\lambda}^2) > 0$, and $\lambda = \tilde{\lambda}/(1 - \tilde{\lambda}^2)^{1/2} \in (-\infty, \infty)$ resolves the difference in parameterizations. We note that our parameterization has three parameters (a, b, λ) compared to the two parameters of the alternative parameterization $(a, \tilde{\lambda})$. Since we have assumed that both $v(\mathbf{s})$ and z have unit scale, the additional b parameter in our parameterization is required to control the precision.

Web Appendix F. Temporal dependence

It is very challenging to derive an analytical expression the temporal extremal dependence at a single site $\mathbf{s}$. However, using simulated data, we have evidence to suggest that the model does exhibits temporal extremal dependence. To demonstrate that our model maintains temporal extremal dependence, we generate lag-m observations for $m = 1, 3, 5, 10$ from our model setting $\phi_w = \phi_z = \phi_\sigma = \varphi$, for $\varphi = 0, 0.02, 0.04, \ldots, 1$. To estimate the lag-$m$ chi-statistic $\chi(m)$ we first estimate the lag-m F-madogram $\nu_F(m)$ (?) using $\hat{\nu}_F(m) = \frac{1}{2n}\sum_{i=1}^{n} |\hat{F}(y_0) - \hat{F}(y_m)|$ where $\hat{F}(\cdot)$ represents an empirical CDF and y_m is the lag-m observation. The F-madogram is related to the χ statistic as follows

$$\chi = 2 - \frac{1 + 2\nu_F}{1 - 2\nu_F}. \tag{12}$$

Web Figure **??** suggests that the extremal dependence increases as $\phi \to 1$, and that the extremal dependence decreases as m increases.

[Figure 2 about here.]

Web Appendix G. Brier scores for ozone prediction

Because typical ozone concentration varies throughout the US, we have included Brier scores by site for two model fits (Gaussian and Skew-t, $K = 5$, $T = 50$) in Web Figure **??**. As we can see in these plots, both models seem to perform similarly across the US with the poorest performance in California. Other methods have similar Brier score maps to these.

[Figure 3 about here.]

Web Appendix H. Simulation study pairwise difference results

The following tables show the methods that have significantly different Brier scores when using a Wilcoxon-Nemenyi-McDonald-Thompson test. In each column, different letters signify that the methods have significantly different Brier scores. For example, there is significant evidence to suggest that method 1 and method 4 have different Brier scores at $q(0.90)$, whereas there is not significant evidence to suggest that method 1 and method 2 have different Brier scores at $q(0.90)$. In each table group A represents the group with the lowest Brier scores. Groups are significant with a familywise error rate of $\alpha = 0.05$.

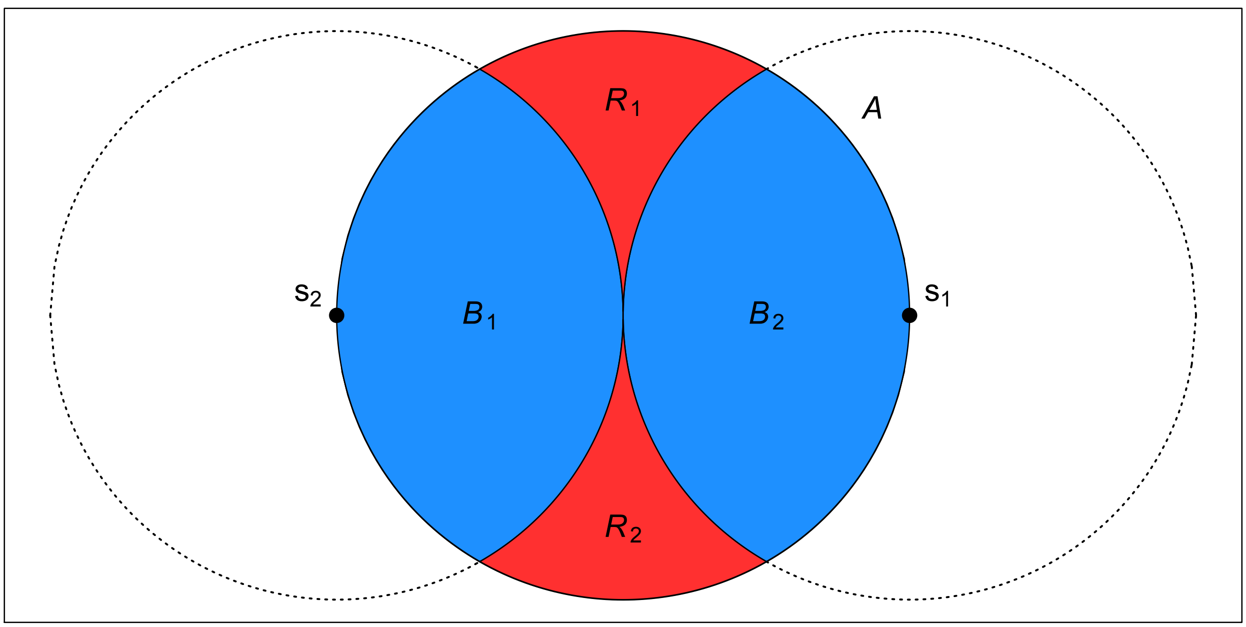
[Table 1 about here.]

[Table 2 about here.]

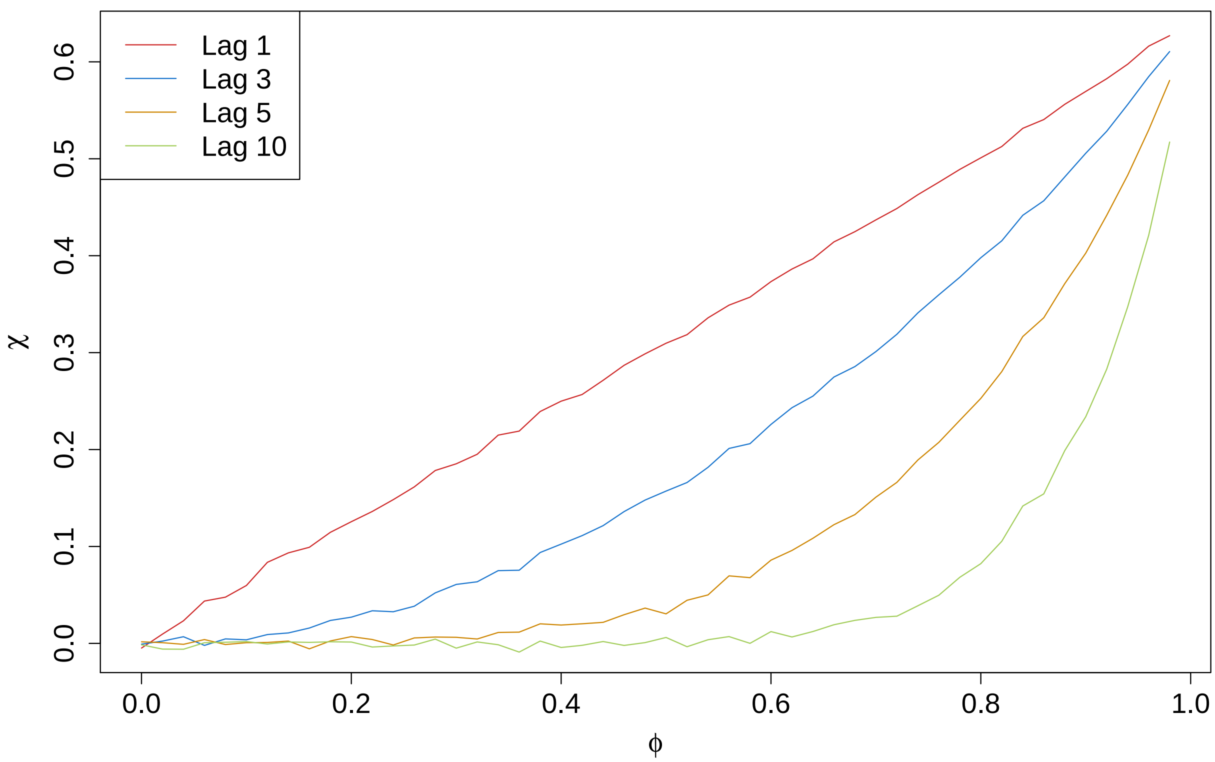
[Table 3 about here.]

[Table 4 about here.]

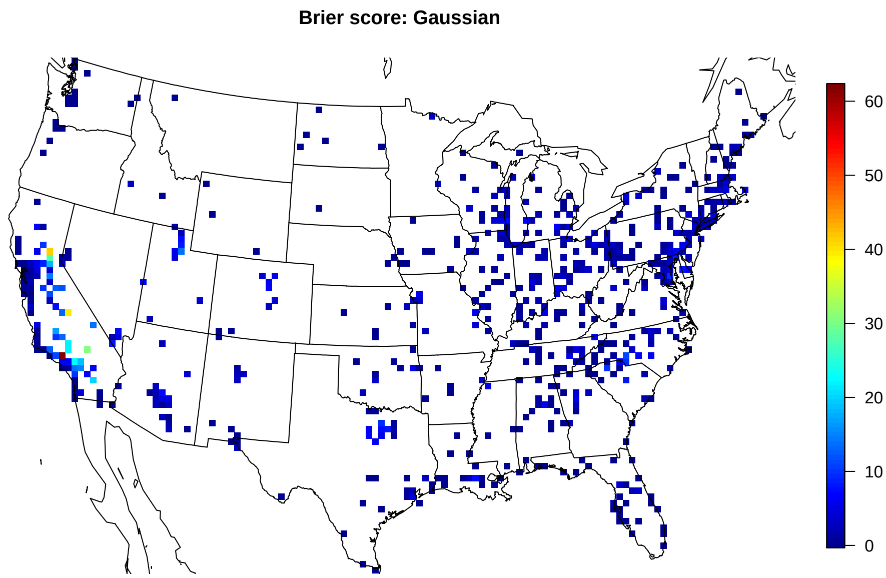
[Table 5 about here.]

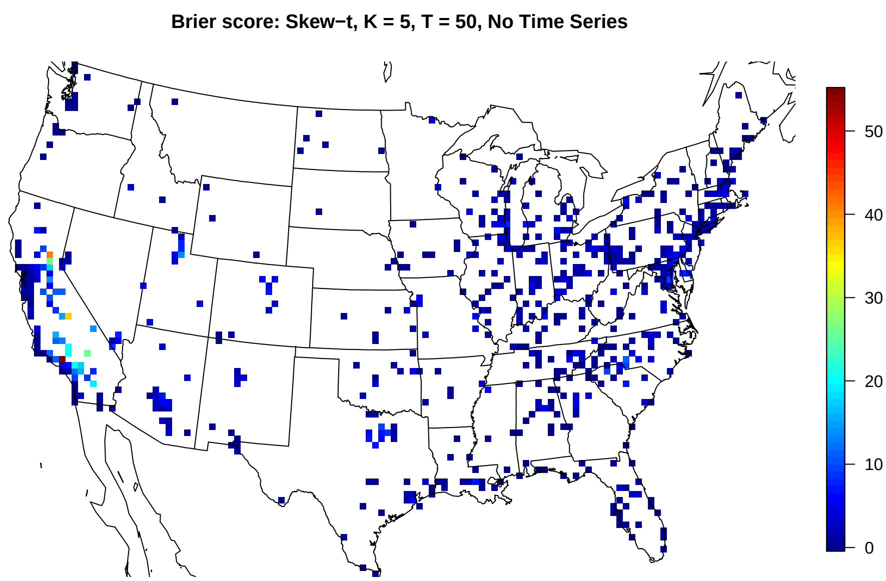


Web Figure 1. Illustration of the partition of A.



Web Figure 2. Simulated lag-m χ for varying levels of φ.





Web Figure 3. Map of Brier scores for Gaussian (top) vs Skew-t, $K = 5$, $T = 50$ (bottom).

Web Table 1

Setting 1 – Gaussian marginal, $K = 1$ knot

	$q(0.90)$	$q(0.95)$	$q(0.98)$	$q(0.99)$
Method 1	A	A	A	A ~~B~~
Method 2	A	A	A	A
Method 3	B	B	C	A ~~B~~
Method 4	A	A	A B	A ~~B~~
Method 5	B	B	B C	A ~~B~~
Method 6	C	C	D	~~C~~ B

Web Table 2

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

	$q(0.90)$	$q(0.95)$	$q(0.98)$	$q(0.99)$
Method 1	B ~~C~~	B	B ~~C~~	B
Method 2	A	A	A	A
Method 3	A B ~~C~~	A B	A B	A B
Method 4	A B	A B	A B	A B
Method 5	C ~~D~~	C	C	~~B~~ C
Method 6	D ~~E~~	D	D	C

Web Table 3

Setting 3 – Skew-t marginal, $K = 5$ knots

	$q(0.90)$	$q(0.95)$	$q(0.98)$	$q(0.99)$
Method 1	~~B~~ C	C	B	B
Method 2	~~B~~ C	C	B	B
Method 3	~~A~~ B	B	A ~~B~~	A ~~B~~
Method 4	A	A	A	A
Method 5	A	A	A	A
Method 6	~~C~~ D	D	C	C

Web Table 4

Setting 4 – Max-stable

	$q(0.90)$	$q(0.95)$	$q(0.98)$	$q(0.99)$
Method 1	A B	B	B	C
Method 2	B	B ~~C~~	B	B C
Method 3	C D	C	B	B
Method 4	D	D	C	C
Method 5	C	B C	B	B C
Method 6	A	A	A	A

Web Table 5

Setting 5 – ~~Transformation below $T = q(0.80)$~~ Brown Resnick

	$q(0.90)$	$q(0.95)$	$q(0.98)$	$q(0.99)$
Method 1	~~C~~ D	~~B~~ C	C	~~C~~ C
Method 2	~~B~~ D	~~B~~ C	~~B~~ C	~~A~~ ~~B~~ C
Method 3	A B	A	A B	~~A~~ B
Method 4	~~B~~ C	B	B	B ~~C~~
Method 5	A ~~B~~	A ~~B~~	A ~~B~~ ~~C~~	A B ~~C~~
Method 6	B C ~~D~~	A ~~C~~	A	~~D~~ A ~~D~~