A new spatial model for points above a threshold

September 8, 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 small $\mathbf{h}$, 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.

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

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

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

2.1 Skew-t process

We assume the data can be modeled as skew-t. Zhang and El-Shaarawi (2010) show that the skew-t can be written as the hierarchical model

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

where $\alpha \in \mathcal{R}$ controls the skewness, $z_t \stackrel{ind}{\sim} N_{(0,\infty)}(0, \sigma_t^2)$ are a random effect from a half-normal distribution (see appendix A.3), $v_t(\mathbf{s})$ is a Gaussian process with mean zero, variance one, and Matérn correlation, and $\sigma_t^2 \stackrel{iid}{\sim} \text{IG}(a, b)$. 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, controlled by both the skewness parameter α and the degrees of freedom $2a$, and its joint distribution is a multivariate skew-t distribution.

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

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

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 popular measure of extremal dependence is the χ statistic. The χ coefficient 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

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

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 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 skew-t distribution 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

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

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

$$z_t(\mathbf{s}) = z_{tk} \tag{5}$$
$$\sigma_t^2(\mathbf{s}) = \sigma_{tk}^2. \tag{6}$$

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

$$\chi(\mathbf{h}) = \lim_{c \to \infty} \pi(\mathbf{h})\chi(\mathbf{h}) \tag{7}$$

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

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

A proof of this is given in Appendix A.2.

2.4 Hierarchical model

Conditioned on $z_{tk}(\mathbf{s}) \overset{iid}{\sim} N(0, \sigma_{tk}^2)$, $\sigma_{tk}^2(\mathbf{s}) \sim IG(a, b)$, 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. Thus, standard geostatistical methods can be used to fit the model, and predictions can be made by Kriging at unobserved locations. We model this with a Bayesian hierarchical model as follows. Let $w_{t1}, \dots, w_{tK}$ be a set of daily spatial knots in a spatial domain of interest, $\mathcal{D}$, so that

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

Then

$$Y_t(\mathbf{s}) \mid \Theta, z_t(\mathbf{s}) = X_t(\mathbf{s})\beta + \alpha z_t(\mathbf{s}) + \sigma_t(\mathbf{s}) v_t(\mathbf{s}) \tag{9}$$
$$z_t(\mathbf{s}) = z_{tk} \text{ if } \mathbf{s} \in P_{tk} \tag{10}$$
$$\sigma_t^2(\mathbf{s}) = \sigma_{tk}^2 \text{ if } \mathbf{s} \in P_{tk} z_{tk} \mid \sigma_{tk}^2 \sim N_{(0,\infty)}(0, \sigma_{tk}^2) \tag{11}$$
$$\sigma_{tk}^2 \overset{iid}{\sim} IG(\alpha, \beta) \tag{12}$$
$$v_t(\mathbf{s}) \mid \Theta \sim \text{Matérn } (0, \Sigma) \tag{13}$$
$$\alpha \sim N(0, 10) \tag{14}$$
$$w_{tk} \sim Unif(\mathcal{D}) \tag{15}$$

where $\Theta = \{w_{t1}, \dots, w_{tK}, \beta, \sigma_t, \alpha, \lambda, \rho, \nu\}$; $k = \arg\min_\ell ||\mathbf{s} - \mathbf{w}_\ell||$; and Σ is a Matérn covariance matrix with variance one, spatial range ρ, smoothness ν.

3 Computation

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

3.1 Imputation

We can use Gibbs sampling to update $Y_t(\mathbf{s})$ for observations that are below T, the thresholded value. Given Θ, $Y_t(\mathbf{s})$ has truncated normal full conditional with these parameter values. So we sample $Y_t(\mathbf{s}) \sim N_{(-\infty, T)}(\mu(\mathbf{s}), \Sigma)$

3.2 Parameter updates

To update Θ given the current value of the complete data $\mathbf{Y}_1, \ldots, \mathbf{Y}_T$, we use a standard Gibbs updates for all parameters except for the knot locations which are done using a Metropolis update. See Appendix A.1 for details regarding Gibbs sampling.

3.3 Spatial prediction

Given $\mathbf{Y}_t$ the usual Kriging equations give the predictive distribution for $Y_t(\mathbf{s}^*)$ at prediction location $(\mathbf{s}^*)$

4 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.

4.1 Design

For all simulation designs, we generate data from the model presented in Section 2.3 using $n_s = 130$ 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 (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. skew-t marginal, $K = 1$ knots, $T = q(0.90)$
4. skew-t marginal, $K = 5$ knots, $T = -\infty$
5. skew-t marginal, $K = 5$ knots, $T = q(0.9)$

where $q(0.9)$ is the 90th sample quantile of the data. The design matrix $\mathbf{X}$ includes on the intercept with a prior of $\beta \sim \mathrm{N}(0, 10)$. The spatial covariance parameters have priors $\log(\nu) \sim \mathrm{N}(-1.2, 1)$, $\gamma \sim \mathrm{U}(0, 1)$, $\log(\rho) \sim \mathrm{N}(-2, 1)$. The skewness parameter has prior $\alpha \sim \mathrm{N}(0, 2)$. The residual variance terms have priors $\sigma^2_t(\mathbf{s}) \sim \mathrm{IG}(0.1, 0.1)$.

4.2 Cross validation

Models were compared using cross validation with 100 sites used as training sites and 30 sites 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 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 τth quantile is $2\{I[y < \widehat{q}(\tau)] - \tau\}(\widehat{q}(\tau) - y)$ where y is a AQS test set value and $\widehat{q}(\tau)$ is the estimated τth 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

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. 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 ???? All methods assume the location can be expressed as

$$\mu_t(\mathbf{s}) = \beta_0 + \beta_1 \cdot \mathrm{CMAQ}_t(\mathbf{s}). \tag{16}$$

To explore the extremal dependence both over space and time, we plot $\chi(\mathbf{s})$ and $\chi(t)$. 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.

5.1 Results

6 Conclusions

Acknowledgments

Appendix A.1: Posterior distributions

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

For simplicity, drop the subscript t and define

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

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

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

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

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

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

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

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

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

where

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

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

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

Appendix A.2: Proof that $\lim_{h \to \infty} \pi(h) = 0$

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 $\overline{S}$ 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 $\overline{W}$ be the perpendicular bisector of the line connecting w_1 and w_2. Then $\overline{W}$ defines a partition dividing $\mathcal{R}^2$ into two half-planes. Let $\lim_{h \to \infty}$ 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 $\overline{W}$ and $\overline{S}$ are not parallel to one another, we are guaranteed that as $h \to \infty$, $\mathbf{s}_1$ and $\mathbf{s}_2$ will be in different partitions.

Appendix A.3: 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

$$f_U(u) = \frac{\sqrt{2}}{\sqrt{\pi\sigma^2}} \exp\left(-\frac{u^2}{2\sigma^2}\right) I(u > 0) \tag{17}$$

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

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.

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.

Smith, R. L. (1990) Max-stable processes and spatial extremes.

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