A supplement to "Statistical modeling of extreme value behavior in North American tree-ring density series"

Elizabeth Mannshardt[1,2]; Peter Craigmile[3,4]; Martin Tingley[5]

[1] Department of Statistics, North Carolina State University

[2] Email: mannshardt@stat.ncsu.edu

[3] Department of Statistics, The Ohio State University

[4] School of Mathematics and Statistics, University of Glasgow

[5] Department of Earth and Planetary Sciences, Harvard University

1 Other extreme value analyses

In the article a block maxima modeling approach is employed. There are two other approaches that were considered:

1. **The running-maxima approach.** Instead of calculating one maxima for each block, the running maxima of block length B are calculated over the entire time period. While this approach generates more maximum values to study, it also introduces dependence between the maxima, and therefore necessitates a more involved model.

2. **The points-over-threshold approach**, which is also known as the peaks over thresholds approach, is based on examining all the observations over a given threshold. Under certain conditions, the distribution of excess values follow the *generalized pareto distribution* (GPD) [Pickands, 1975]. There are differing methods for the choice of threshold and the bias-variance trade-off must be considered. The threshold must be high enough to ensure that exceedances reflect the tail of the distribution, otherwise the resulting

inferences will be biased towards non-extreme behavior, yet not so high that sparse observations cause an inflated variance. Drawbacks to the peaks-over-threshold method include the dependence of parameter estimates on the value of the chosen threshold, and the fact that, in practice, there are often very few observations to analyze. The lack of observations is a problem for a points-over-threshold tree ring density analysis.

It is important to stress that the block maxima, as well as the approaches described above, are appropriate for *marginal* extreme value analysis. In the paleoclimate context, these models are appropriate for describing and characterizing the extreme values at a single spatial location, which for this analysis is considered conditionally independent of other sites. Spatial and temporal dependence will enter into the analysis by modeling the *parameters* of the GEV distribution as varying in both space and time. As the climate system displays both temporal and spatial dependencies which carry over to the occurrences and values of extremes, an analysis strategy that models the distribution of extremes jointly, at all sites, may be more appropriate. However, the modeling of multivariate extreme values, such as the joint occurrence of extreme values at a number of sites, is an active field of statistical research, and is outside the scope of this paper; see Coles [2001, Chapter 8] for a general review. A limiting factor is that most realistic spatio-temporal multivariate extreme value distributions cannot be fit exactly in the Bayesian hierarchical modeling context [see, e.g., Ribatet et al., 2012, for further details, and an example of an approximation]. Examples of modeling joint dependence can be found in, for example, Ledford and Tawn [1996] and Heffernan and Tawn [2004].

2 Supplemental exploratory data analysis

Maps of the ML estimates of the $\beta(\boldsymbol{s}_i)$ and the associated standard errors indicate how the temporal slopes in the model for the GEV location parameters vary as a function of space,

for both the maxima model (Fig. 1, upper panel) and the minima model (Fig. 1, lower panel). As a rough guide to statistical significance, a 90% confidence interval is calculated for the true slope parameter at each location and we indicate if the interval contains only positive values (Fig. 1, pink shading), only negative values (Fig. 1, blue shading), or if the interval contains zero (Fig. 1, white).

Figures 2 and 3, respectively, displays spatial maps of the ML estimates of the shape and scale parameter for the decadal maxima and decadal minima GEV models.

3 A discussion of the choice of spatial model

It is useful to briefly discuss how the spatial domain of the data set motivated the modeling decisions described in the main text. Since the proxy data are analyzed on a 5° by 5° grid there are a number of different ways to define the spatial models (see, e.g., Cressie [1993] and Banerjee et al. [2004] for reviews) for the intercept and slope:

1. A **lattice (or areal) model** for the spatially-varying coefficients is defined by modeling the conditional distribution of each grid cell given a neighborhood of adjacent grid cells. Lattice models are only appropriate for gridded or area-referenced data, so while applicable here, are not appropriate more generally for the analysis of heterogeneously distributed proxy observations. This approach can be computationally efficient due to sparse covariance matrices, but has the disadvantage that prediction at unobserved grid cells can be non-trivial, unless carefully planned for (e.g., suppose we wish to predict the parameters of the extreme value distribution in the center of the map shown in Fig. 1**(a)**). Banerjee et al. [2004, p.163] lists other computational and theoretical problems with using the models in practice.

2. A **geostatistical model** defines the covariance between the spatially varying coeffi-

cients in terms of functions of the locations and unknown statistical parameters. In general, the geostatistical approach allows for more flexibility, is easier to adapt, and is appropriate for dealing with point-referenced and/or Non-Gaussian data. Computation for these models can be more involved, but prediction at unobserved locations is relatively straightforward.

The second approach is used here, making sure to only define the geostatistical model at the centroids of each grid cell. While geostatistical models are defined over a continuum of spatial locations, it make no sense for this analysis to make inference at non-centroid locations, given that the observations are formed as averages over the gridboxes. Geostatistical models have the advantage that they are a natural model if the observations are not gridded data sources – as is common for paleoclimate series.

3.1 The chordal distance

The chordal distance, $||\cdot||$, used in the paper is defined by

$$\begin{aligned} ||\boldsymbol{s}_i - \boldsymbol{s}_{i'}||^2 &= 6371^2 \left[(\cos(\boldsymbol{s}_{i,1}\zeta)\cos(\boldsymbol{s}_{i,2}\zeta) - \cos(\boldsymbol{s}_{i',1}\zeta)\cos(\boldsymbol{s}_{i',2}\zeta))^2 + \right. \\ &\qquad \left. (\cos(\boldsymbol{s}_{i,1}\zeta)\sin(\boldsymbol{s}_{i,2}\zeta) - \cos(\boldsymbol{s}_{i',1}\zeta)\sin(\boldsymbol{s}_{i',2}\zeta))^2 + (\sin(\boldsymbol{s}_{1,i}\zeta) - \sin(\boldsymbol{s}_{1',i}\zeta))^2\right], \end{aligned}$$

with $\zeta = \pi/180$.

4 Prior specifications

Independent priors are assumed for the scale and shape parameters, $1/\sigma \sim \text{Gamma}(r_\sigma, s_\sigma); \xi \sim N(m_\xi, v_\xi^2)$, where $\text{Gamma}(r, s)$ denotes a Gamma distribution with rate r and shape s, and $N(m, v^2)$ denotes a normal distribution with mean m and variance v^2. The six parameters which define the spatially varying coefficient models for the intercepts

and slopes are given priors that are mutually independent, and independent of σ and ξ, with the following parametric forms:

$$\begin{aligned}
\boldsymbol{\lambda}_\alpha &\sim N_3(\boldsymbol{m}_{\alpha,\lambda}, v^2_{\alpha,\lambda}\boldsymbol{I}); \\
\phi_\alpha &\sim \text{Gamma}(r_{\alpha,\phi}, s_{\alpha,\phi}); \\
1/\tau^2_\alpha &\sim \text{Gamma}(r_{\alpha,\tau^2}, s_{\alpha,\tau^2}); \\
1/\omega^2_\alpha &\sim \text{Gamma}(r_{\alpha,\omega^2}, s_{\alpha,\omega^2}); \\
\boldsymbol{\lambda}_\beta &\sim N_3(\boldsymbol{m}_{\beta,\lambda}, v^2_{\beta,\lambda}\boldsymbol{I}); \\
\phi_\beta &\sim \text{Gamma}(r_{\beta,\phi}, s_{\beta,\phi}); \\
1/\tau^2_\beta &\sim \text{Gamma}(r_{\beta,\tau^2}, s_{\beta,\tau^2}). \\
1/\omega^2_\alpha &\sim \text{Gamma}(r_{\alpha,\omega^2}, s_{\alpha,\omega^2});
\end{aligned}$$

These priors distribution in turn depend on hyperparameters, which will be treated as fixed constants, and are presented in Table 1.

With the exception of the spatial correlation parameters ϕ_α and ϕ_β, the hyperparameters correspond to fairly vague prior distributions. Initial results using vague priors for the spatial correlation parameters indicated that there is limited information in the data to update the prior belief about these parameters. To allow for useful inference on other model parameters, scientifically informative priors are used for the spatial correlation parameters. In the absence of other information, the e-folding length scale (that is, the separation at which the correlation falls to e^{-1}) of annual mean surface temperature anomalies was used as a rough guide in specifying these priors. Figure 3 of Hansen and Lebedeff [1987] suggests an e-folding length scale of about 1750km in the mid- to high- latitudes, while Mann and Park [1993] and Tingley and Huybers [2010] estimate length scales of 1500km and 1800km, respectively, using global data. Informed by these studies, identical Gamma priors are chosen for ϕ_α and ϕ_β, with means of 1000km and, to allow for a fair amount of uncertainty, 5th and 95th

Table 1: A list of the hyperparameters used in the Bayesian hierarchical models for both the decadal maxima and decadal minima.

Parameter	Hyperparameters
σ	$r_\sigma = 0.01 \quad ; \quad s_\sigma = 0.01$
ξ	$m_\xi = 0 \quad ; \quad v_\xi^2 = 1$
$\boldsymbol{\lambda}_\alpha$	$\boldsymbol{m}_{\alpha,\lambda} = (0,0)^T \quad ; \quad v_{\alpha,\lambda}^2 = 100$
ϕ_α	$r_{\alpha,\phi} = 10 \quad ; \quad s_{\alpha,\phi} = 0.01$
τ_α^2	$r_{\alpha,\tau^2} = 0.01 \quad ; \quad s_{\alpha,\tau^2} = 0.01$
ω_α^2	$r_{\alpha,\omega^2} = 0.01 \quad ; \quad s_{\alpha,\omega^2} = 0.01$
$\boldsymbol{\lambda}_\beta$	$\boldsymbol{m}_{\beta,\lambda} = (0,0)^T \quad ; \quad v_{\beta,\lambda}^2 = 100$
ϕ_β	$r_{\beta,\phi} = 10 \quad ; \quad s_{\beta,\phi} = 0.01$
τ_β^2	$r_{\beta,\tau^2} = 0.01 \quad ; \quad s_{\beta,\tau^2} = 0.01$
ω_β^2	$r_{\beta,\omega^2} = 0.01 \quad ; \quad s_{\beta,\omega^2} = 0.01$

percentiles of 542.5km and 1570.5km. Fig. 4**(a)** shows the prior density for ϕ_β (equivalently for ϕ_α). A prior mean is chosen smaller than the values suggested in Hansen and Lebedeff [1987], Mann and Park [1993] and Tingley and Huybers [2010] as our aim is to model spatial structure in the limited spatial domain considered here. In addition, there is no simple way to link the characteristic length scale of instrumental temperature anomalies to that for the distributional parameters governing extremal behavior in climate-sensitive tree ring series, so the studies of the instrumental record are used as no more than a rough guide. A sensitivity study indicated the results are robust to moderately different priors for ϕ_α and ϕ_β.

5 The full conditional distributions used in the Markov chain Monte Carlo algorithm

Here are the steps required to carry out the MCMC algorithm for the results presented in the main article. Throughout, let $\boldsymbol{V}(\tau_\alpha^2, \phi_\alpha, \omega_\alpha^2) = \tau_\alpha^2 \boldsymbol{R}(\phi_\alpha) + \omega_\alpha^2 \boldsymbol{I}$.

Update σ and ξ:

$$\pi(\sigma, \xi | \boldsymbol{y}, \boldsymbol{\theta} \setminus \{\sigma, \xi\}) \quad \propto \quad \left[\prod_{i=1}^{I} \prod_{j=1}^{N(\boldsymbol{s}_i)} f(M_j(\boldsymbol{s}_i) | \eta_j(\boldsymbol{s}_i), \sigma, \xi) \right] \pi(\sigma)\pi(\xi).$$

(In the above equation, $\boldsymbol{\theta} \setminus \{\sigma, \xi\}$ denotes the parameter vector $\boldsymbol{\theta}$ excluding the parameters σ, and ξ). A Metropolis-Hastings symmetric random walk update is used on the $\log \sigma$ and ξ parameters. Suppose the walk is at $\log \sigma$ and ξ. For constants κ_1 and κ_2, $\log \sigma^{new} \sim N(\log \sigma, \kappa_1)$ and $\xi^{new} \sim N(\xi, \kappa_2)$ are propsed. The new values are accepted with probability $\min(e^\alpha, 1)$ where

$$\begin{aligned} \alpha \quad = \quad & \left[\sum_i \sum_j \log(M_j(\boldsymbol{s}_i) | \eta_j(\boldsymbol{s}_i), \sigma^{new}, \xi^{new}) + \log \pi(\sigma^{new}) + \log \sigma^{new} + \log \pi(\xi^{new}) \right] - \\ & \left[\sum_i \sum_j \log f(M_j(\boldsymbol{s}_i) | \eta_j(\boldsymbol{s}_i), \sigma, \xi) + \log \pi(\sigma) + \log \sigma + \log \pi(\xi) \right], \end{aligned}$$

and stay at the current values otherwise.

Update $\alpha(\boldsymbol{s}_i)$ and $\beta(\boldsymbol{s}_i)$ for each location $\boldsymbol{s}_i$ ($i = 1, \ldots, I$): Let $\boldsymbol{\alpha}^{-i}$ denote the vector $\boldsymbol{\alpha}$ without the ith element $\alpha(\boldsymbol{s}_i)$, and let $\boldsymbol{\beta}^{-i}$ denote the vector $\boldsymbol{\beta}$ without the ith element $\beta(\boldsymbol{s}_i)$. Then

$$\begin{aligned} & \pi(\alpha(\boldsymbol{s}_i), \beta(\boldsymbol{s}_i) | \boldsymbol{y}, \boldsymbol{\theta} \setminus \{\alpha(\boldsymbol{s}_i), \beta(\boldsymbol{s}_i)\}) \\ & \quad \propto \quad \left[\prod_{j=1}^{N(\boldsymbol{s}_i)} f(M_j(\boldsymbol{s}_i) | \eta_j(\boldsymbol{s}_i), \sigma, \xi) \right] \pi(\alpha(\boldsymbol{s}_i) | \boldsymbol{\alpha}^{-i}, \boldsymbol{\lambda}_\alpha, \tau_\alpha^2, \phi_\alpha, \omega_\alpha^2) \pi(\beta(\boldsymbol{s}_i) | \boldsymbol{\beta}^{-i}, \boldsymbol{\lambda}_\beta, \tau_\beta^2, \phi_\beta, \omega_\beta^2). \end{aligned}$$

Note that

$$\alpha(\boldsymbol{s}_i) | \boldsymbol{\alpha}^{-i}, \boldsymbol{\lambda}_\alpha, \tau_\alpha^2, \phi_\alpha \sim N(m_i, v_i^2),$$

where

$$\begin{aligned} m_i &= (\boldsymbol{X}\boldsymbol{\lambda}_\alpha)_i + \boldsymbol{V}(\tau_\alpha^2, \phi_\alpha, \omega_\alpha^2)^{i,-i}\boldsymbol{V}(\tau_\alpha^2, \phi_\alpha, \omega_\alpha^2)^{-i,-i}(\boldsymbol{\alpha}^{-i} - (\boldsymbol{X}\boldsymbol{\lambda}_\alpha)^{-i}), \\ v_i^2 &= (\boldsymbol{V}(\tau_\alpha^2, \phi_\alpha, \omega_\alpha^2)^{-i,-i} - \boldsymbol{V}(\tau_\alpha^2, \phi_\alpha, \omega_\alpha^2)^{i,-i}\boldsymbol{V}(\tau_\alpha^2, \phi_\alpha, \omega_\alpha^2)^{-i,-i}\boldsymbol{V}(\tau_\alpha^2, \phi_\alpha, \omega_\alpha^2)^{-i,i}). \end{aligned}$$

A similar result holds for $\boldsymbol{\beta}$. In the above formulae, $M^{-i,-j}$ means a matrix M without row i and without column j, and $M^{i,-j}$ means select row i of matrix M but leave out column j.

A Metropolis-Hastings symmetric random walk update is used on these two parameters. Supposing the walk is at $\alpha(\boldsymbol{s}_i)$ and $\beta(\boldsymbol{s}_i)$, $\alpha^{new}(\boldsymbol{s}_i) \sim N(\alpha(\boldsymbol{s}_i), \kappa_1)$ and $\beta^{new}(\boldsymbol{s}_i) \sim N(\beta(\boldsymbol{s}_i), \kappa_2)$ are proposed for positive variances κ_1 and κ_2, and these new values are accepted with probability $\min(e^\alpha, 1)$ where

$$\begin{aligned} \alpha = & \left\{ \sum_{j=1}^{N(\boldsymbol{s}_i)} \log f(M_j(\boldsymbol{s}_i)|\alpha^{new}(\boldsymbol{s}_i) + \beta^{new}(\boldsymbol{s}_i)a_j(\boldsymbol{s}_i), \sigma, \xi) + \right. \\ & \left. \log \pi(\alpha^{new}(\boldsymbol{s}_i)|\boldsymbol{\alpha}^{-i}, \boldsymbol{\lambda}_\alpha, \tau_\alpha^2, \phi_\alpha, \omega_\alpha^2) + \log \pi(\beta^{new}(\boldsymbol{s}_i)|\boldsymbol{\beta}^{-i}, \boldsymbol{\lambda}_\beta, \tau_\beta, \phi_\beta, \omega_\beta^2) \right\} - \\ & \left\{ \sum_{j=1}^{N(\boldsymbol{s}_i)} \log f(M_j(\boldsymbol{s}_i)|\alpha(\boldsymbol{s}_i) + \beta(\boldsymbol{s}_i)a_j(\boldsymbol{s}_i), \sigma, \xi) + \right. \\ & \left. \log \pi(\alpha(\boldsymbol{s}_i)|\boldsymbol{\alpha}^{-i}, \boldsymbol{\lambda}_\alpha, \tau_\alpha^2, \phi_\alpha, \omega_\alpha^2) + \log \pi(\beta(\boldsymbol{s}_i)|\boldsymbol{\beta}^{-i}, \boldsymbol{\lambda}_\beta, \tau_\beta^2, \phi_\beta, \omega_\beta^2) \right\}. \end{aligned}$$

Update $\boldsymbol{\lambda}_\alpha$: (The update for $\boldsymbol{\lambda}_\beta$ is similar.)

$$\pi(\boldsymbol{\lambda_\alpha}|\boldsymbol{y}, \boldsymbol{\theta} \setminus \boldsymbol{\lambda_\alpha}) \propto \pi(\boldsymbol{\alpha}|\boldsymbol{\lambda}_\alpha, \tau_\alpha^2, \phi_\alpha)\pi(\boldsymbol{\lambda}_\alpha),$$

and hence we sample $\boldsymbol{\lambda}_\alpha$ from a $N_I(P^{-1}\boldsymbol{q}, P^{-1})$ distribution where

$$P = \boldsymbol{X}^T\boldsymbol{V}(\tau_\alpha^2, \phi_\alpha, \omega_\alpha)^{-1}\boldsymbol{X} + Iv_{\alpha,\lambda}^{-1}; \quad \boldsymbol{q} = \boldsymbol{X}^T\boldsymbol{V}(\tau_\alpha^2, \phi_\alpha, \omega_\alpha)^{-1}\boldsymbol{\alpha} + Iv_{\alpha,\lambda}^{-1}\boldsymbol{m}_{\alpha,\lambda}.$$

Update τ_α^2, ϕ_α, and ω_α^2: (The update for τ_β^2, ϕ_β, and ω_β^2 is similar.)

$$\pi(\tau_\alpha^2, \phi_\alpha, \omega_\alpha^2|\boldsymbol{y}, \boldsymbol{\theta} \setminus \{\tau_\alpha^2, \phi_\alpha, \omega_\alpha^2\}) \propto \pi(\boldsymbol{\alpha}|\boldsymbol{\lambda}_\alpha, \tau_\alpha^2, \phi_\alpha, \omega_\alpha^2)\pi(\tau_\alpha^2)\pi(\phi_\alpha)\pi(\omega_\alpha^2)$$

A Metropolis-Hastings symmetric random walk update is used on the log scales for these three parameters. Suppose the walk is at $\log \tau_\alpha^2$, $\log \phi_\alpha$, and $\log \omega_\alpha^2$. For a constant 3×3 positive definite matrix Λ, $(\log \tau_\alpha^{2\,new}, \log \phi_\alpha^{new}, \log \omega_\alpha^{2\,new})^T \sim N_3((\log \tau_\alpha^2, \log \phi_\alpha, \log \omega_\alpha^2)^T, \Lambda)$ are proposed. The new values are accepted with probability $\min(e^\alpha, 1)$ where

$$\begin{aligned}
\alpha \;=\; & \big[\log \pi(\boldsymbol{\alpha}|\boldsymbol{\lambda}_\alpha, \tau_\alpha^{2\,new}, \phi_\alpha^{new}) + \\
& \log \pi(\tau_\alpha^{2\,new}) + \log \tau_\alpha^{2\,new} + \log \pi(\phi_\alpha^{new}) + \log \phi_\alpha^{new} + \log \pi(\omega_\alpha^{2\,new}) + \log \omega_\alpha^{2\,new}\big] - \\
& \big[\log \pi(\boldsymbol{\alpha}|\boldsymbol{\lambda}_\alpha, \tau_\alpha^2, \phi_\alpha) + \\
& \log \pi(\tau_\alpha^2) + \log \tau_\alpha^2 + \log \pi(\phi_\alpha) + \log \phi_\alpha + \log \pi(\omega_\alpha^2) + \log \omega_\alpha^2\big]\,,
\end{aligned}$$

and stay at the current values otherwise.

6 A graphical illustration of the Bayesian learning

As an illustration of the range of learning in our model, Fig. 4 compares the prior density (in gray) and posterior density (in black) for the parameters ϕ_β and $\lambda_{\beta,3}$ of the decadal maxima model. There is a minor positive shift in the posterior values for the spatial range parameter ϕ_β as compared to the prior, indicating very weak learning about the length scales in the decadal maxima model. In contrast, the posterior distribution for the spatial intercept parameter $\lambda_{\beta,3}$ changes dramatically from the vague prior distribution, indicating strong learning.

7 Diagnostics and model comparisons

The posterior means of the scale and shape parameters (σ and ξ), as well as the location parameters, $\eta_j(\boldsymbol{s})$, for each decade j and centroid $\boldsymbol{s}$, are calculated for the fits to both the maxima and the minima. Following [Coles, 2001, Section 6.2.3, p.110], we calculate the

standardized variable, $Z_j(\boldsymbol{s}_i)$, at each decade j and centroid $\boldsymbol{s}_i$ via

$$Z_j(\boldsymbol{s}_i) \;=\; \frac{1}{\widetilde{\xi}}\log\left\{1+\widetilde{\xi}\left(\frac{M_j(\boldsymbol{s}_i)-\widetilde{\eta}_j(\boldsymbol{s}_i)}{\widetilde{\sigma}}\right)\right\},$$

where here a bar over a parameter is taken to denote the posterior mean of that parameter calculated from the MCMC output. We obtain a quantile-quantile plot for the maxima for each location by ordering the $\{Z_j(\boldsymbol{s}_i)\}$ values into order statistics $\{Z_{(j)}(\boldsymbol{s}_i)\}$ and plotting the pairs

$$\{(Z_{(j)}(\boldsymbol{s}_i), -\log(-\log(j/(N(\boldsymbol{s}_i)+1)))) : j = 1,\ldots,N(\boldsymbol{s}_i)\},$$

on a scatterplot. Fig. 5 (respectively Fig. 6) shows the quantile-quantile plots for the maxima (respectively minima) models fit in the main article. These quantile plots show that the GEV models fit better for the decadal maxima than for the decadal minima. The lack of fit for the minima generally occurs for only a small fraction of points in the lower tail, indicating that, overall, the GEV models are still a reasonably good choice for the minima. In particular, the model misspecification is not sufficiently substantial to impact the general conclusions about spatial and temporal trends discussed above. In spite of evidence for the model misspecification being weak, we nevertheless fit a number of other models to the decadal maxima and minima series to ensure that our conclusions are robust to variations of the modeling assumptions discussed in the text:

1. A simpler model that included no covariate for latitude and no nugget terms (ω_α^2 and ω_β^2) in the models for the spatially-varying intercept and slope parameters.

2. A model that included the covariate for latitude but removed both nugget effects.

3. A more general model that allowed the scale parameters to vary by spatial location, with $\log\sigma(\boldsymbol{s}_i) \sim N(m_\sigma, v_\sigma)$, conditionally independent over locations. Priors were then necessary for m_σ and v_σ in order to learn across locations.

While there were minor differences in the posterior summaries of certain the parameters under these different GEV models, we did not detect any significant differences in the quantile-quantile plots, or in the posterior summaries of the slope parameter, $\beta(\boldsymbol{s})$. We chose the model presented in the article, because we believe it accurately allows us to model *and interpret* the nonstationary spatial effects in the intercept and slope parameters (e.g., the effect of latitude is important and should be included in the model), without overly complicating matters by including a spatially varying scale parameter.

References

S. Banerjee, B. P. Carlin, and A. E. Gelfand. *Hierarchical modeling and analysis for spatial data*. Chapman and Hall, Boca Raton, Florida, 2004.

S. Coles. *An Introduction to Statistical Modeling of Extreme Values*. Springer-Verlag, New York, 2001.

N. Cressie. *Statistics for Spatial Data (Revised edition)*. Wiley-Interscience, New York, 1993.

J. Hansen and S. Lebedeff. Global trends of measured surface air temperature. *JGR*, 92:345–13, 1987.

J. E. Heffernan and J. A. Tawn. A conditional approach for multivariate extreme values. *J. of the Royal Statistical Society: Series B*, 66:497–546, 2004.

A. W. Ledford and J. A. Tawn. Statistics for near independence in multivariate extreme values. *Biometrika*, 83:169–187, 1996.

M.E. Mann and J. Park. Spatial correlations of interdecadal variation in global surface temperatures. *Geophysical Research Letters*, 20:1055–1058, 1993.

J. Pickands. Statistical inference using extreme order statistics. *Annals of Statistics*, 3:119–131, 1975.

M. Ribatet, D. Cooley, and A. C. Davison. Bayesian inference from composite likelihoods, with an application to spatial extremes. *Statistica Sinica*, 22:813–845, 2012.

M. P. Tingley and P. Huybers. A Bayesian algorithm for reconstructing climate anomalies in space and time. Part 1: Development and applications to paleoclimate reconstruction problems. *J. Clim*, 23:2759–2781, 2010.

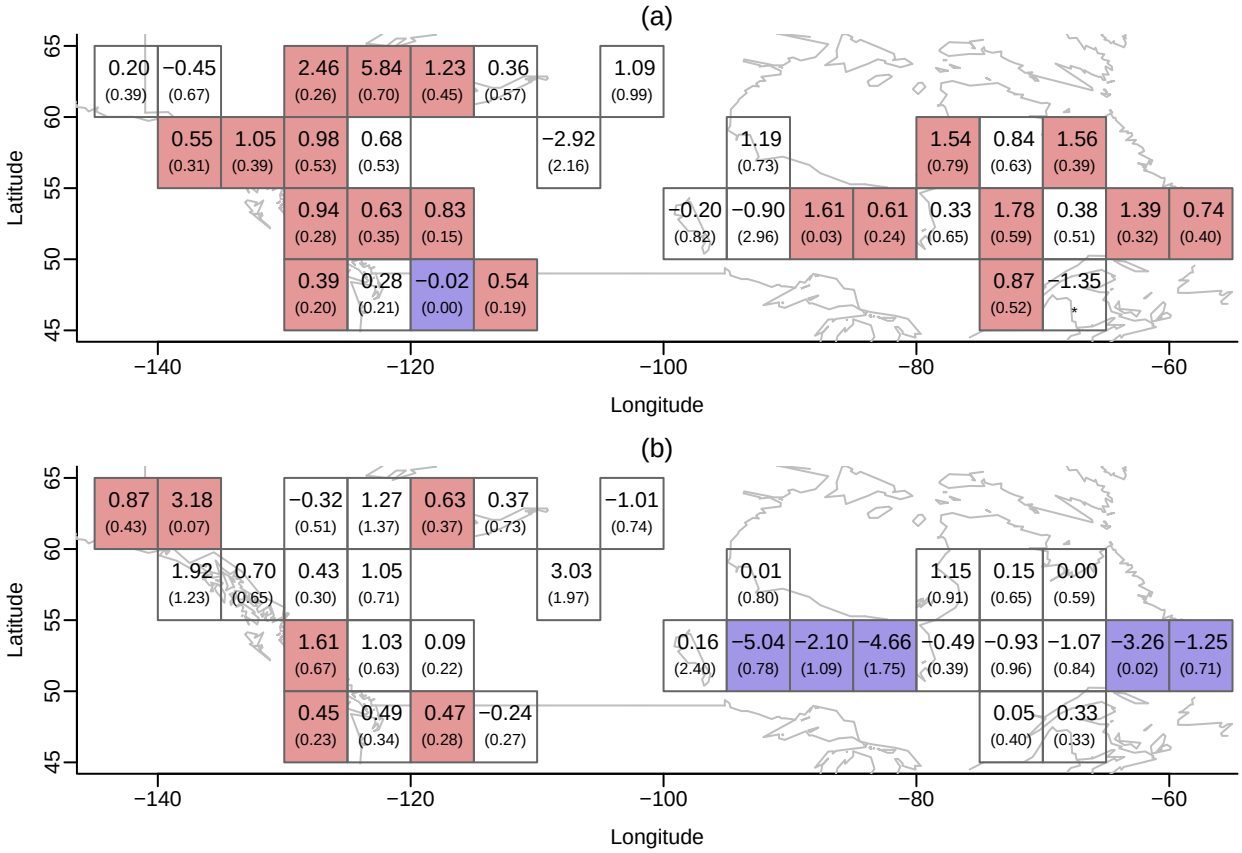

Figure 1: **(a)** ML estimates of the temporal trend in the location parameter for the site-by-site maxima GEV models (Eqs. (2) and (3)). In each box, the top number is the ML estimate of the slope, and the bracketed lower number is the standard error (an asterisk indicates that there was not enough data to reliably estimate the standard error). Pink shading indicates that the 90% confidence interval for the slope parameter at that location contains only positive values (i.e. interval is entirely above 0), while blue shading indicates that the 90% confidence interval contains only negative values. **(b)** As in (a), but for the GEV model for the minima.

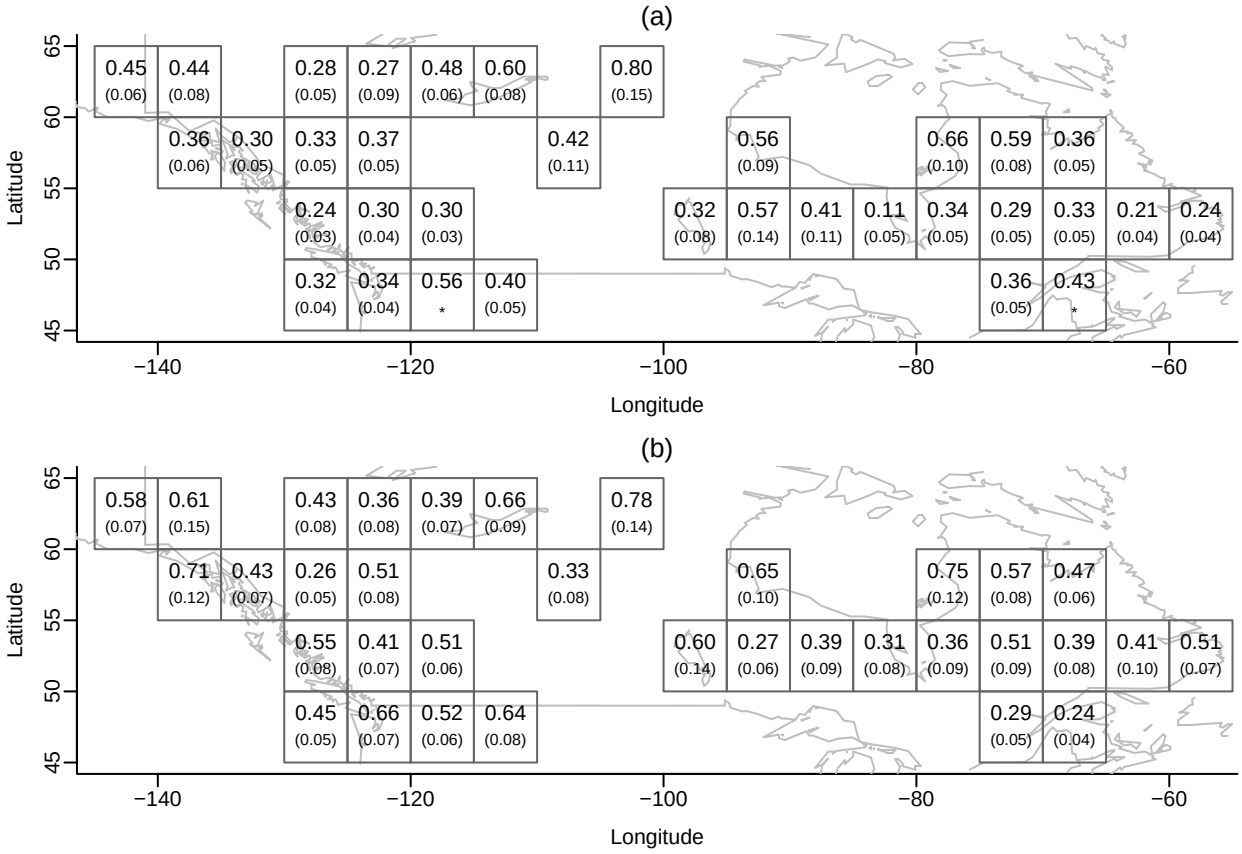

Figure 2: **(a)** The ML estimates of the scale parameter in the site-by-site maxima GEV models. In each box, the top number is the ML estimate of the scale parameter, and the bracketed lower number is the standard error (an asterisk indicates that there was not enough data to reliably estimate the standard error) **(b)** As in (a), but for the GEV model for the minima.

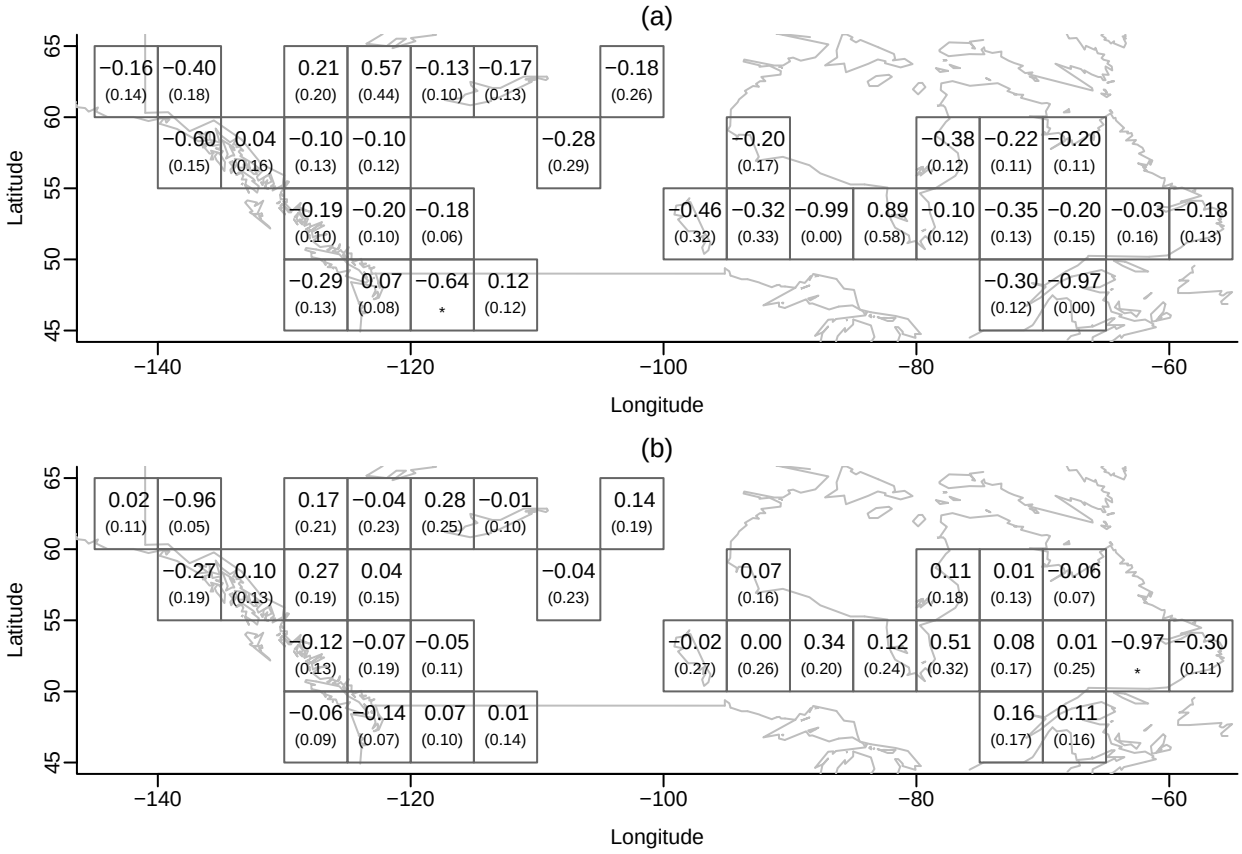

Figure 3: **(a)** The ML estimates of the shape parameter in the site-by-site maxima GEV models. In each box, the top number is the ML estimate of the shape parameter, and the bracketed lower number is the standard error (an asterisk indicates that there was not enough data to reliably estimate the standard error) **(b)** As in (a), but for the GEV model for the minima.

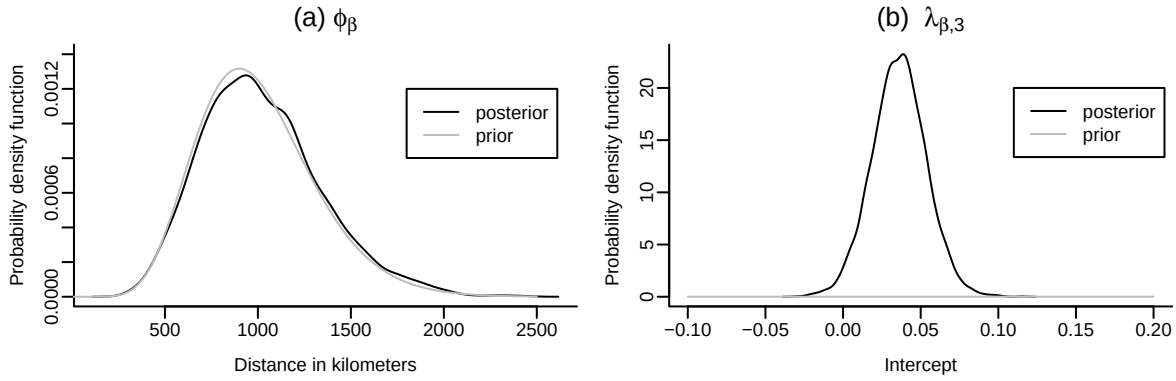

Figure 4: **(a)** A comparison of the prior (in gray) and posterior (in black) densities for the spatial correlation parameter ϕ_β in the decadal maxima model. **(b)** The corresponding comparison for the spatial intercept parameter $\lambda_{\beta,3}$.

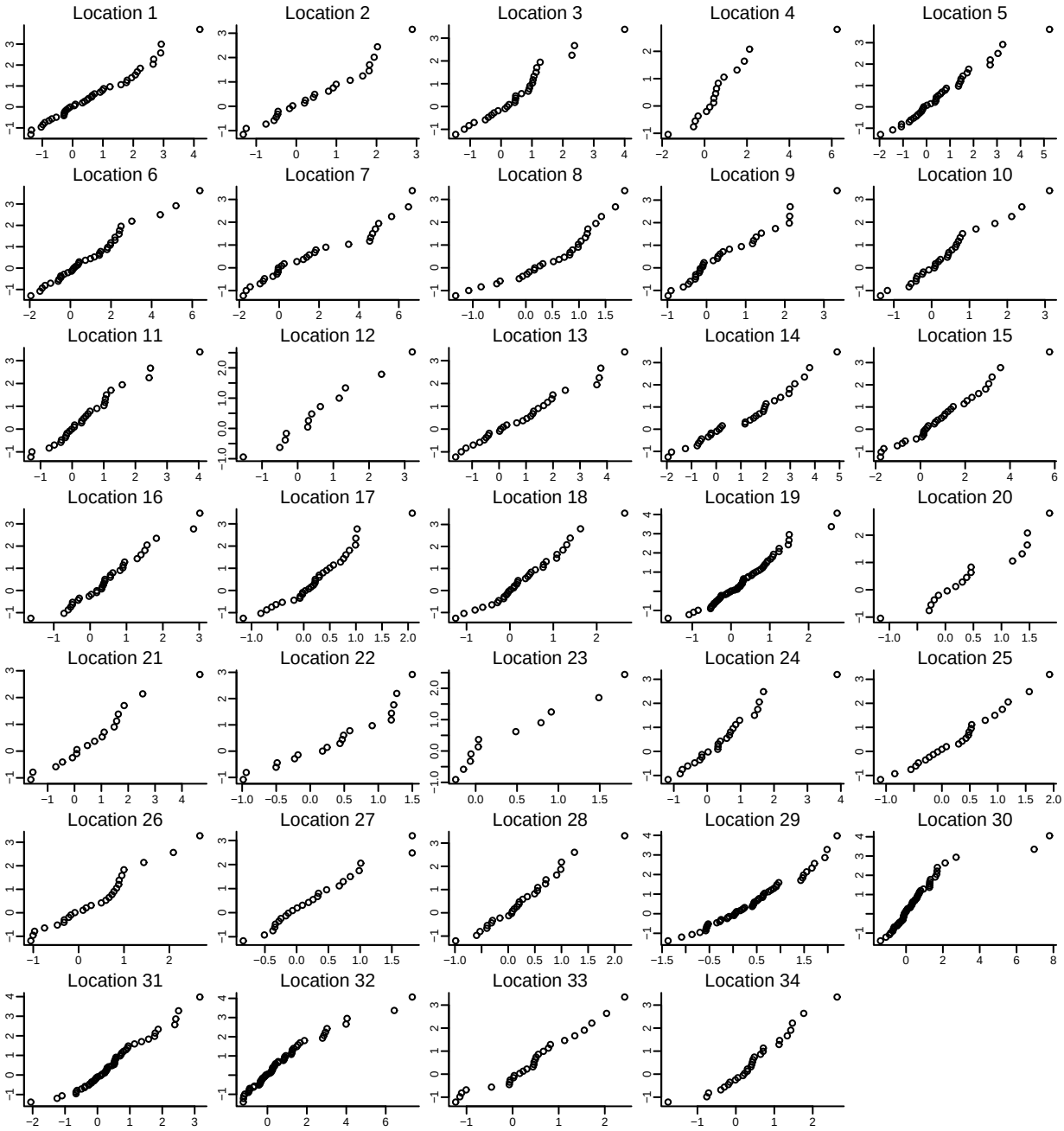

Figure 5: Quantile-quantile plots at all locations evaluatedat the posterior mean values of the parameters of the maxima model. The x scales shows the ordered standardized value $z_{(j)}$, and the y scale shows the $-\log(-\log(j/(N(\boldsymbol{s}_i)+1)))$ [Coles, 2001, p.110]

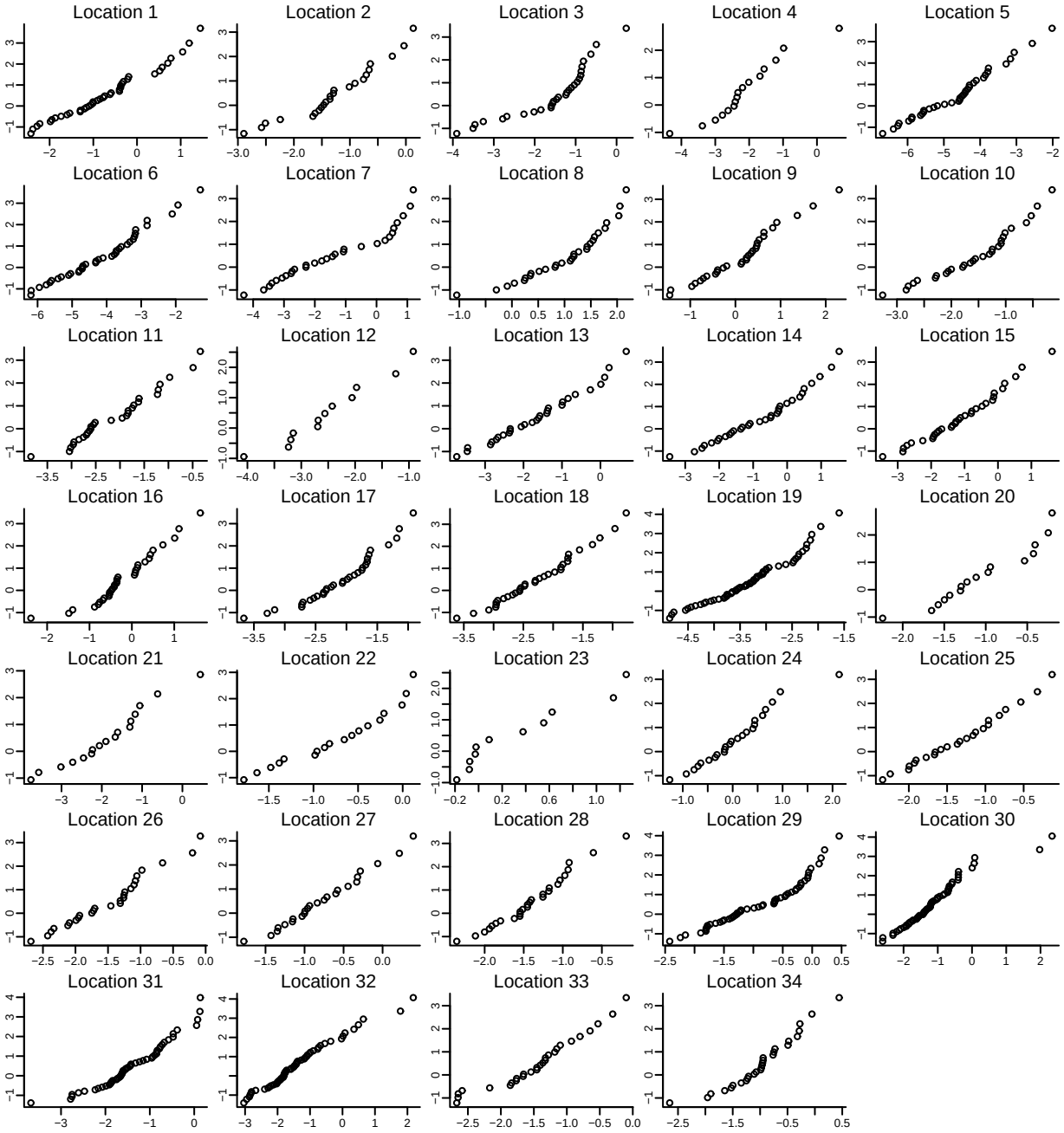

Figure 6: As Fig. 5, but for the parameters of the minima model.