Supporting Information

Spatial Modelling Framework for the Characterisation of Rainfall Extremes at Different Durations and under Climate Change

Eric A. Lehmann[1], Aloke Phatak[2], Alec Stephenson[3], Rex Lau[4]

[1]CSIRO Data61, Acton ACT, Australia
[2]Dept. of Mathematics and Statistics, Curtin University, Bentley WA, Australia
[3]CSIRO Data61, Clayton South VIC, Australia
[4]CSIRO Data61, Floreat WA, Australia

This supplementary document provides supporting material for the above paper, detailing some technical aspects of the proposed methodology. The various equations and sections referenced herein can be found in the main article. The sections in this document are not meant to logically follow one another, with the material organised so that the characteristics of the datasets and proposed methods are presented first (Sections S1 – S3), while more mathematical derivations are given at the end (Sections S4 and S5).

S1 Bayesian inference and prior distributions

Based on the notation used in Sec. 3, and defining the set $\boldsymbol{\Omega}$ of parameters as:

$$\boldsymbol{\Omega} = \bigcup_{\chi \in \{\tilde{\mu},\sigma,\xi,\theta,\eta\}} \{\alpha_\chi, \lambda_\chi, \boldsymbol{\beta}_\chi\} \tag{S1}$$

the posterior distribution of the model parameters is determined as follows:

$$p(\tilde{\boldsymbol{\mu}}, \boldsymbol{\sigma}, \boldsymbol{\xi}, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\Omega} | \mathbf{Y}) \propto p(\mathbf{Y}, \tilde{\boldsymbol{\mu}}, \boldsymbol{\sigma}, \boldsymbol{\xi}, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\Omega}) \tag{S2}$$

$$= p(\mathbf{Y} | \tilde{\boldsymbol{\mu}}, \boldsymbol{\sigma}, \boldsymbol{\xi}, \boldsymbol{\theta}, \boldsymbol{\eta}) \cdot p(\boldsymbol{\Omega}) \cdot \prod_{\chi \in \{\tilde{\mu},\sigma,\xi,\theta,\eta\}} p(\boldsymbol{\chi} | \boldsymbol{\beta}_\chi, \alpha_\chi, \lambda_\chi) \tag{S3}$$

where $p(\boldsymbol{\Omega})$ represents the prior density function:

$$p(\boldsymbol{\Omega}) = \prod_{\chi \in \{\tilde{\mu},\sigma,\xi,\theta,\eta\}} p(\boldsymbol{\beta}_\chi)\ p(\alpha_\chi)\ p(\lambda_\chi). \tag{S4}$$

Following Davison *et al.* (2012), we use conjugate Gamma, inverse Gamma and multivariate-normal priors, respectively:

$$\lambda_\chi \sim \text{Gamma}(\kappa_{\lambda_\chi}, \gamma_{\lambda_\chi}) \tag{S5}$$

$$\alpha_\chi \sim \text{InvGamma}(\kappa_{\alpha_\chi}, \gamma_{\alpha_\chi}) \tag{S6}$$

$$\boldsymbol{\beta}_\chi \sim \text{MVN}(\boldsymbol{\mu}_{\beta_\chi}, \boldsymbol{\Sigma}_{\beta_\chi}) \tag{S7}$$

where $\kappa_{(\cdot)}$ and $\gamma_{(\cdot)}$ represent the shape and scale hyper-parameters of the respective distributions, and $\boldsymbol{\mu}_{\beta_\chi}$ and $\boldsymbol{\Sigma}_{\beta_\chi}$ are the prior mean and variance matrix of $\boldsymbol{\beta}_\chi$, respectively.

Inferences from the posterior distribution in Eq. (S3) are obtained using Markov chain Monte Carlo (MCMC) simulation, with implementation details provided in Sec. S5. Selection of the hyper-parameters in our implementation was performed as described in Davison *et al.* (2012). Priors for the regression coefficients $\boldsymbol{\beta}_\chi$ were set to have very large variances. Suitable prior densities were chosen for α_χ and λ_χ on the basis of exploratory analyses of fitted marginal distributions: the priors were set to have means equal to the average marginal maximum likelihood estimates (MLEs), with much larger variances.

S2 Implementation and computational aspects

This section provides specific implementation details for the BHM-HF and BHM-EI methods, as well as their computational requirements when applied to the SYD and SEQ datasets. The following results are based on the models described in Sections 3 and 4 of the main manuscript, with elevation and geographical location used as covariates. The results are based on 150,000 iterations of the MCMC algorithms for each model parameter, with the first 10,000 iterations discarded (burn-in), and thinned by a factor 40. Step sizes for proposal samples within the Metropolis–Hastings steps were set so as to ensure an acceptance rate of about 45%.

For each dataset, MCMC sampling is used to fit a total of P model parameters:

$$P = S \cdot 5 + U \cdot 3 + \sum_{\chi \in \{\tilde{\mu}, \tilde{\sigma}^{(24)}, \xi, \theta, \eta\}} (N_\chi + 1) + 10 \tag{S8}$$

where the first term on the right-hand side represents the number of parameters for the pluviometer stations, the second term is that of the daily stations, the third term corresponds to the regression coefficients ($\boldsymbol{\beta}$) for N_χ covariates and an intercept term, and the last term relates to the spatial model parameters (α and λ). This results in a total of $P_{\text{SYD}} = 3266$ model parameters for the SYD dataset, and $P_{\text{SEQ}} = 4554$ for SEQ.

S2.1 Chain diagnostics

Table S1 provides the lag for which the autocorrelation in the MCMC chains drops below a negligible level. To summarise the results obtained for a large number of model parameters (P), we present averaged values for three different categories, namely GEV parameters, regression coefficients and spatial process parameters. Further results (not presented here) show that the corresponding integrated autocorrelation times averaged over these categories fall within the range between 1.15 and 2.28. These results indicate that an acceptably low autocorrelation is achieved for the chains of all parameters on average.

Convergence results for the same categories of model variables are provided in Table S2. It presents the percentage of MCMC chains having passed the Geweke (Geweke, 1992) and Heidelberger–Welch (Heidelberger and Welch, 1983) diagnostic tests of convergence. These results show no evidence of systematic convergence issues with the simulated algorithms. Results from the Raftery–Lewis diagnostic (Raftery and Lewis, 1992b), not included here, show that the average dependence factor for all parameter categories falls within the range between 1.02 and 1.5, thereby pointing to satisfactory chain lengths and mixing properties; typically, dependence factors larger than 5 would point to a potential convergence failure (Raftery and Lewis, 1992a).

	SYD		SEQ	
	BHM-HF	BHM-EI	BHM-HF	BHM-EI
GEV parameters – pluviometer	2.8 (1550)	2.6 (1550)	2.4 (1200)	2.4 (1200)
GEV parameters – daily	1.6 (1686)	2.0 (1686)	1.7 (3324)	2.3 (3324)
Regression coefficients (20)	1.6	2.6	1.7	3.0
Spatial model parameters (10)	4.7	3.2	2.7	4.2

Table S1: Average lag for which the MCMC chain autocorrelation becomes negligible. Results are averaged for three different categories of parameters: GEV parameters $\tilde{\mu}$, $\tilde{\sigma}^{(24)}$, ξ, θ and η (further split between daily and pluviometer stations), regression coefficients $\boldsymbol{\beta}$, and spatial process parameters α and λ. Numbers in parentheses indicate how many parameters are used in the averaging.

	SYD				SEQ			
	BHM-HF		BHM-EI		BHM-HF		BHM-EI	
	GD	HD	GD	HD	GD	HD	GD	HD
GEV parameters – pluviometer	94.3	99.2	93.2	99.4	94.1	99.3	95.3	99.6
GEV parameters – daily	94.9	99.4	94.8	99.2	94.1	99.4	94.5	99.2
Regression coefficients	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Spatial model parameters	100.0	100.0	90.0	100.0	100.0	90.0	100.0	100.0

Table S2: Proportion of model variables whose MCMC chain passed the Geweke (GD) and Heidelberger–Welch (HD) diagnostic test. All values are given in % as averages over the respective parameter categories (the number of parameters in each category is provided in Table S1).

S2.2 Computational requirements

The BHM-HF and BHM-EI algorithms were implemented in R (R Core Team, 2015) and optimised for computational efficiency. Parallelisation of the main MCMC sampling loop was considered (sampling of the $\boldsymbol{\beta}_{(\cdot)}$, $\alpha_{(\cdot)}$ and $\lambda_{(\cdot)}$ coefficients can be parallelised) but did not substantially reduce the computational efforts due to large overheads. An implementation using block sampling of the model variables was also tested but did not lead to substantial improvements in execution times either: the computational advantages gained through block sampling were offset manyfold by much longer convergence times.

The BHM fitting routines were thus executed sequentially on a single processing core (of a Dual Xeon 8-core E5-2650 2GHz compute node) on CSIRO's high-performance computing infrastructure 'Bragg'. Table S3 shows the detail of the resulting computing requirements for each of the algorithms applied to each of the selected regions of interest. Long chain lengths were specifically used in this work to monitor and investigate potential model convergence issues. Without major degradation in model performance, it is expected that these computation times can be reduced considerably by using a combination of fewer MCMC iterations, a smaller thinning factor (MacEachern and Berliner, 1994), and an implementation based on the simulation of multiple parallel chains.

Once the model parameters are fitted, further calculations (such as IFD curves at ungauged locations, future-climate IFD curves, or gridded outputs of return levels) can be parallelised and were executed on a 16-core node of the Bragg computing infrastructure. Such results were typically achieved in a matter of minutes, up to a maximum of one hour (depending on the number of locations involved).

	SYD		SEQ	
	BHM-HF	BHM-EI	BHM-HF	BHM-EI
median loop time (s)	1.752	1.776	2.760	2.878
total CPU time (h)	75.58	82.52	135.20	131.98

Table S3: Average MCMC loop time and total processing time (based on a total of 150,000 iterations) for model fitting of $P_{\text{SYD}} = 3266$, respectively $P_{\text{SEQ}} = 4554$ parameters.

S3 Model evaluation

S3.1 Koutsoyiannis *et al.* relationship

Exploratory data analysis was carried out on our precipitation datasets in order to assess the dependence model across duration provided in Koutsoyiannis *et al.* (1998). Fig. S1 provides an overview of these results for four randomly selected pluviometer stations in the SEQ dataset. The black dots (with error bars) show the MLEs of the GEV parameters obtained for each duration independently. The red lines are based on MLEs of the GEV parameters using the Koutsoyiannis *et al.* parameterisation of Sec. 2.2 to combine all the durations into a single composite likelihood. Both of these results are obtained for each station separately.

The Koutsoyiannis *et al.* model defines the GEV parameters $\tilde{\mu}$ and ξ as constant across duration, while the duration dependence for $\tilde{\sigma}$ is modelled as per Eq. (2). The plots in Fig. S1 indicate that this dependence model can be reasonably assumed to be valid for the considered stations. With similar results obtained for a vast majority of stations in both the SYD and SEQ datasets, the Koutsoyiannis *et al.* relationship used in Sec. 2.2 was thus found to provide an adequate fit for our data.

For insight, the plots in Fig. S1 also include the Bayesian estimates (with 95% credible intervals) of the same parameters obtained with the BHM-HF method. These model-based estimates draw on the information available across durations and neighbouring stations (through the spatial model), and consequently differ from the MLEs in general, most notably by exhibiting substantially reduced levels of uncertainty. Similar results are obtained with BHM-EI.

S3.2 Cumulative distribution plots

Cumulative distribution function (CDF) plots can be used to provide insight into how well the extreme rainfall data fit under a specific model. Plotting the empirical CDF of the data together with the corresponding model-estimated CDF allows the identification of potential discrepancies between the data and the model outputs.

The dataset $\mathbf{Y}$ used in this work is bi-variate, corresponding to the top $r = 2$ maxima in the time series of precipitation data. A CDF display would thus involve three-dimensional plots of cumulative distribution surfaces over the $(y^{(1)}, y^{(2)})$ domain, leading to considerable plotting and interpretation difficulties. Instead, here we present results in terms of the marginal distributions for $y^{(1)}$, i.e. the top maximum only. On the basis of the r-largest density $\text{GEV}_r(\mathbf{y}; \tilde{\mu}, \tilde{\sigma}, \xi)$, it can be shown that the marginal density for $y^{(1)}$ is the univariate GEV density with same parameters, i.e. $\text{GEV}(y^{(1)}; \tilde{\mu}, \tilde{\sigma}, \xi)$.

Fig. S2 presents the BHM-HF results for three pluviometer stations in the SYD dataset, selected to represent a variety of geographical locations and number of available data points (years). The CDFs are evaluated at four different durations, namely 5 min, 1, 12 and 48 h, as a representative subset of the range of durations in the dataset. These plots indicate that the model provides a satisfactory fit of the data. Similar results (not shown here) are achieved for

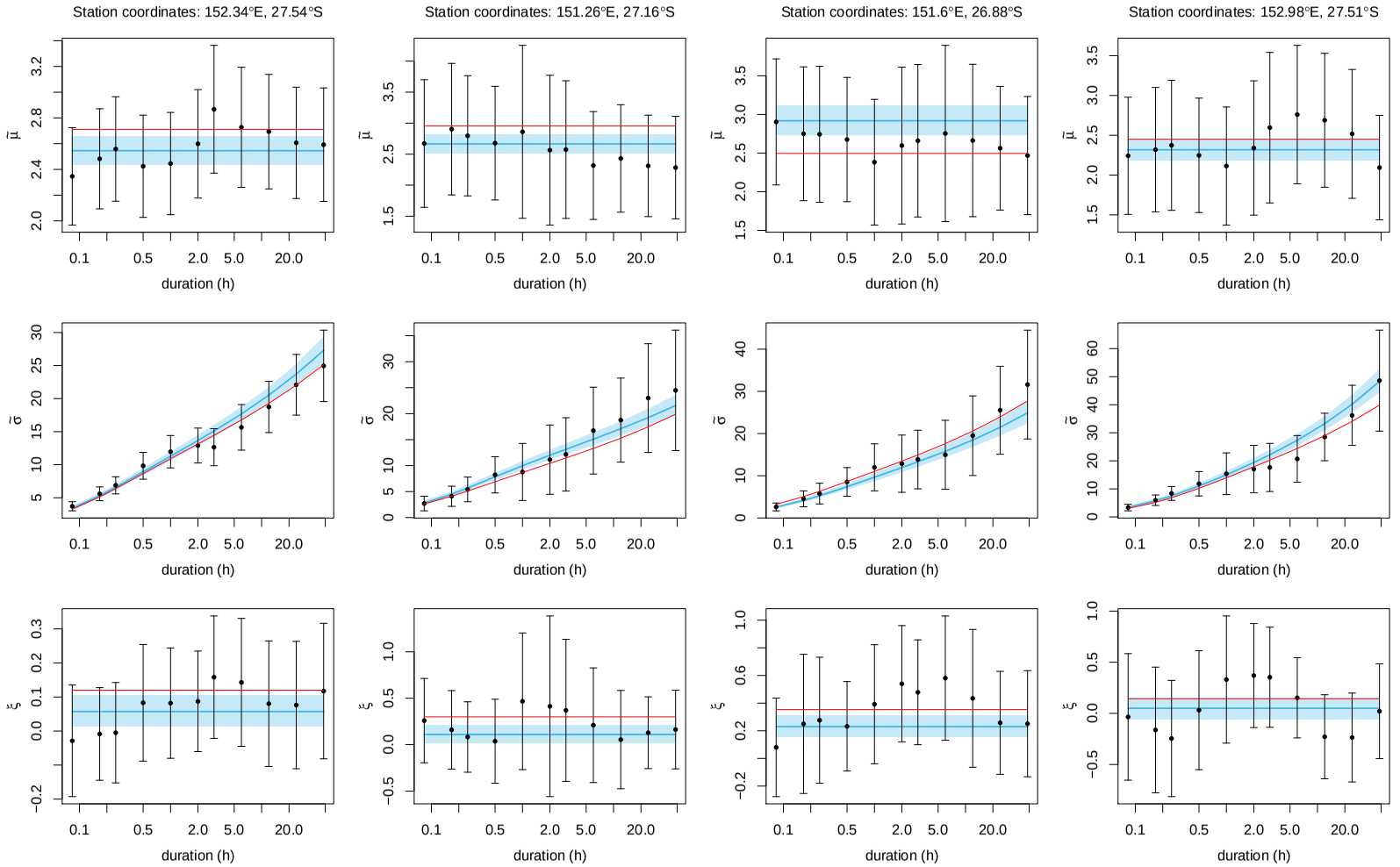

Figure S1: Estimates of GEV parameters $\tilde{\mu}$, $\tilde{\sigma}$ and ξ (rows) for four pluviometer stations (columns) in the SEQ dataset. The black dots with error bars (asymptotic 95% confidence intervals) show the MLEs for each duration separately, while the red lines represent MLEs of the GEV parameters using the Koutsoyiannis *et al.* relationship (each station being considered independently). The blue lines show the model-based estimates (with shaded regions representing the 95% posterior credible intervals) obtained with BHM-HF.

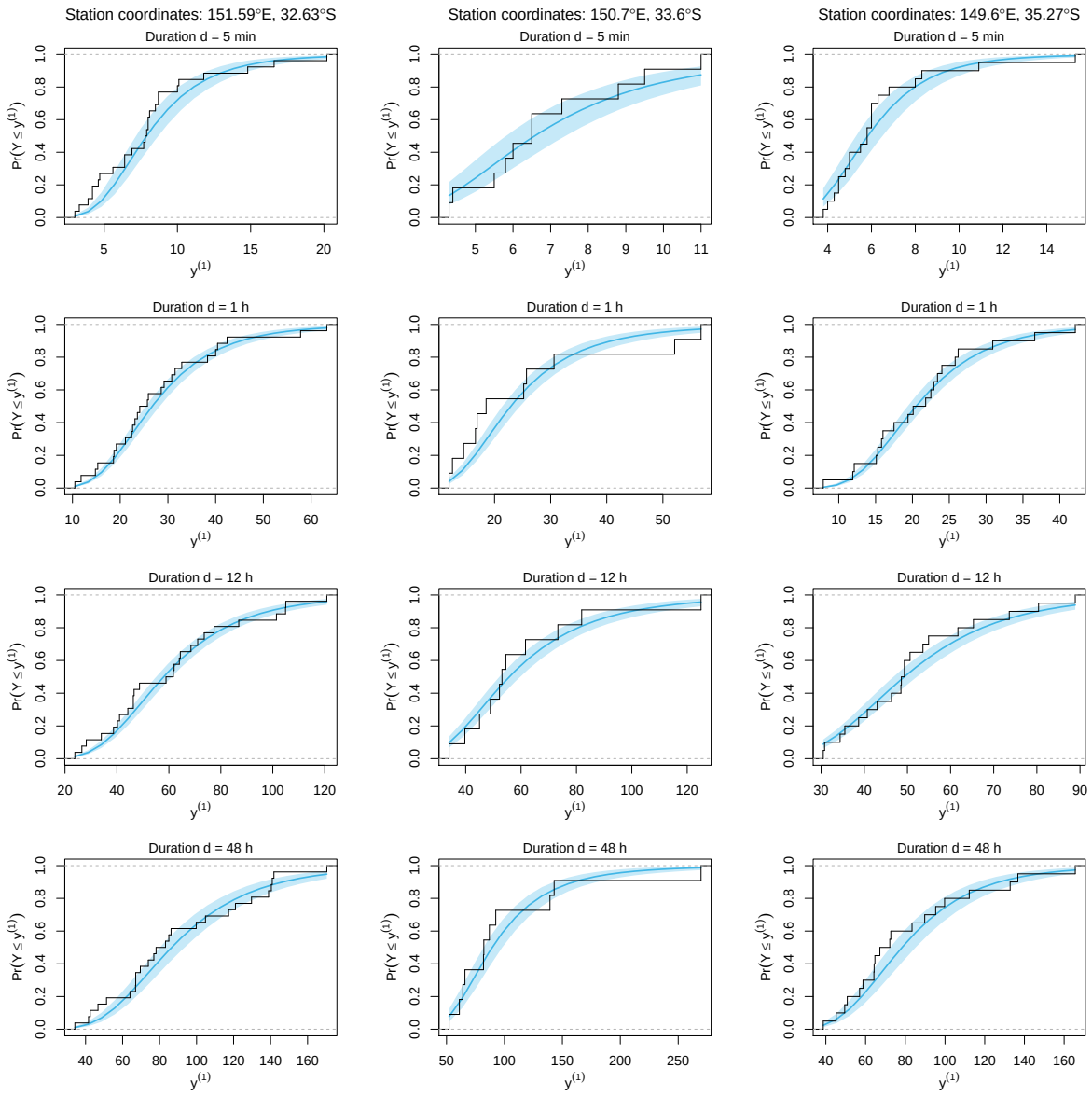

Figure S2: CDF plots for three pluviometer stations (columns) in the SYD dataset and four durations (rows). The black lines show the empirical CDFs for the top maximum data ($y^{(1)}$). The blue lines show the model-based CDFs (with shaded regions representing the 95% posterior credible intervals) obtained with the BHM-HF algorithm.

other durations and other pluviometer stations, as well as for the BHM-EI algorithm and the SEQ dataset.

S3.3 Comparison of posterior densities

Fig. S3 shows plots of the posterior densities (kernel estimates) for each of the main model parameters, for both regions and both approaches (BHM-HF and BHM-EI). Results in this figure suggest that the geographical covariates (easting, northing and elevation in the second, third and fourth columns, respectively) have more inferential power in the SYD region (blue curves), with about 11 of the fitted regression coefficients ($\boldsymbol{\beta}$) being different from zero (i.e. credible intervals exclude zero), compared to about 5 in the SEQ region (red curves).

BHM-HF (solid lines) and BHM-EI (dashed lines) yield similar results for most of the parameters, and particularly for the scale parameter $\tilde{\sigma}^{(24)}$, which represents the main influencing factor in the calculation of IFD curves. Some differences can be seen for estimates related to $\tilde{\mu}$, with the most notable difference occurring with $\lambda_{\tilde{\mu}}$. Here, the BHM-HF results point to a high level of spatial correlation for $\tilde{\mu}$, whereas $\lambda_{\tilde{\mu}}$ becomes virtually negligible under BHM-EI. In contrast, the exact opposite can be seen to occur for λ_{ξ}.

S4 Model selection and validation

Fitting a Bayesian hierarchical model should be followed by an assessment of how well the model fits the data, and of its adequacy in the application for which it was implemented. In this section, we review some aspects of model checking and model comparison which we apply to our BHMs. The following descriptions are mostly based on the review presented in Stephenson (2016, Supplement section) and the references therein, adapted specifically to the Bayesian models proposed in this work.

S4.1 Posterior predictive checking

Posterior predictive checking is a diagnostic that can be used to determine whether the observed data $\mathbf{Y}$ looks plausible under the posterior predictive distribution $p(\check{\mathbf{Y}}|\mathbf{Y})$ (Stephenson, 2016; Gelman *et al.*, 2015). If the model is appropriate, then the simulated data samples $\check{\mathbf{Y}}$ are expected to be similar to the observed maxima $\mathbf{Y}$.

Assume that the MCMC sampling procedure produces K draws (MCMC chains) of the model parameters:

$$\left\{\tilde{\boldsymbol{\mu}}_{+}^{(k)}, \tilde{\boldsymbol{\sigma}}_{+}^{(24)(k)}, \boldsymbol{\xi}_{+}^{(k)}, \boldsymbol{\theta}^{(k)}, \boldsymbol{\eta}^{(k)}\right\}, \qquad k = 1, \ldots, K \tag{S9}$$

from the respective marginal posterior distributions. Simulated data samples $\check{\mathbf{y}}^{(k)}$ from the posterior predictive distribution can then simply be obtained by drawing from

$$p\big(\check{\mathbf{Y}}|\tilde{\boldsymbol{\mu}}_{+}^{(k)}, \tilde{\boldsymbol{\sigma}}_{+}^{(24)(k)}, \boldsymbol{\xi}_{+}^{(k)}, \boldsymbol{\theta}^{(k)}, \boldsymbol{\eta}^{(k)}\big) \tag{S10}$$

for each $k = 1, \ldots, K$.

For pluviometer station $s \in \{1, \ldots, S\}$, year $t \in \{1, \ldots, T_s\}$, and accumulation duration $d_{s,i}$, $i \in \{1, \ldots, D_s\}$, posterior predictive samples are thus generated according to:

$$\check{\mathbf{y}}_{s,t,i}^{(k)} \sim \mathrm{GEV}_r^{*}\left(\tilde{\mu}_s^{(k)}, \tilde{\sigma}_s^{(24)(k)}, \xi_s^{(k)}, \theta_s^{(k)}, \eta_s^{(k)}\right). \tag{S11}$$

For daily stations $u \in \{S+1, \ldots, S+U\}$, generating the posterior predictive samples depends on the considered model type. For BHM-HF, we have:

$$H \cdot \check{\mathbf{y}}_{u,t}^{(\mathrm{D})(k)} \sim \mathrm{GEV}_r\left(\tilde{\mu}_u^{(k)}, \tilde{\sigma}_u^{(24)(k)}, \xi_u^{(k)}\right) \tag{S12}$$

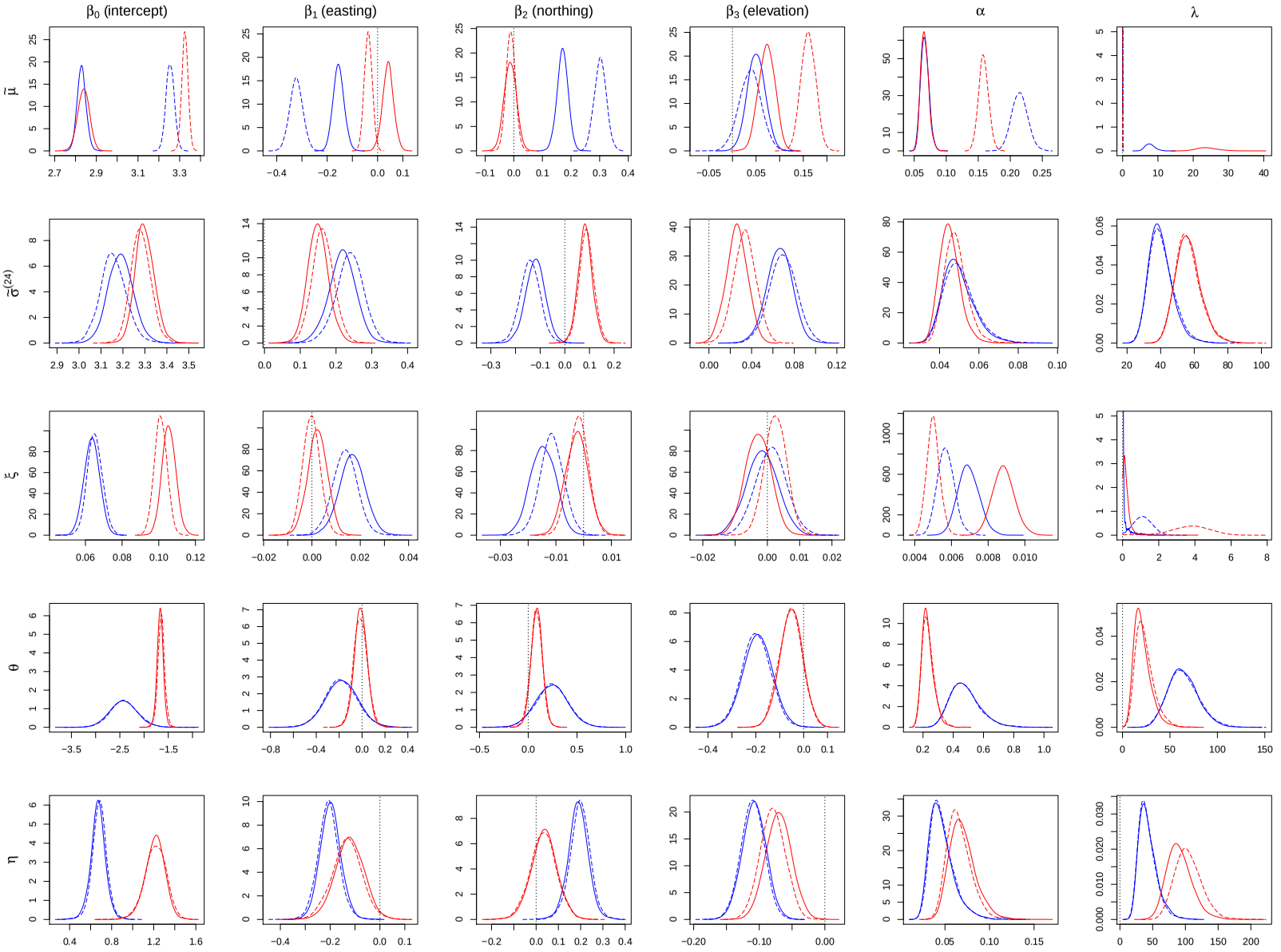

Figure S3: Estimated posterior densities of model parameters for the SYD region (blue) and SEQ region (red), using BHM-HF (solid lines) and BHM-EI (dashed lines). Each column represents a different coefficient for each GEV parameter (one per row).

while the following sampling process is used for BHM-EI:

$$\check{\mathbf{y}}_{u,t}^{(\mathrm{D})(k)} \sim \mathrm{GEV}_r\left(\tilde{\mu}_u^{(\mathrm{D})(k)}, \sigma_u^{(\mathrm{D})(k)}, \xi_u^{(\mathrm{D})(k)}\right). \tag{S13}$$

A comparison of the K posterior predictive samples $\{\check{\mathbf{y}}^{(k)}\}$ with the observed data $\mathbf{Y}$ can be achieved using a specific test statistic $\tau(\cdot)$, and calculating the proportion p_c of simulated samples for which the test statistic is greater than the value calculated for the actual data. A value of p_c close to 0 or 1 indicates that the test statistic corresponds to a feature that is poorly fitted by the model. $\tau(\cdot)$ should thus be defined to reflect aspects of the model that are relevant to the purposes for which the inference is applied, and for extreme value models, Stephenson (2016) suggests defining $\tau(\cdot)$ on the basis of the maximum value. In this work, we thus use:

$$p_c = \frac{1}{r \cdot K \cdot \left(\sum_{s=1}^{S} D_s + U\right)} \times \sum_{j=1}^{r}\sum_{k=1}^{K}\left[\sum_{s=1}^{S}\sum_{i=1}^{D_s}\tau\left(\left[\check{\mathbf{y}}_{s,t,i}^{(k)}\right]_{(j)}, \left[\mathbf{y}_{s,t,i}\right]_{(j)}\right) + \sum_{u=S+1}^{S+U}\tau\left(\left[\check{\mathbf{y}}_{u,t}^{(\mathrm{D})(k)}\right]_{(j)}, \left[\mathbf{y}_{u,t}^{(\mathrm{D})}\right]_{(j)}\right)\right] \tag{S14}$$

with the following test statistic applied to vectors of precipitation maxima for T_s (respectively T_u) years:

$$\tau(\mathbf{x}, \mathbf{y}) = \begin{cases} 1 & \text{if } \max(\mathbf{x}) > \max(\mathbf{y}) \\ 0 & \text{otherwise.} \end{cases} \tag{S15}$$

S4.2 Posterior predictive coverage

An alternative approach to the task of posterior predictive checking is to consider how well the observed data $\mathbf{Y}$ fits under the estimated posterior predictive density. A measure of this fit is to provide the percentage of observed data points contained within the 95% probability interval (highest posterior density) of $p\big(\mathbf{Y}|\hat{\tilde{\boldsymbol{\mu}}}_+, \hat{\tilde{\boldsymbol{\sigma}}}_+^{(24)}, \hat{\boldsymbol{\xi}}_+, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}}\big)$, with $\hat{\boldsymbol{\chi}}$ representing the estimated posterior mean for parameter $\boldsymbol{\chi} \in \{\tilde{\boldsymbol{\mu}}_+, \tilde{\boldsymbol{\sigma}}_+^{(24)}, \boldsymbol{\xi}_+, \boldsymbol{\theta}, \boldsymbol{\eta}\}$:

$$\hat{\boldsymbol{\chi}} = \frac{1}{K}\sum_{k=1}^{K}\boldsymbol{\chi}^{(k)}. \tag{S16}$$

For pluviometer station $s \in \{1, \ldots, S\}$ at accumulation duration $d_{s,i}$, $i \in \{1, \ldots, D_s\}$, the predictive coverage score is calculated by determining how many of the data points $\mathbf{y}_{s,t,i}$, $t \in \{1, \ldots, T_s\}$, fall within the 95% probability region of $\mathrm{GEV}_r^*\big(\hat{\tilde{\mu}}_s, \hat{\tilde{\sigma}}_s^{(24)}, \hat{\xi}_s, \hat{\theta}_s, \hat{\eta}_s\big)$. For daily station $u \in \{S+1, \ldots, S+U\}$, the coverage of $\mathbf{y}_{u,t}^{(\mathrm{D})}$, $t \in \{1, \ldots, T_u\}$, is checked against $\mathrm{GEV}_r\big(\hat{\tilde{\mu}}_u, \hat{\tilde{\sigma}}_u^{(24)}, \hat{\xi}_u\big)$ for BHM-HF, or $\mathrm{GEV}_r\big(\hat{\tilde{\mu}}_u^{(\mathrm{D})}, \hat{\sigma}_u^{(\mathrm{D})}, \hat{\xi}_u^{(\mathrm{D})}\big)$ for BHM-EI. The contributions from all stations are then pooled to calculate a single overall percentage score for the posterior predictive coverage of observed data for a given model.

S4.3 Predictive accuracy score

A measure of predictive accuracy can be obtained by setting aside Q validation stations (randomly selected), and fitting each model to the rest of the precipitation dataset ("training" data, with $S + U - Q$ stations). In this work, we only select pluviometer stations to provide the validation data: $\check{\mathbf{y}}_{q,t,i}$, $q = 1, \ldots, Q$.

Once the model is fitted to the training dataset, spatial interpolation of the results is carried out through the BHM's process layer to infer the value of the GEV parameters:

$$\left\{\breve{\tilde{\mu}}_q^{(k)}, \breve{\tilde{\sigma}}_q^{(24)(k)}, \breve{\xi}_q^{(k)}, \breve{\theta}_q^{(k)}, \breve{\eta}_q^{(k)}\right\} \tag{S17}$$

at the locations of the validation stations $q = 1, \ldots, Q$, for each MCMC iteration $k = 1, \ldots, K$. Given these inferred parameters, a global measure of predictive accuracy can be obtained by averaging the accuracy "score" achieved for each individual validation data point (Stephenson, 2016):

$$\text{pred. acc.} = \frac{1}{K \cdot \sum_{q=1}^{Q} D_q T_q} \sum_{q=1}^{Q} \sum_{t=1}^{T_q} \sum_{i=1}^{D_q} \sum_{k=1}^{K} \text{GEV}_r^* \left(\breve{\mathbf{y}}_{q,t,i}; \breve{\tilde{\mu}}_q^{(k)}, \breve{\tilde{\sigma}}_q^{(24)(k)}, \breve{\xi}_q^{(k)}, \breve{\theta}_q^{(k)}, \breve{\eta}_q^{(k)}\right). \tag{S18}$$

This result provides a measure of the overall fit of the validation data, and can thus be used for model comparison purposes, with larger values indicating better models.

S4.4 Deviance information criterion (DIC)

Different models can also be compared by means of information criteria, which can be regarded as a simplified approach for the evaluation of predictive accuracy (Stephenson, 2016). One such measure is the deviance information criterion (Spiegelhalter *et al.*, 2002), which is defined as follows for the considered models:

$$\text{DIC} = 2 \cdot \log p\left(\mathbf{Y} \middle| \hat{\tilde{\boldsymbol{\mu}}}_+, \hat{\tilde{\boldsymbol{\sigma}}}_+^{(24)}, \hat{\boldsymbol{\xi}}_+, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}}\right) - \frac{4}{K} \sum_{k=1}^{K} \log p\left(\mathbf{Y} \middle| \tilde{\boldsymbol{\mu}}_+^{(k)}, \tilde{\boldsymbol{\sigma}}_+^{(24)(k)}, \boldsymbol{\xi}_+^{(k)}, \boldsymbol{\theta}^{(k)}, \boldsymbol{\eta}^{(k)}\right). \tag{S19}$$

The composite likelihood function, applicable to both terms in the above equation, is defined as follows:

$$p\left(\mathbf{Y} | \tilde{\boldsymbol{\mu}}_+, \tilde{\boldsymbol{\sigma}}_+^{(24)}, \boldsymbol{\xi}_+, \boldsymbol{\theta}, \boldsymbol{\eta}\right) = \prod_{s=1}^{S} \prod_{t=1}^{T_s} \prod_{i=1}^{D_s} p\left(\left[\mathbf{Y}^{(\text{P})}\right]_{(s,t,i)} \middle| \tilde{\boldsymbol{\mu}}_+, \tilde{\boldsymbol{\sigma}}_+^{(24)}, \boldsymbol{\xi}_+, \boldsymbol{\theta}, \boldsymbol{\eta}\right) \times \prod_{u=S+1}^{S+U} \prod_{t=1}^{T_u} p\left(\left[\mathbf{Y}^{(\text{D})}\right]_{(u,t)} \middle| \tilde{\boldsymbol{\mu}}_+, \tilde{\boldsymbol{\sigma}}_+^{(24)}, \boldsymbol{\xi}_+\right) \tag{S20}$$

where we have

$$p\left(\left[\mathbf{Y}^{(\text{P})}\right]_{(s,t,i)} \middle| \tilde{\boldsymbol{\mu}}_+, \tilde{\boldsymbol{\sigma}}_+^{(24)}, \boldsymbol{\xi}_+, \boldsymbol{\theta}, \boldsymbol{\eta}\right) = \text{GEV}_r^*\left(\mathbf{y}_{s,t,i}; \tilde{\mu}_s, \tilde{\sigma}_s^{(24)}, \xi_s, \theta_s, \eta_s\right) \tag{S21}$$

$$p\left(\left[\mathbf{Y}^{(\text{D})}\right]_{(u,t)} \middle| \tilde{\boldsymbol{\mu}}_+, \tilde{\boldsymbol{\sigma}}_+^{(24)}, \boldsymbol{\xi}_+\right) = \begin{cases} H^r \cdot \text{GEV}_r\left(H \cdot \mathbf{y}_{u,t}^{(\text{D})}; \tilde{\mu}_u, \tilde{\sigma}_u^{(24)}, \xi_u\right) & \text{for BHM-HF} \\ \text{GEV}_r\left(\mathbf{y}_{u,t}^{(\text{D})}; \tilde{\mu}_u^{(\text{D})}, \sigma_u^{(\text{D})}, \xi_u^{(\text{D})}\right) & \text{for BHM-EI.} \end{cases} \tag{S22}$$

The DIC parameter is defined on the deviance scale, with smaller DIC values indicating better models.

S4.5 Widely applicable information criterion (WAIC)

An alternative criterion for model comparison is the widely applicable information criterion (Watanabe, 2010). For the proposed models, this criterion is given by

$$\begin{aligned}\mathrm{WAIC} = &\sum_{s=1}^{S}\sum_{t=1}^{T_s}\sum_{i=1}^{D_s}\left\{2\cdot\log\left[\frac{1}{K}\sum_{k=1}^{K} p\left(\left[\mathbf{Y}^{(\mathrm{P})}\right]_{(s,t,i)}\middle|\tilde{\boldsymbol{\mu}}_+^{(k)},\tilde{\boldsymbol{\sigma}}_+^{(24)(k)},\boldsymbol{\xi}_+^{(k)},\boldsymbol{\theta}^{(k)},\boldsymbol{\eta}^{(k)}\right)\right] - \right.\\ &\left.\frac{4}{K}\sum_{k=1}^{K}\log p\left(\left[\mathbf{Y}^{(\mathrm{P})}\right]_{(s,t,i)}\middle|\tilde{\boldsymbol{\mu}}_+^{(k)},\tilde{\boldsymbol{\sigma}}_+^{(24)(k)},\boldsymbol{\xi}_+^{(k)},\boldsymbol{\theta}^{(k)},\boldsymbol{\eta}^{(k)}\right)\right\} + \\ &\sum_{u=S+1}^{S+U}\sum_{t=1}^{T_u}\left\{2\cdot\log\left[\frac{1}{K}\sum_{k=1}^{K} p\left(\left[\mathbf{Y}^{(\mathrm{D})}\right]_{(u,t)}\middle|\tilde{\boldsymbol{\mu}}_+^{(k)},\tilde{\boldsymbol{\sigma}}_+^{(24)(k)},\boldsymbol{\xi}_+^{(k)}\right)\right] - \right.\\ &\left.\frac{4}{K}\sum_{k=1}^{K}\log p\left(\left[\mathbf{Y}^{(\mathrm{D})}\right]_{(u,t)}\middle|\tilde{\boldsymbol{\mu}}_+^{(k)},\tilde{\boldsymbol{\sigma}}_+^{(24)(k)},\boldsymbol{\xi}_+^{(k)}\right)\right\} \end{aligned} \tag{S23}$$

using the likelihood definitions given in Eqs. (S21) and (S22).

S5 Full conditionals

This section provides an overview of the full conditional distributions derived for MCMC implementation of Eq. (S3). Unless otherwise mentioned, standard Gibbs sampling is used. Otherwise, derivation of the acceptance probability is also provided for those variables requiring a Metropolis–Hastings (MH) step within the Gibbs sampler.

The derivations presented below are based on the model of Sec. 3, which assumes a single dataset of S pluviometer stations. Extension of this framework to account for an additional dataset of daily measurements (as described in Sec. 4), using either the Hershfield factor or extremal index approach, is straightforward.

S5.1 GEV parameters $\tilde{\boldsymbol{\mu}}$ and $\boldsymbol{\xi}$

The process model for $\boldsymbol{\xi}$ is similar to that of $\tilde{\boldsymbol{\mu}}$ (see Eq. (6) in Sec. 3.2); the same derivations can thus be used to determine the respective full conditional distribution. Let us define the generic variable $\chi \in \{\tilde{\mu}, \xi\}$ and the related complementary variable:

$$\bar{\chi} \triangleq \begin{cases} \xi & \text{if } \chi \triangleq \tilde{\mu} \\ \tilde{\mu} & \text{if } \chi \triangleq \xi. \end{cases} \tag{S24}$$

With the spatial process definition given as follows (see Sec. 3.2):

$$\mathcal{P}_\chi(\boldsymbol{\ell}, \alpha_\chi, \lambda_\chi) \sim \mathrm{MVN}\big(\mathbf{0}, \boldsymbol{\Sigma}_\chi(\boldsymbol{\ell}, \alpha_\chi, \lambda_\chi)\big), \tag{S25}$$

the full conditional distribution for the model parameter $\boldsymbol{\chi}$ results from Eqs. (S3), (5) and (6) as:

$$p(\boldsymbol{\chi}|\mathrm{rest}) \propto p(\mathbf{Y}|\tilde{\boldsymbol{\mu}}, \boldsymbol{\sigma}, \boldsymbol{\xi}, \boldsymbol{\theta}, \boldsymbol{\eta}) \cdot p(\boldsymbol{\chi}|\boldsymbol{\beta}_\chi, \alpha_\chi, \lambda_\chi) \tag{S26}$$

$$= \prod_{s=1}^{S}\prod_{t=1}^{T_s}\prod_{i=1}^{D_s} \mathrm{GEV}_r^*(\mathbf{y}_{s,t,i}; \tilde{\mu}_s, \sigma_s, \xi_s, \theta_s, \eta_s) \cdot \mathrm{MVN}\big(\boldsymbol{\chi}; \mathbf{X}_\chi\boldsymbol{\beta}_\chi, \boldsymbol{\Sigma}_\chi(\boldsymbol{\ell}, \alpha_\chi, \lambda_\chi)\big). \tag{S27}$$

As Eq. (S27) does not represent a standard distribution, a MH step is used to update each component of $\boldsymbol{\chi}$ in turn during the MCMC sampling. At the k-th MCMC iteration, a proposal sample is generated for the n-th element of $\boldsymbol{\chi}$, $n \in \{1, \ldots, S\}$, using a symmetric random walk process:

$$\chi_n^* \sim \mathrm{N}\left(\chi_n^{(k-1)}, \sigma^2_{\mathrm{prop},\chi}\right) \tag{S28}$$

where $\chi_n^{(k-1)}$ represents the n-th element of $\boldsymbol{\chi}$ at the previous MCMC iteration, and with $\sigma_{\mathrm{prop},\chi}$ representing the "step size" of the proposal distribution. From the current state $\boldsymbol{\chi}'$ of the $\boldsymbol{\chi}$ vector:

$$\boldsymbol{\chi}' = \left[\chi_1^{(k)}, \ldots, \chi_{n-1}^{(k)}, \chi_n^{(k-1)}, \chi_{n+1}^{(k-1)}, \ldots, \chi_S^{(k-1)}\right]^T \tag{S29}$$

the proposal $\boldsymbol{\chi}$ vector thus results as:

$$\boldsymbol{\chi}^* = \left[\chi_1^{(k)}, \ldots, \chi_{n-1}^{(k)}, \chi_n^*, \chi_{n+1}^{(k-1)}, \ldots, \chi_S^{(k-1)}\right]^T. \tag{S30}$$

The proposal sample χ_n^* is subsequently accepted or rejected according to:

$$\chi_n^{(k)} = \begin{cases} \chi_n^{(k-1)} & \text{with probability } 1 - P_{\mathrm{acc},\chi} \\ \chi_n^* & \text{with probability } P_{\mathrm{acc},\chi} \end{cases} \tag{S31}$$

where the acceptance probability is defined as:

$$P_{\mathrm{acc},\chi} = \min\left\{1, \frac{p(\mathbf{Y}|\boldsymbol{\chi}^*, \bar{\boldsymbol{\chi}}, \boldsymbol{\sigma}, \boldsymbol{\theta}, \boldsymbol{\eta})}{p(\mathbf{Y}|\boldsymbol{\chi}', \bar{\boldsymbol{\chi}}, \boldsymbol{\sigma}, \boldsymbol{\theta}, \boldsymbol{\eta})} \cdot \frac{p(\boldsymbol{\chi}^*|\boldsymbol{\beta}_\chi, \alpha_\chi, \lambda_\chi)}{p(\boldsymbol{\chi}'|\boldsymbol{\beta}_\chi, \alpha_\chi, \lambda_\chi)}\right\} \tag{S32}$$

$$= \min\left\{1, \underbrace{\frac{\prod_s \prod_t \prod_i \mathrm{GEV}_r^*\left(\mathbf{y}_{s,t,i}; [\boldsymbol{\chi}^*]_{(s)}, \bar{\chi}_s, \sigma_s, \theta_s, \eta_s\right)}{\prod_s \prod_t \prod_i \mathrm{GEV}_r^*\left(\mathbf{y}_{s,t,i}; [\boldsymbol{\chi}']_{(s)}, \bar{\chi}_s, \sigma_s, \theta_s, \eta_s\right)}}_{A_0} \cdot \underbrace{\frac{\mathrm{MVN}(\boldsymbol{\chi}^*; \mathbf{X}_\chi \boldsymbol{\beta}_\chi, \boldsymbol{\Sigma}_\chi)}{\mathrm{MVN}(\boldsymbol{\chi}'; \mathbf{X}_\chi \boldsymbol{\beta}_\chi, \boldsymbol{\Sigma}_\chi)}}_{A_1}\right\}. \tag{S33}$$

Since all but one element of the vectors $\boldsymbol{\chi}^*$ and $\boldsymbol{\chi}'$ are identical, the term A_0 in Eq. (S33) can be simplified through some simple algebra:

$$A_0 = \frac{\prod_{t=1}^{T_s} \prod_{i=1}^{D_s} \mathrm{GEV}_r^*\left(\mathbf{y}_{n,t,i}; \chi_n^*, \bar{\chi}_n, \sigma_n, \theta_n, \eta_n\right)}{\prod_{t=1}^{T_s} \prod_{i=1}^{D_s} \mathrm{GEV}_r^*\left(\mathbf{y}_{n,t,i}; \chi_n^{(k-1)}, \bar{\chi}_n, \sigma_n, \theta_n, \eta_n\right)}. \tag{S34}$$

Due to $\boldsymbol{\Sigma}_\chi$ being symmetric, the second fraction A_1 can also be simplified further (leading to a substantial reduction in computational requirements):

$$A_1 = \exp\left\{\left(\mathbf{X}_\chi \boldsymbol{\beta}_\chi - \boldsymbol{\chi}''\right)^T \cdot \left[\boldsymbol{\Sigma}_\chi^{-1}\right]_{(\cdot,n)} \cdot \left(\chi_n^* - \chi_n^{(k-1)}\right)\right\} \tag{S35}$$

where $[\boldsymbol{\Sigma}]_{(\cdot,n)}$ extracts the n-th column of the matrix $\boldsymbol{\Sigma}$, and with

$$\boldsymbol{\chi}'' = \left[\chi_1^{(k)}, \ldots, \chi_{n-1}^{(k)}, \frac{1}{2} \cdot \left(\chi_n^* + \chi_n^{(k-1)}\right), \chi_{n+1}^{(k-1)}, \ldots, \chi_S^{(k-1)}\right]^T. \tag{S36}$$

S5.2 GEV parameters $\boldsymbol{\sigma}$ and $\boldsymbol{\theta}$

Similar definitions of the process model for $\boldsymbol{\sigma}$ and $\boldsymbol{\theta}$ in Sec. 3.2 allow for the same full conditionals derivations for both parameters. For $\chi \in \{\sigma, \theta\}$ (with $\bar{\chi}$ defined similarly to Eq. (S24)), the full conditional distribution for $\boldsymbol{\chi}$ results from Eq. (S3) as:

$$p(\boldsymbol{\chi}|\mathrm{rest}) \propto p(\mathbf{Y}|\tilde{\boldsymbol{\mu}}, \boldsymbol{\sigma}, \boldsymbol{\xi}, \boldsymbol{\theta}, \boldsymbol{\eta}) \cdot p(\boldsymbol{\chi}|\boldsymbol{\beta}_\chi, \alpha_\chi, \lambda_\chi). \tag{S37}$$

From Eqs. (6) and (S25), we have:

$$\log(\boldsymbol{\chi}) \sim \text{MVN}\big(\mathbf{X}_\chi \boldsymbol{\beta}_\chi, \boldsymbol{\Sigma}_\chi(\boldsymbol{\ell}, \alpha_\chi, \lambda_\chi)\big) \tag{S38}$$

from which $p(\boldsymbol{\chi}|\boldsymbol{\beta}_\chi, \alpha_\chi, \lambda_\chi)$ must be derived.

Lemma S1. Given a collection of random variables $(X_1, \ldots, X_N)$ with joint density function $p_{(X_1,\ldots,X_N)}(\cdot)$, it can be shown that the probability density function of the transformed set of variables $(Z_1, \ldots, Z_N) = (f(X_1), \ldots, f(X_N))$ is given by:

$$p_{(Z_1,\ldots,Z_N)}(z_1, \ldots, z_N) = p_{(X_1,\ldots,X_N)}\left(f^{-1}(z_1), \ldots, f^{-1}(z_N)\right) \cdot |J(z_1, \ldots, z_N)| \tag{S39}$$

where $J(z_1, \ldots, z_N)$ represents the Jacobian of the transformation.

Defining $f(\cdot) \triangleq \exp(\cdot)$ and $z_i \triangleq \chi_i$ in conjunction with Eq. (S39), the desired density can thus be derived from Eq. (S38) as:

$$p(\boldsymbol{\chi}|\boldsymbol{\beta}_\chi, \alpha_\chi, \lambda_\chi) = \text{MVN}\big(\log(\boldsymbol{\chi}); \mathbf{X}_\chi \boldsymbol{\beta}_\chi, \boldsymbol{\Sigma}_\chi\big) \cdot \left(\prod_{s=1}^{S} \chi_s\right)^{-1}. \tag{S40}$$

Hence, Eq. (S37) becomes:

$$p(\boldsymbol{\chi}|\text{rest}) \propto \prod_{s=1}^{S}\prod_{t=1}^{T_s}\prod_{i=1}^{D_s} \text{GEV}_r^*(\mathbf{y}_{s,t,i}; \tilde{\mu}_s, \sigma_s, \xi_s, \theta_s, \eta_s) \cdot \text{MVN}\big(\log(\boldsymbol{\chi}); \mathbf{X}_\chi \boldsymbol{\beta}_\chi, \boldsymbol{\Sigma}_\chi\big) \cdot \left(\prod_{s=1}^{S} \chi_s\right)^{-1}. \tag{S41}$$

A MH step is used to sample from the non-standard density in Eq. (S41). So as to improve the MCMC sampling convergence, we generate proposal samples $\chi_n^* > 0$ for the n-th component of $\boldsymbol{\chi}$ using a log-normal distribution:

$$p\left(\chi_n^* | \chi_n^{(k-1)}, \sigma^2_{\text{prop},\chi}\right) = \text{logN}\left(\chi_n^*; \log\left(\chi_n^{(k-1)}\right), \sigma^2_{\text{prop},\chi}\right). \tag{S42}$$

The proposal sample is then accepted with probability $P_{\text{acc},\chi}$ defined as:

$$P_{\text{acc},\chi} = \min\left\{1, \underbrace{\frac{p(\mathbf{Y}|\boldsymbol{\chi}^*, \bar{\boldsymbol{\chi}}, \tilde{\boldsymbol{\mu}}, \boldsymbol{\xi}, \boldsymbol{\eta})}{p(\mathbf{Y}|\boldsymbol{\chi}', \bar{\boldsymbol{\chi}}, \tilde{\boldsymbol{\mu}}, \boldsymbol{\xi}, \boldsymbol{\eta})}}_{B_0} \cdot \underbrace{\frac{p(\boldsymbol{\chi}^*|\boldsymbol{\beta}_\chi, \alpha_\chi, \lambda_\chi)}{p(\boldsymbol{\chi}'|\boldsymbol{\beta}_\chi, \alpha_\chi, \lambda_\chi)}}_{B_1} \cdot \underbrace{\frac{p\left(\chi_n^{(k-1)}|\chi_n^*, \sigma^2_{\text{prop},\chi}\right)}{p\left(\chi_n^*|\chi_n^{(k-1)}, \sigma^2_{\text{prop},\chi}\right)}}_{B_2}\right\} \tag{S43}$$

with $\boldsymbol{\chi}'$ and $\boldsymbol{\chi}^*$ defined similarly to Eqs. (S29) and (S30), respectively. As per Eq. (S34), the term B_0 becomes:

$$B_0 = \frac{\prod_{t=1}^{T_s}\prod_{i=1}^{D_s} \text{GEV}_r^*\left(\mathbf{y}_{n,t,i}; \chi_n^*, \bar{\chi}_n, \tilde{\mu}_n, \xi_n, \eta_n\right)}{\prod_{t=1}^{T_s}\prod_{i=1}^{D_s} \text{GEV}_r^*\left(\mathbf{y}_{n,t,i}; \chi_n^{(k-1)}, \bar{\chi}_n, \tilde{\mu}_n, \xi_n, \eta_n\right)}. \tag{S44}$$

Analogous to Eq. (S35), and using Eq. (S40), the term B_1 simplifies to:

$$B_1 = \underbrace{\exp\left\{\left(\mathbf{X}_\chi \boldsymbol{\beta}_\chi - \log(\boldsymbol{\chi}'')\right)^T \cdot \left[\boldsymbol{\Sigma}_\chi^{-1}\right]_{(\cdot,n)} \cdot \log\left(\frac{\chi_n^*}{\chi_n^{(k-1)}}\right)\right\}}_{B_3} \cdot \frac{\chi_n^{(k-1)}}{\chi_n^*} \tag{S45}$$

with

$$\boldsymbol{\chi}'' = \left[\chi_1^{(k)}, \ldots, \chi_{n-1}^{(k)}, \sqrt{\chi_n^* \cdot \chi_n^{(k-1)}}, \chi_{n+1}^{(k-1)}, \ldots, \chi_S^{(k-1)}\right]^T. \tag{S46}$$

The term B_2 in Eq. (S43) is introduced to compensate for the use of a non-symmetric proposal distribution. Based on Eq. (S42) and following some algebraic manipulations, this term can be shown to simplify to:

$$B_2 = \frac{\chi_n^*}{\chi_n^{(k-1)}} \tag{S47}$$

which cancels out the last term on the right-hand side of Eq. (S45). Consequently, Eq. (S43) becomes:

$$P_{\text{acc},\chi} = \min\{1, B_0 \cdot B_3\}. \tag{S48}$$

S5.3 GEV parameter $\boldsymbol{\eta}$

The full conditional density for $\boldsymbol{\eta}$ can be determined following derivations similar to those given in Sec. S5.2. From Eq. (S3), we have:

$$p(\boldsymbol{\eta}|\text{rest}) \propto p(\mathbf{Y}|\tilde{\boldsymbol{\mu}}, \boldsymbol{\sigma}, \boldsymbol{\xi}, \boldsymbol{\theta}, \boldsymbol{\eta}) \cdot p(\boldsymbol{\eta}|\boldsymbol{\beta}_\eta, \alpha_\eta, \lambda_\eta) \tag{S49}$$

with Eqs. (6) and (S25) leading to:

$$\text{logit}(\boldsymbol{\eta}) \sim \text{MVN}\big(\mathbf{X}_\eta \boldsymbol{\beta}_\eta, \boldsymbol{\Sigma}_\eta(\boldsymbol{\ell}, \alpha_\eta, \lambda_\eta)\big). \tag{S50}$$

Using Eq. (S39) on the transformed variable $\text{logit}(\boldsymbol{\eta})$, the density of $\boldsymbol{\eta}$ can be shown to be:

$$p(\boldsymbol{\eta}|\boldsymbol{\beta}_\eta, \alpha_\eta, \lambda_\eta) = \text{MVN}\big(\text{logit}(\boldsymbol{\eta}); \mathbf{X}_\eta \boldsymbol{\beta}_\eta, \boldsymbol{\Sigma}_\eta\big) \cdot \left(\prod_{s=1}^{S} \eta_s \cdot (1-\eta_s)\right)^{-1}. \tag{S51}$$

MH sampling for the n-th component of $\boldsymbol{\eta}$ is implemented using a logit-normal distribution to generate proposals $\eta_n^* \in [0,1]$:

$$p\left(\eta_n^* | \eta_n^{(k-1)}, \sigma^2_{\text{prop},\eta}\right) = \text{logitN}\left(\eta_n^*; \text{logit}\left(\eta_n^{(k-1)}\right), \sigma^2_{\text{prop},\eta}\right) \tag{S52}$$

leading to the acceptance probability:

$$P_{\text{acc},\eta} = \min\left\{1, \underbrace{\frac{p(\mathbf{Y}|\tilde{\boldsymbol{\mu}}, \boldsymbol{\sigma}, \boldsymbol{\xi}, \boldsymbol{\theta}, \boldsymbol{\eta}^*)}{p(\mathbf{Y}|\tilde{\boldsymbol{\mu}}, \boldsymbol{\sigma}, \boldsymbol{\xi}, \boldsymbol{\theta}, \boldsymbol{\eta}')}}_{C_0} \cdot \underbrace{\frac{p(\boldsymbol{\eta}^*|\boldsymbol{\beta}_\eta, \alpha_\eta, \lambda_\eta)}{p(\boldsymbol{\eta}'|\boldsymbol{\beta}_\eta, \alpha_\eta, \lambda_\eta)}}_{C_1} \cdot \underbrace{\frac{p\left(\eta_n^{(k-1)}|\eta_n^*, \sigma^2_{\text{prop},\eta}\right)}{p\left(\eta_n^*|\eta_n^{(k-1)}, \sigma^2_{\text{prop},\eta}\right)}}_{C_2}\right\} \tag{S53}$$

with $\boldsymbol{\eta}'$ and $\boldsymbol{\eta}^*$ defined similarly to Eqs. (S29) and (S30), respectively. As per Eq. (S34), the term C_0 becomes:

$$C_0 = \frac{\prod_{t=1}^{T_s}\prod_{i=1}^{D_s} \text{GEV}_r^*\left(\mathbf{y}_{n,t,i}; \tilde{\mu}_n, \sigma_n, \xi_n, \theta_n, \eta_n^*\right)}{\prod_{t=1}^{T_s}\prod_{i=1}^{D_s} \text{GEV}_r^*\left(\mathbf{y}_{n,t,i}; \tilde{\mu}_n, \sigma_n, \xi_n, \theta_n, \eta_n^{(k-1)}\right)} \tag{S54}$$

and analogous to Eq. (S35) with Eq. (S51), the term C_1 simplifies to:

$$C_1 = \underbrace{\exp\left\{\left(\mathbf{X}_\eta \boldsymbol{\beta}_\eta - \text{logit}(\boldsymbol{\eta}'')\right)^T \cdot \left[\boldsymbol{\Sigma}_\eta^{-1}\right]_{(\cdot,n)} \cdot \left(\text{logit}(\eta_n^*) - \text{logit}\left(\eta_n^{(k-1)}\right)\right)\right\}}_{C_3} \times \frac{\eta_n^{(k-1)} \cdot \left(1 - \eta_n^{(k-1)}\right)}{\eta_n^* \cdot (1-\eta_n^*)} \tag{S55}$$

with

$$\boldsymbol{\eta}'' = \left[\eta_1^{(k)}, \ldots, \eta_{n-1}^{(k)}, \eta_n'', \eta_{n+1}^{(k-1)}, \ldots, \eta_S^{(k-1)}\right]^T \tag{S56}$$

$$\eta_n'' = \left(1 + \sqrt{\frac{(1-\eta_n^*)\left(1-\eta_n^{(k-1)}\right)}{\eta_n^* \cdot \eta_n^{(k-1)}}}\right)^{-1}. \tag{S57}$$

Using Eq. (S52), the term C_2 can be shown to simplify as follows:

$$C_2 = \frac{\eta_n^* \cdot (1-\eta_n^*)}{\eta_n^{(k-1)} \cdot \left(1-\eta_n^{(k-1)}\right)} \tag{S58}$$

resulting in:

$$P_{\text{acc},\eta} = \min\{1, C_0 \cdot C_3\}. \tag{S59}$$

S5.4 Regression parameters $\boldsymbol{\beta}_{\tilde{\mu}}$ and $\boldsymbol{\beta}_{\xi}$

For $\chi \in \{\tilde{\mu}, \xi\}$, the full conditional distribution of the respective regression coefficient results from Eqs. (S3), (S7) and (S25) as well as Eq. (6):

$$\begin{aligned} p(\boldsymbol{\beta}_\chi|\text{rest}) &\propto p(\boldsymbol{\chi}|\boldsymbol{\beta}_\chi, \alpha_\chi, \lambda_\chi) \cdot p(\boldsymbol{\beta}_\chi) && \text{(S60)}\\ &= \text{MVN}\left(\boldsymbol{\chi}; \mathbf{X}_\chi\boldsymbol{\beta}_\chi, \boldsymbol{\Sigma}_\chi(\boldsymbol{\ell}, \alpha_\chi, \lambda_\chi)\right) \cdot \text{MVN}\left(\boldsymbol{\beta}_\chi; \boldsymbol{\mu}_{\beta_\chi}, \boldsymbol{\Sigma}_{\beta_\chi}\right) && \text{(S61)}\\ &= \text{MVN}\left(\boldsymbol{\beta}_\chi; \mathbf{m}_\chi, \mathbf{S}_\chi\right) && \text{(S62)} \end{aligned}$$

where Eq. (S62) results from conjugacy with:

$$\begin{aligned} \mathbf{m}_\chi &= \mathbf{S}_\chi \cdot \left(\boldsymbol{\Sigma}_{\beta_\chi}^{-1}\boldsymbol{\mu}_{\beta_\chi} + \mathbf{X}_\chi^T\boldsymbol{\Sigma}_\chi^{-1}\boldsymbol{\chi}\right) && \text{(S63)}\\ \mathbf{S}_\chi &= \left(\boldsymbol{\Sigma}_{\beta_\chi}^{-1} + \mathbf{X}_\chi^T\boldsymbol{\Sigma}_\chi^{-1}\mathbf{X}_\chi\right)^{-1}. && \text{(S64)} \end{aligned}$$

S5.5 Regression parameters $\boldsymbol{\beta}_{\sigma}$ and $\boldsymbol{\beta}_{\theta}$

Using $\chi \in \{\sigma, \theta\}$, the full conditional distribution of the respective regression coefficient can be derived from Eqs. (S3), (S7) and (S40) as:

$$\begin{aligned} p(\boldsymbol{\beta}_\chi|\text{rest}) &\propto p(\boldsymbol{\chi}|\boldsymbol{\beta}_\chi, \alpha_\chi, \lambda_\chi) \cdot p(\boldsymbol{\beta}_\chi) && \text{(S65)}\\ &\propto \text{MVN}\left(\log(\boldsymbol{\chi}); \mathbf{X}_\chi\boldsymbol{\beta}_\chi, \boldsymbol{\Sigma}_\chi(\boldsymbol{\ell}, \alpha_\chi, \lambda_\chi)\right) \cdot \text{MVN}\left(\boldsymbol{\beta}_\chi; \boldsymbol{\mu}_{\beta_\chi}, \boldsymbol{\Sigma}_{\beta_\chi}\right) && \text{(S66)}\\ &= \text{MVN}\left(\boldsymbol{\beta}_\chi; \mathbf{m}_\chi, \mathbf{S}_\chi\right) && \text{(S67)} \end{aligned}$$

where Eq. (S67) results from conjugacy with:

$$\begin{aligned} \mathbf{m}_\chi &= \mathbf{S}_\chi \cdot \left(\boldsymbol{\Sigma}_{\beta_\chi}^{-1}\boldsymbol{\mu}_{\beta_\chi} + \mathbf{X}_\chi^T\boldsymbol{\Sigma}_\chi^{-1}\log(\boldsymbol{\chi})\right) && \text{(S68)}\\ \mathbf{S}_\chi &= \left(\boldsymbol{\Sigma}_{\beta_\chi}^{-1} + \mathbf{X}_\chi^T\boldsymbol{\Sigma}_\chi^{-1}\mathbf{X}_\chi\right)^{-1}. && \text{(S69)} \end{aligned}$$

S5.6 Regression parameter $\boldsymbol{\beta}_\eta$

The full conditional distribution for $\boldsymbol{\beta}_\eta$ can be derived from Eqs. (S3), (S7) and (S51) as:

$$p(\boldsymbol{\beta}_\eta|\text{rest}) \propto p(\boldsymbol{\eta}|\boldsymbol{\beta}_\eta, \alpha_\eta, \lambda_\eta) \cdot p(\boldsymbol{\beta}_\eta) \tag{S70}$$

$$\propto \text{MVN}\left(\text{logit}(\boldsymbol{\eta}); \mathbf{X}_\eta\boldsymbol{\beta}_\eta, \boldsymbol{\Sigma}_\eta(\boldsymbol{\ell}, \alpha_\eta, \lambda_\eta)\right) \cdot \text{MVN}\left(\boldsymbol{\beta}_\eta; \boldsymbol{\mu}_{\beta_\eta}, \boldsymbol{\Sigma}_{\beta_\eta}\right) \tag{S71}$$

$$= \text{MVN}\left(\boldsymbol{\beta}_\eta; \mathbf{m}_\eta, \mathbf{S}_\eta\right) \tag{S72}$$

where Eq. (S72) results from conjugacy with:

$$\mathbf{m}_\eta = \mathbf{S}_\eta \cdot \left(\boldsymbol{\Sigma}_{\beta_\eta}^{-1}\boldsymbol{\mu}_{\beta_\eta} + \mathbf{X}_\eta^T\boldsymbol{\Sigma}_\eta^{-1}\,\text{logit}(\boldsymbol{\eta})\right) \tag{S73}$$

$$\mathbf{S}_\eta = \left(\boldsymbol{\Sigma}_{\beta_\eta}^{-1} + \mathbf{X}_\eta^T\boldsymbol{\Sigma}_\eta^{-1}\mathbf{X}_\eta\right)^{-1}. \tag{S74}$$

S5.7 Sill parameters $\alpha_{\tilde{\mu}}$ and α_ξ

Using $\chi \in \{\tilde{\mu}, \xi\}$, the full conditional distribution of the respective sill parameter results from Eqs. (S3), (S6), (S25) and (6) in:

$$p(\alpha_\chi|\text{rest}) \propto p(\boldsymbol{\chi}|\boldsymbol{\beta}_\chi, \alpha_\chi, \lambda_\chi) \cdot p(\alpha_\chi) \tag{S75}$$

$$= \text{MVN}\left(\boldsymbol{\chi}; \mathbf{X}_\chi\boldsymbol{\beta}_\chi, \boldsymbol{\Sigma}_\chi(\boldsymbol{\ell}, \alpha_\chi, \lambda_\chi)\right) \cdot \text{InvGamma}\left(\alpha_\chi; \kappa_{\alpha_\chi}, \gamma_{\alpha_\chi}\right). \tag{S76}$$

Given the definition of correlation function used in Eq. (7) (exponential correlation), the matrix $\boldsymbol{\Sigma}_\phi(\boldsymbol{\ell}, \alpha_\phi, \lambda_\phi)$ can be decomposed as follows, for $\phi \in \{\tilde{\mu}, \sigma, \xi, \theta, \eta\}$:

$$\boldsymbol{\Sigma}_\phi(\boldsymbol{\ell}, \alpha_\phi, \lambda_\phi) = \alpha_\phi \cdot \tilde{\boldsymbol{\Sigma}}_\phi(\boldsymbol{\ell}, \lambda_\phi) \tag{S77}$$

$$[\tilde{\boldsymbol{\Sigma}}_\phi]_{(i,j)} = \exp\left(\frac{\|\boldsymbol{\ell}_i - \boldsymbol{\ell}_j\|}{\lambda_\phi}\right), \quad i, j = 1, \ldots, S. \tag{S78}$$

Consequently, Eq. (S76) results from conjugacy as:

$$p(\alpha_\chi|\text{rest}) \propto \text{InvGamma}\left(\alpha_\chi; K_\chi, G_\chi\right) \tag{S79}$$

$$K_\chi = \kappa_{\alpha_\chi} + \frac{S}{2} \tag{S80}$$

$$G_\chi = \frac{1}{2} \cdot \left(\boldsymbol{\chi} - \mathbf{X}_\chi\boldsymbol{\beta}_\chi\right)^T \cdot \tilde{\boldsymbol{\Sigma}}_\chi^{-1} \cdot \left(\boldsymbol{\chi} - \mathbf{X}_\chi\boldsymbol{\beta}_\chi\right) + \gamma_{\alpha_\chi}. \tag{S81}$$

S5.8 Sill parameters α_σ and α_θ

For $\chi \in \{\sigma, \theta\}$, the full conditional density of the respective sill parameter can be derived from Eqs. (S3), (S6) and (S40) as:

$$p(\alpha_\chi|\text{rest}) \propto p(\boldsymbol{\chi}|\boldsymbol{\beta}_\chi, \alpha_\chi, \lambda_\chi) \cdot p(\alpha_\chi) \tag{S82}$$

$$\propto \text{MVN}\left(\log(\boldsymbol{\chi}); \mathbf{X}_\chi\boldsymbol{\beta}_\chi, \boldsymbol{\Sigma}_\chi(\boldsymbol{\ell}, \alpha_\chi, \lambda_\chi)\right) \cdot \text{InvGamma}\left(\alpha_\chi; \kappa_{\alpha_\chi}, \gamma_{\alpha_\chi}\right) \tag{S83}$$

$$= \text{InvGamma}\left(\alpha_\chi; K_\chi, G_\chi\right) \tag{S84}$$

where Eq. (S84) results from conjugacy and Eq. (S77), with:

$$K_\chi = \kappa_{\alpha_\chi} + \frac{S}{2} \tag{S85}$$

$$G_\chi = \frac{1}{2} \cdot \left(\log(\boldsymbol{\chi}) - \mathbf{X}_\chi\boldsymbol{\beta}_\chi\right)^T \cdot \tilde{\boldsymbol{\Sigma}}_\chi^{-1} \cdot \left(\log(\boldsymbol{\chi}) - \mathbf{X}_\chi\boldsymbol{\beta}_\chi\right) + \gamma_{\alpha_\chi}. \tag{S86}$$

S5.9 Sill parameter α_η

The full conditional density for α_η can be derived from Eqs. (S3), (S6) and (S51) as:

$$p(\alpha_\eta|\text{rest}) \propto p(\boldsymbol{\eta}|\boldsymbol{\beta}_\eta, \alpha_\eta, \lambda_\eta) \cdot p(\alpha_\eta) \tag{S87}$$

$$\propto \text{MVN}\left(\text{logit}(\boldsymbol{\eta}); \mathbf{X}_\eta\boldsymbol{\beta}_\eta, \boldsymbol{\Sigma}_\eta(\boldsymbol{\ell}, \alpha_\eta, \lambda_\eta)\right) \cdot \text{InvGamma}\left(\alpha_\eta; \kappa_{\alpha_\eta}, \gamma_{\alpha_\eta}\right) \tag{S88}$$

$$= \text{InvGamma}\left(\alpha_\eta; K_\eta, G_\eta\right) \tag{S89}$$

where Eq. (S89) results from conjugacy and Eq. (S77), with:

$$K_\eta = \kappa_{\alpha_\eta} + \frac{S}{2} \tag{S90}$$

$$G_\eta = \frac{1}{2} \cdot \left(\text{logit}(\boldsymbol{\eta}) - \mathbf{X}_\eta\boldsymbol{\beta}_\eta\right)^T \cdot \tilde{\boldsymbol{\Sigma}}_\eta^{-1} \cdot \left(\text{logit}(\boldsymbol{\eta}) - \mathbf{X}_\eta\boldsymbol{\beta}_\eta\right) + \gamma_{\alpha_\eta}. \tag{S91}$$

S5.10 Range parameters $\lambda_{\tilde{\mu}}$ and λ_ξ

For $\chi \in \{\tilde{\mu}, \xi\}$, the full conditional distribution of the respective range parameter results from Eqs. (S3), (S5), (S25) and (6) as:

$$p(\lambda_\chi|\text{rest}) \propto p(\boldsymbol{\chi}|\boldsymbol{\beta}_\chi, \alpha_\chi, \lambda_\chi) \cdot p(\lambda_\chi) \tag{S92}$$

$$= \text{MVN}\left(\boldsymbol{\chi}; \mathbf{X}_\chi\boldsymbol{\beta}_\chi, \boldsymbol{\Sigma}_\chi(\boldsymbol{\ell}, \alpha_\chi, \lambda_\chi)\right) \cdot \text{Gamma}\left(\lambda_\chi; \kappa_{\lambda_\chi}, \gamma_{\lambda_\chi}\right) \tag{S93}$$

which represents a non-standard distribution (due to λ_χ being "embedded" within $\boldsymbol{\Sigma}_\chi$), and thus requires a MH sampling step. A non-symmetric proposal distribution is used to generate proposal samples $\lambda_\chi^* > 0$ at the k-th MCMC iteration:

$$p\left(\lambda_\chi^* | \lambda_\chi^{(k-1)}, \sigma^2_{\text{prop},\lambda_\chi}\right) = \text{logN}\left(\lambda_\chi^*; \log\left(\lambda_\chi^{(k-1)}\right), \sigma^2_{\text{prop},\lambda_\chi}\right) \tag{S94}$$

where $\lambda_\chi^{(k-1)}$ is the value of the range parameter at the previous iteration in the MCMC chain. The acceptance probability for the proposal sample is:

$$P_{\text{acc},\lambda_\chi} = \min\left\{1, \frac{\text{MVN}\left(\boldsymbol{\chi}; \mathbf{X}_\chi\boldsymbol{\beta}_\chi, \boldsymbol{\Sigma}_\chi(\boldsymbol{\ell}, \alpha_\chi, \lambda_\chi^*)\right)}{\text{MVN}\left(\boldsymbol{\chi}; \mathbf{X}_\chi\boldsymbol{\beta}_\chi, \boldsymbol{\Sigma}_\chi(\boldsymbol{\ell}, \alpha_\chi, \lambda_\chi^{(k-1)})\right)} \times \underbrace{\frac{\text{Gamma}\left(\lambda_\chi^*; \kappa_{\lambda_\chi}, \gamma_{\lambda_\chi}\right)}{\text{Gamma}\left(\lambda_\chi^{(k-1)}; \kappa_{\lambda_\chi}, \gamma_{\lambda_\chi}\right)} \cdot \frac{p\left(\lambda_\chi^{(k-1)} | \lambda_\chi^*, \sigma^2_{\text{prop},\lambda_\chi}\right)}{p\left(\lambda_\chi^* | \lambda_\chi^{(k-1)}, \sigma^2_{\text{prop},\lambda_\chi}\right)}}_{D_0}\right\} \tag{S95}$$

where D_0 can be shown to simplify to:

$$D_0 = \left(\frac{\lambda_\chi^*}{\lambda_\chi^{(k-1)}}\right)^{\kappa_{\lambda_\chi}} \cdot \exp\left(\frac{\lambda_\chi^{(k-1)} - \lambda_\chi^*}{\gamma_{\lambda_\chi}}\right). \tag{S96}$$

S5.11 Range parameters λ_σ and λ_θ

With $\chi \in \{\sigma, \theta\}$, the full conditional distribution of the range parameters results from Eqs. (S3), (S5) and (S40) as:

$$p(\lambda_\chi|\text{rest}) \propto p(\boldsymbol{\chi}|\boldsymbol{\beta}_\chi, \alpha_\chi, \lambda_\chi) \cdot p(\lambda_\chi) \tag{S97}$$

$$\propto \text{MVN}\left(\log(\boldsymbol{\chi}); \mathbf{X}_\chi\boldsymbol{\beta}_\chi, \boldsymbol{\Sigma}_\chi(\boldsymbol{\ell}, \alpha_\chi, \lambda_\chi)\right) \cdot \text{Gamma}\left(\lambda_\chi; \kappa_{\lambda_\chi}, \gamma_{\lambda_\chi}\right). \tag{S98}$$

Eq. (S98) again represents a non-standard distribution requiring the use of a MH step, which is here also implemented as per Eq. (S94) using a non-symmetric proposal distribution to generate $\lambda_\chi^* > 0$. As per Eqs. (S95) and (S96), the acceptance probability for the proposal sample can be shown to be as follows:

$$P_{\text{acc},\lambda_\chi} = \min\left\{1, \frac{\text{MVN}\left(\log(\boldsymbol{\chi}); \mathbf{X}_\chi\boldsymbol{\beta}_\chi, \boldsymbol{\Sigma}_\chi(\boldsymbol{\ell}, \alpha_\chi, \lambda_\chi^*)\right)}{\text{MVN}\left(\log(\boldsymbol{\chi}); \mathbf{X}_\chi\boldsymbol{\beta}_\chi, \boldsymbol{\Sigma}_\chi\left(\boldsymbol{\ell}, \alpha_\chi, \lambda_\chi^{(k-1)}\right)\right)} \times \left(\frac{\lambda_\chi^*}{\lambda_\chi^{(k-1)}}\right)^{\kappa_{\lambda_\chi}} \cdot \exp\left(\frac{\lambda_\chi^{(k-1)} - \lambda_\chi^*}{\gamma_{\lambda_\chi}}\right)\right\}. \quad \text{(S99)}$$

S5.12 Range parameter λ_η

Finally, the full conditional distribution for λ_η results from Eqs. (S3), (S5) and (S51) as:

$$p(\lambda_\eta|\text{rest}) \propto p(\boldsymbol{\eta}|\boldsymbol{\beta}_\eta, \alpha_\eta, \lambda_\eta) \cdot p(\lambda_\eta) \quad \text{(S100)}$$

$$\propto \text{MVN}\left(\text{logit}(\boldsymbol{\eta}); \mathbf{X}_\eta\boldsymbol{\beta}_\eta, \boldsymbol{\Sigma}_\eta(\boldsymbol{\ell}, \alpha_\eta, \lambda_\eta)\right) \cdot \text{Gamma}\left(\lambda_\eta; \kappa_{\lambda_\eta}, \gamma_{\lambda_\eta}\right). \quad \text{(S101)}$$

As per Eq. (S94), a non-symmetric proposal distribution is used in the MH step for Eq. (S101) so as to generate $\lambda_\eta^* > 0$. And as per Eqs. (S95) and (S96), the acceptance probability for this proposal sample is:

$$P_{\text{acc},\lambda_\eta} = \min\left\{1, \frac{\text{MVN}\left(\text{logit}(\boldsymbol{\eta}); \mathbf{X}_\eta\boldsymbol{\beta}_\eta, \boldsymbol{\Sigma}_\eta(\boldsymbol{\ell}, \alpha_\eta, \lambda_\eta^*)\right)}{\text{MVN}\left(\text{logit}(\boldsymbol{\eta}); \mathbf{X}_\eta\boldsymbol{\beta}_\eta, \boldsymbol{\Sigma}_\eta\left(\boldsymbol{\ell}, \alpha_\eta, \lambda_\eta^{(k-1)}\right)\right)} \times \left(\frac{\lambda_\eta^*}{\lambda_\eta^{(k-1)}}\right)^{\kappa_{\lambda_\eta}} \cdot \exp\left(\frac{\lambda_\eta^{(k-1)} - \lambda_\eta^*}{\gamma_{\lambda_\eta}}\right)\right\}. \quad \text{(S102)}$$

References

Davison A, Padoan S, Ribatet M, 2012. Statistical modelling of spatial extremes. *Statistical Science* **27**(2): 161–186.

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB, 2015. *Bayesian Data Analysis*. Boca Raton: CRC Press, 3rd edition.

Geweke J, 1992. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In Bernardo JM, Berger JO, Dawid AP, Smith AFM (eds.), *Bayesian Statistics 4*, Clarendon Press, Oxford, UK.

Heidelberger P, Welch P, 1983. Simulation run length control in the presence of an initial transient. *Operations Research* **31**: 1109–1144.

Koutsoyiannis D, Kozonis D, Manetas A, 1998. A mathematical framework for studying rainfall intensity-duration-frequency relationships. *Journal of Hydrology* **206**(1): 118–135.

MacEachern S, Berliner L, 1994. Subsampling the gibbs sampler. *The American Statistician* **48**(3): 188–190.

R Core Team, 2015. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Raftery AL, Lewis S, 1992a. How many iterations in the Gibbs sampler? In Bernardo JM, Berger JO, Dawid AP, Smith AFM (eds.), *Bayesian Statistics 4*, Oxford University Press, 763–774.

Raftery AL, Lewis S, 1992b. One long run with diagnostics: Implementation strategies for Markov chain Monte Carlo. *Statistical Science* **7**: 493–497.

Spiegelhalter DJ, Best NG, Carlin BP, Linde AVD, 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**(4): 583–639.

Stephenson AG, 2016. Bayesian inference for extreme value modelling. In Dey D, Yan J (eds.), *Extreme value modeling and risk analysis: methods and applications*, Chapman & Hall/CRC, 257–280.

Watanabe S, 2010. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research* **11**: 3571–3594.