Supplementary Material for “Bayesian model averaging over tree-based dependence structures for multivariate extremes”

Sabrina Vettori[1], Raphaël Huser[1], Johan Segers[2] and Marc G. Genton[1]

1 Recursive formulas for the nested logistic distribution

We first recall in Section 1.1 and Section 1.2 some notation and preliminary results introduced in the appendix of the main paper, and then derive a recursive formula for the partial derivatives of the nested logistic model exponent function.

1.1 Notation

For simplicity, we denote the distribution of the nested logistic model is as $G = \exp(-V)$, where

$$V = \left(\sum_{k=1}^{K} V_k\right)^{\alpha_0}, \qquad V_k = \left\{\sum_{i_k=1}^{D_k} z_{k;i_k}^{-1/(\alpha_0\alpha_k)}\right\}^{\alpha_k}, \qquad k = 1, 2, \ldots, K,$$

with dependence parameters $0 < \alpha_0, \alpha_1, \ldots, \alpha_K \leq 1$, and where, for clarity, we have omitted the function arguments. Moreover, we use the following vector notation:

$$\boldsymbol{z}_k = (z_{k;1}, \ldots, z_{k;D_k})^\top \qquad \textit{All variables in cluster } k = 1, \ldots, K;$$

$$\boldsymbol{z}_{k,1:d_k} = (z_{k;1}, \ldots, z_{k;d_k})^\top \qquad \textit{Sub-vector of the first } d_k \textit{ variables in cluster } k = 1, \ldots, K;$$

$$\boldsymbol{i}_{1:\kappa} = (i_1, \ldots, i_\kappa)^\top \qquad \textit{Vector of } \kappa \textit{ indices.}$$

[1] King Abdullah University of Science and Technology (KAUST), Computer, Electrical and Mathematical Science and Engineering Division (CEMSE) Thuwal 23955-6900, Saudi Arabia.
E-mails: sabrina.vettori@kaust.edu.sa, raphael.huser@kaust.edu.sa, marc.genton@kaust.edu.sa.
[2] Université catholique de Louvain, Institut de Statistique, Biostatistique et Sciences Actuarielles (ISBA) Louvain-la-Neuve B-1348, Belgium.
E-mail: johan.segers@uclouvain.be.

1.2 Preliminary results

From the definition in Section 1.1, we deduce the following derivatives:

$$\frac{\partial V_k}{\partial z_{k;i_k}} = \alpha_k \left\{ \sum_{i_k=1}^{D_k} z_{k;i_k}^{-1/(\alpha_0\alpha_k)} \right\}^{\alpha_k-1} \times \frac{-1}{\alpha_0\alpha_k} z_{k;i_k}^{-1/(\alpha_0\alpha_k)-1} = -\frac{1}{\alpha_0} z_{k;i_k}^{-1/(\alpha_0\alpha_k)-1} V_k^{1-1/\alpha_k}; \quad (1)$$

$$\frac{\partial V}{\partial z_{k;i_k}} = \alpha_0 \left(\sum_{k=1}^{K} V_k \right)^{\alpha_0-1} \frac{\partial V_k}{\partial z_{k;i_k}} = -z_{k;i_k}^{-1/(\alpha_0\alpha_k)-1} V_k^{1-1/\alpha_k} V^{1-1/\alpha_0}; \quad (2)$$

$$\frac{\partial G}{\partial z_{k;i_k}} = -\frac{\partial V}{\partial z_{k;i_k}} \exp(-V) = z_{k;i_k}^{-1/(\alpha_0\alpha_k)-1} G V_k^{1-1/\alpha_k} V^{1-1/\alpha_0}. \quad (3)$$

1.3 Partial derivatives of V

The exponent function V is a function of $D = \sum_{k=1}^{K} D_k$ variables, namely $\boldsymbol{z}_1 = (z_{1;1}, \ldots, z_{1;D_1})^T, \ldots, \boldsymbol{z}_K = (z_{K;1}, \ldots, z_{K;D_K})^\top$. The partial derivative of V with respect to any subset of variables $\boldsymbol{z}_{1,1:d_1}, \ldots, \boldsymbol{z}_{\kappa,1:d_\kappa}$ in $1 \leq \kappa \leq K$ clusters of dimensions $1 \leq d_k \leq D_k$, $k = 1, \ldots, \kappa$, may be expressed as

$$\frac{\partial^{\sum_{k=1}^{\kappa} d_k} V}{\partial \prod_{k=1}^{\kappa} \mathbf{z}_{k;1:d_k}} = \prod_{i_1=1}^{d_1} z_{1;i_1}^{-\frac{1}{\alpha_0\alpha_1}-1} \cdots \prod_{i_\kappa=1}^{d_\kappa} z_{\kappa;i_\kappa}^{-\frac{1}{\alpha_0\alpha_\kappa}-1} \sum_{i_1=1}^{d_1} \cdots \sum_{i_\kappa=1}^{d_\kappa} \gamma_{i_1;\ldots;i_\kappa}^{(d_1;\ldots;d_\kappa)} V_1^{i_1-\frac{d_1}{\alpha_1}} \cdots V_\kappa^{i_\kappa-\frac{d_\kappa}{\alpha_\kappa}} V^{1-\frac{\sum_{k=1}^{\kappa} i_k}{\alpha_0}}, \quad (4)$$

where the coefficients $\gamma_{i_1;\ldots;i_\kappa}^{(d_1;\ldots;d_\kappa)}$ can be computed recursively as demonstrated below.

Proof Equation (4) may be proven by double induction over $\kappa \in \{1, \ldots, K\}$ and $d_k \in \{1, \ldots, D_k\}$, $k = 1, \ldots, \kappa$. The proof also naturally provides a constructive approach to the recursive computation of the coefficients $\gamma_{i_1;\ldots;i_\kappa}^{(d_1;\ldots;d_\kappa)}$. More precisely, we demonstrate the following four steps:

1. (4) holds for $\kappa = 1$ and $d_1 = 1$;
2. If (4) holds for $\kappa = 1$ and $d_1 \in \{1, \ldots, D_1 - 1\}$, then it also holds for $d_1 \mapsto d_1 + 1$;
3. If (4) holds for $\kappa \in \{1, \ldots, K - 1\}$, then it also holds for $\kappa \mapsto \kappa + 1$ with $d_{\kappa+1} = 1$;
4. If (4) holds for $\kappa \in \{1, \ldots, K\}$ and $d_\kappa \in \{1, \ldots, D_{\kappa-1}\}$, then it also holds for $d_\kappa \mapsto d_\kappa + 1$.

Step 1. From (2) we have

$$\frac{\partial V}{\partial z_{1;1}} = -z_{1;1}^{-\frac{1}{\alpha_0\alpha_1}-1} V_1^{1-\frac{1}{\alpha_k}} V^{1-\frac{1}{\alpha_0}},$$

which proves the first step by setting $\gamma_1^{(1)} = -1$.

Step 2. Assuming that (4) holds for $\kappa = 1$ and $d_1 \in \{1, \ldots, D_1 - 1\}$, and using (1) and (2), we obtain

$$\begin{aligned}
\frac{\partial^{d_1+1} V}{\partial \mathbf{z}_{1;1:d_1+1}} &= \prod_{i_1=1}^{d_1} z_{1;i_1}^{-\frac{1}{\alpha_0\alpha_1}-1} \left\{ \sum_{i_1=1}^{d_1} \gamma_{i_1}^{(d_1)} \frac{\partial V_1^{i_1-\frac{d_1}{\alpha_1}}}{\partial z_{1;d_1+1}} V^{1-\frac{i_1}{\alpha_0}} + \sum_{i_1=1}^{d_1} \gamma_{i_1}^{(d_1)} V_1^{i_1-\frac{d_1}{\alpha_1}} \frac{\partial V^{1-\frac{i_1}{\alpha_0}}}{\partial z_{1;d_1+1}} \right\} \\
&= \prod_{i_1=1}^{d_1+1} z_{1;i_1}^{-\frac{1}{\alpha_0\alpha_1}-1} \left\{ -\sum_{i_1=1}^{d_1} \gamma_{i_1}^{(d_1)} \left(i_1 - \frac{d_1}{\alpha_1}\right)\left(\frac{1}{\alpha_0}\right) V_1^{i_1-\frac{d_1+1}{\alpha_1}} V^{1-\frac{i_1}{\alpha_0}} - \sum_{i_1=1}^{d_1} \gamma_{i_1}^{(d_1)} V_1^{(i_1+1)-\frac{d_1+1}{\alpha_1}} \left(1 - \frac{i_1}{\alpha_0}\right) V^{1-\frac{i_1+1}{\alpha_0}} \right\} \\
&= \prod_{i_1=1}^{d_1+1} z_{1;i_1}^{-\frac{1}{\alpha_0\alpha_1}-1} \left\{ -\sum_{i_1=1}^{d_1} \gamma_{i_1}^{(d_1)} \left(i_1 - \frac{d_1}{\alpha_1}\right)\left(\frac{1}{\alpha_0}\right) V_1^{i_1-\frac{d_1+1}{\alpha_1}} V^{1-\frac{i_1}{\alpha_0}} - \sum_{i_1=2}^{d_1+1} \gamma_{i_1-1}^{(d_1)} V_1^{i_1-\frac{d_1+1}{\alpha_1}} \left(1 - \frac{i_1-1}{\alpha_0}\right) V^{1-\frac{i_1}{\alpha_0}} \right\} \\
&= \prod_{i_1=1}^{d_1+1} z_{1;i_1}^{-\frac{1}{\alpha_0\alpha_1}-1} \sum_{i_1=1}^{d_1+1} \gamma_{i_1}^{(d_1+1)} V_1^{i_1-\frac{d_1+1}{\alpha_1}} V^{1-\frac{i_1}{\alpha_0}}
\end{aligned}$$

where

$$\gamma_{i_1}^{(d_1+1)} = -\frac{1}{\alpha_0}\left(i_1 - \frac{d_1}{\alpha_1}\right)\gamma_{i_1}^{(d_1)} - \left(1 - \frac{i_1-1}{\alpha_0}\right)\gamma_{i_1-1}^{(d_1)}, \quad 1 \le i_1 \le d_1 + 1 \tag{5}$$

with

$$\gamma_{i_1}^{(d_1)} = 0, \qquad \text{for all} \quad i_1 \notin \{1, \ldots, d_1\}. \tag{6}$$

Hence, (4) holds by induction for $\kappa = 1$ and any $1 \le d_1 \le D_1$, and the recursive formula to compute coefficients $\gamma_{i_1}^{(d_1)}$ is given by (5) and (6) with the initial condition $\gamma_1^{(1)} = -1$.

Step 3. Assuming that (4) holds for $\kappa \in \{1, \dots, K-1\}$, we obtain

$$\frac{\partial^{1+\sum_{k=1}^{\kappa} d_k} V}{\partial \prod_{k=1}^{\kappa} \mathbf{z}_{k;1:d_k} \partial z_{\kappa+1;1}} = \prod_{i_1=1}^{d_1} z_{1;i_1}^{-\frac{1}{\alpha_0\alpha_1}-1} \cdots \prod_{i_\kappa=1}^{d_\kappa} z_{\kappa;i_\kappa}^{-\frac{1}{\alpha_0\alpha_\kappa}-1} \sum_{i_1=1}^{d_1} \cdots \sum_{i_\kappa=1}^{d_\kappa} \gamma_{i_1;\dots;i_\kappa}^{(d_1;\dots;d_\kappa)} V_1^{i_1-\frac{d_1}{\alpha_1}} \cdots V_\kappa^{i_\kappa-\frac{d_\kappa}{\alpha_\kappa}} \frac{\partial V^{1-\frac{\sum_{k=1}^{\kappa} i_k}{\alpha_0}}}{\partial z_{\kappa+1;1}}$$

$$= -\prod_{i_1=1}^{d_1} z_{1;i_1}^{-\frac{1}{\alpha_0\alpha_1}-1} \cdots \prod_{i_\kappa=1}^{d_\kappa} z_{\kappa;i_\kappa}^{-\frac{1}{\alpha_0\alpha_\kappa}-1} z_{\kappa+1;1}^{-\frac{1}{\alpha_0\alpha_{\kappa+1}}-1} \sum_{i_1=1}^{d_1} \cdots \sum_{i_\kappa=1}^{d_\kappa} \gamma_{i_1;\dots;i_\kappa}^{(d_1;\dots;d_\kappa)} V_1^{i_1-\frac{d_1}{\alpha_1}} \cdots V_\kappa^{i_\kappa-\frac{d_\kappa}{\alpha_\kappa}}$$

$$\times \left(1 - \frac{\sum_{k=1}^{\kappa} i_k}{\alpha_0}\right) V_{\kappa+1}^{1-\frac{1}{\alpha_{\kappa+1}}} V^{1-\frac{1+\sum_{k=1}^{\kappa} i_k}{\alpha_0}} = \prod_{i_1=1}^{d_1} z_{1;i_1}^{-\frac{1}{\alpha_0\alpha_1}-1} \cdots \prod_{i_\kappa=1}^{d_\kappa} z_{\kappa;i_\kappa}^{-\frac{1}{\alpha_0\alpha_\kappa}-1} z_{\kappa+1;1}^{-\frac{1}{\alpha_0\alpha_{\kappa+1}}-1}$$

$$\times \sum_{i_1=1}^{d_1} \cdots \sum_{i_\kappa=1}^{d_\kappa} \gamma_{i_1;\dots;i_\kappa;1}^{(d_1;\dots;d_\kappa;1)} V_1^{i_1-\frac{d_1}{\alpha_1}} \cdots V_\kappa^{i_\kappa-\frac{d_\kappa}{\alpha_\kappa}} V_{\kappa+1}^{1-\frac{1}{\alpha_{\kappa+1}}} V^{1-\frac{1+\sum_{k=1}^{\kappa} i_k}{\alpha_0}}$$

where

$$\gamma_{i_1;\dots;i_\kappa;1}^{(d_1;\dots;d_\kappa;1)} = -\left(1 - \frac{\sum_{k=1}^{\kappa} i_k}{\alpha_0}\right) \gamma_{i_1;\dots;i_\kappa}^{(d_1;\dots;d_\kappa)}, \quad 1 \le i_k \le d_k,\ k = 1, \dots, \kappa \tag{7}$$

with

$$\gamma_{i_1;\dots;i_\kappa;1}^{(d_1;\dots;d_\kappa;1)} = 0, \quad \text{for all} \quad i_k \notin \{1, \dots, d_k\},\ k = 1, \dots, \kappa. \tag{8}$$

Hence, (4) holds by induction for $\kappa \mapsto \kappa + 1$ with $d_{\kappa+1} = 1$ and the recursive formula to compute coefficients $\gamma_{i_1;\dots;i_\kappa;1}^{(d_1;\dots;d_\kappa,1)}$ is given by (7) and (8).

Step 4. Assuming that (4) holds for $\kappa \in \{1,\ldots,K\}$ and $d_\kappa \in \{1,\ldots,D_\kappa - 1\}$, we obtain

$$\frac{\partial^{1+\sum_{k=1}^{\kappa} d_k} V}{\partial \prod_{k=1}^{\kappa} \mathbf{z}_{k;1:d_k} \partial z_{\kappa;d_\kappa+1}} = \prod_{i_1=1}^{d_1} z_{1;i_1}^{-\frac{1}{\alpha_0\alpha_1}-1} \cdots \prod_{i_\kappa=1}^{d_\kappa} z_{\kappa;i_\kappa}^{-\frac{1}{\alpha_0\alpha_\kappa}-1} \left\{ \sum_{i_1=1}^{d_1} \cdots \sum_{i_\kappa=1}^{d_\kappa} \gamma_{i_1;\ldots;i_\kappa}^{(d_1;\ldots;d_\kappa)} V_1^{i_1-\frac{d_1}{\alpha_1}} \cdots \frac{\partial V_\kappa^{i_\kappa-\frac{d_\kappa}{\alpha_\kappa}}}{\partial z_{\kappa;d_\kappa+1}} \right.$$

$$\left. V^{1-\frac{\sum_{k=1}^{\kappa} i_k}{\alpha_0}} + \sum_{i_1=1}^{d_1} \cdots \sum_{i_\kappa=1}^{d_\kappa} \gamma_{i_1;\ldots;i_\kappa}^{(d_1;\ldots;d_\kappa)} V_1^{i_1-\frac{d_1}{\alpha_1}} \cdots V_\kappa^{i_\kappa-\frac{d_\kappa}{\alpha_\kappa}} \frac{\partial V^{1-\frac{\sum_{k=1}^{\kappa} i_k}{\alpha_0}}}{\partial z_{\kappa;d_\kappa+1}} \right\} = \prod_{i_1=1}^{d_1} z_{1;i_1}^{-\frac{1}{\alpha_0\alpha_1}-1} \cdots \prod_{i_\kappa=1}^{d_\kappa+1} z_{\kappa;i_\kappa}^{-\frac{1}{\alpha_0\alpha_\kappa}-1}$$

$$\times \left\{ -\sum_{i_1=1}^{d_1} \cdots \sum_{i_\kappa=1}^{d_\kappa} \gamma_{i_1;\ldots;i_\kappa}^{(d_1;\ldots;d_\kappa)} V_1^{i_1-\frac{d_1}{\alpha_1}} \cdots \left(i_\kappa - \frac{d_\kappa}{\alpha_\kappa}\right)\left(\frac{1}{\alpha_0}\right) V_\kappa^{i_\kappa - \frac{d_\kappa+1}{\alpha_\kappa}} V^{1-\frac{\sum_{k=1}^{\kappa} i_k}{\alpha_0}} \right.$$

$$\left. - \sum_{i_1=1}^{d_1} \cdots \sum_{i_\kappa=1}^{d_\kappa} \gamma_{i_1;\ldots;i_\kappa}^{(d_1;\ldots;d_\kappa)} V_1^{i_1-\frac{d_1}{\alpha_1}} \cdots V_\kappa^{(i_\kappa+1)-\frac{d_\kappa+1}{\alpha_\kappa}} \left(1 - \frac{\sum_{k=1}^{\kappa} i_k}{\alpha_0}\right) V^{1-\frac{1+\sum_{k=1}^{\kappa} i_k}{\alpha_0}} \right\}$$

$$= \prod_{i_1=1}^{d_1} z_{1;i_1}^{-\frac{1}{\alpha_0\alpha_1}-1} \cdots \prod_{i_\kappa=1}^{d_\kappa+1} z_{\kappa;i_\kappa}^{-\frac{1}{\alpha_0\alpha_\kappa}-1} \left\{ -\sum_{i_1=1}^{d_1} \cdots \sum_{i_\kappa=1}^{d_\kappa} \gamma_{i_1;\ldots;i_\kappa}^{(d_1;\ldots;d_\kappa)} V_1^{i_1-\frac{d_1}{\alpha_1}} \cdots \left(i_\kappa - \frac{d_\kappa}{\alpha_\kappa}\right)\left(\frac{1}{\alpha_0}\right) V_\kappa^{i_\kappa-\frac{d_\kappa+1}{\alpha_\kappa}} V^{1-\frac{\sum_{k=1}^{\kappa} i_k}{\alpha_0}} \right.$$

$$\left. - \sum_{i_1=1}^{d_1} \cdots \sum_{i_\kappa=2}^{d_\kappa+1} \gamma_{i_1;\ldots;i_\kappa-1}^{(d_1;\ldots;d_\kappa)} V_1^{i_1-\frac{d_1}{\alpha_1}} \cdots V_\kappa^{i_\kappa-\frac{d_\kappa+1}{\alpha_\kappa}} \left(1 - \frac{i_1+\cdots+i_\kappa-1}{\alpha_0}\right) V^{1-\frac{\sum_{k=1}^{\kappa} i_k}{\alpha_0}} \right\}$$

$$= \prod_{i_1=1}^{d_1} z_{1;i_1}^{-\frac{1}{\alpha_0\alpha_1}-1} \cdots \prod_{i_\kappa=1}^{d_\kappa+1} z_{\kappa;i_\kappa}^{-\frac{1}{\alpha_0\alpha_\kappa}-1} \sum_{i_1=1}^{d_1} \cdots \sum_{i_\kappa=1}^{d_\kappa} \gamma_{i_1;\ldots;i_\kappa}^{(d_1;\ldots;d_\kappa+1)} V_1^{i_1-\frac{d_1}{\alpha_1}} \cdots V_\kappa^{i_\kappa-\frac{d_\kappa+1}{\alpha_\kappa}} V^{1-\frac{\sum_{k=1}^{\kappa} i_k}{\alpha_0}}$$

where

$$\gamma_{i_1;\ldots;i_\kappa}^{(d_1;\ldots;d_\kappa+1)} = -\frac{1}{\alpha_0}\left(i_\kappa - \frac{d_\kappa}{\alpha_\kappa}\right)\gamma_{i_1;\ldots;i_\kappa}^{(d_1;\ldots;d_\kappa)} - \left(1 - \frac{i_1+\cdots+i_\kappa-1}{\alpha_0}\right)\gamma_{i_1;\ldots;i_\kappa-1}^{(d_1;\ldots;d_\kappa)}, \tag{9}$$

if $1 \leq i_k \leq d_k,\ k = 1,\ldots,\kappa$, with

$$\gamma_{i_1;\ldots;i_\kappa}^{(d_1;\ldots;d_\kappa+1)} = 0, \quad \text{for all} \quad i_k \notin \{1,\ldots,d_k\},\ k = 1,\ldots,\kappa, \tag{10}$$

Hence, (4) holds by induction for $\in \{1,\ldots,K\}$ with $d_\kappa \mapsto d_\kappa + 1$ and the recursive formula to compute coefficients $\gamma_{i_1;\ldots;i_\kappa}^{(d_1;\ldots;d_\kappa+1)}$ is given by (9) and (10).

1.4 Complexity

If $\kappa = 1$, the number of coefficients $\gamma_{i_1}^{(d_1)}$ to be computed recursively in (4) is

$$\sum_{h=1}^{d_1}\left(\sum_{i_1=1}^{h} 1\right) = \frac{d_1(d_1+1)}{2}$$

which implies that the complexity is $\mathcal{O}(d_1^2)$. For $1 \leq \kappa \leq K$ clusters of size $d_1, \ldots, d_\kappa$, the number of coefficients $\gamma_{i_1;\ldots;i_\kappa}^{(d_1;\ldots;d_\kappa)}$ to be computed in (4) is

$$\sum_{h=1}^{d_\kappa} \left(\sum_{i_1=1}^{d_1} \cdots \sum_{i_{\kappa-1}=1}^{d_{\kappa-1}} \sum_{i_\kappa=1}^{h} 1 \right) = \sum_{h=1}^{d_\kappa} \sum_{i_\kappa=1}^{h} (d_1 \cdots d_{\kappa-1}) = (d_1 \cdots d_{\kappa-1}) \frac{d_\kappa(d_\kappa+1)}{2}$$

which implies that the total complexity for the computation of the coefficients $\gamma_{i_1;\ldots;i_\kappa}^{(d_1;\ldots;d_\kappa)}$ in (4) is

$$\mathcal{O}\left(\sum_{k=1}^{\kappa} d_1 \cdots d_{k-1} d_k^2 \right).$$

If the cluster size is the same for all clusters, *i.e.*, $d_k = d_1$, for all $k = 2, \ldots, \kappa$, then we have $\mathcal{O}\left(\sum_{k=1}^{\kappa} d_1^{k+1} \right)$.

2 Updating the model parameters given the tree structure using a Metropolis-Hastings algorithm

1. Set initial values for the parameters vector $\boldsymbol{\alpha}^{(0)} = (\alpha_0^0, \alpha_1^0 \ldots, \alpha_K^0)^\top$.

2. Then, repeat the following steps for $r = 1, \ldots, R$ iterations for each parameter α_k, $k = 0, 1, \ldots, K$:

   (a) Simulate candidate parameter value α_k^* based on a proposal distribution $q(\alpha_k^* \mid \alpha_k^c)$, where $\alpha_k^c = \alpha_k^{(r-1)}$ indicates the value of the parameter at the previous iteration.

   (b) Define the current parameter vector $\boldsymbol{\alpha}^c = \boldsymbol{\alpha}^{(r-1)}$ and replace its k-th component by α_k^* to define the candidate parameter vector $\boldsymbol{\alpha}^*$.

   (c) If the acceptance probability ratio

   $$\min\left\{ \frac{\pi(\boldsymbol{\alpha}^*, \mathcal{T}^c \mid \mathbf{m}_1, \ldots, \mathbf{m}_N) q(\alpha_k^c \mid \alpha_k^*)}{\pi(\boldsymbol{\alpha}^c, \mathcal{T}^c \mid \mathbf{m}_1, \ldots, \mathbf{m}_N) q(\alpha_k^* \mid \alpha_k^c)}, 1 \right\} > U \sim \mathrm{Unif}(0, 1),$$

   where $\mathcal{T}^c$ denote the (fixed) tree structure, then accept the proposed parameter value, *i.e.*, $\boldsymbol{\alpha}^{(r)} = \boldsymbol{\alpha}^*$, otherwise reject, *i.e.*, $\boldsymbol{\alpha}^{(r)} = \boldsymbol{\alpha}^{(r-1)}$.

After some burn-in iterations, this provides a dependent sample from the posterior distribution $\pi(\boldsymbol{\alpha}, \mathcal{T}^c \mid \mathbf{m}_1, \ldots, \mathbf{m}_N)$ of the parameters given the data $\mathbf{m}_1, \ldots, \mathbf{m}_N$ and some fixed tree structure $\mathcal{T}^c$. In our algorithm, the tree structure $\mathcal{T}^c$ is also updated at every iteration r, and the next section describes how this is performed based on the reversible jump MCMC.

3 Reversible jump MCMC algorithm

1. Set an initial configuration for the state space $\mathbf{x}^0 = (\boldsymbol{\alpha}^0, \mathbf{u}^0)^\top$ and choose a corresponding initial tree structure $\mathcal{T}^0$.

2. Repeat the following steps for $r = 1, \ldots, R$ iterations:

   (a) Update each parameter α_k, $k = 0, 1, \ldots, K$ as in the previous section.

   (b) Select a move type from the set of reversible moves which defines a proposed transition from the current state $\mathbf{x}^c = (\boldsymbol{\alpha}^c, \mathbf{u}^c)$ of dimension $\dim(\boldsymbol{\alpha}^c)$ (based on tree $\mathcal{T}^c$) to the proposed state $\mathbf{x}^* = (\boldsymbol{\alpha}^*, \mathbf{u}^*)$ of dimension $\dim(\boldsymbol{\alpha}^*)$ (based on tree $\mathcal{T}^*$), where $\mathbf{x}^c = \mathbf{x}^{(r-1)}$ indicates the state space configuration at the previous iteration.

   (c) If the transition involves a dimension change, generate a random auxiliary variable $\mathbf{u}^c$ based on the proposal distribution $q(\mathbf{u}^c, \boldsymbol{\alpha}^c \mid \mathbf{u}^*, \boldsymbol{\alpha}^*)$ in order to meet the reversibility condition $\dim(\boldsymbol{\alpha}^c)$ =$\dim(\boldsymbol{\alpha}^*)$.

   (d) If the acceptance probability ratio

   $$\min\left\{\frac{\pi(\boldsymbol{\alpha}^*, \mathcal{T}^* \mid \mathbf{m}_1, \ldots, \mathbf{m}_N) q(\mathbf{u}^c, \boldsymbol{\alpha}^c \mid \mathbf{u}^*, \boldsymbol{\alpha}^*) \pi_{\mathbf{x}^c \to \mathbf{x}^*}}{\pi(\boldsymbol{\alpha}^c, \mathcal{T}^c \mid \mathbf{m}_1, \ldots, \mathbf{m}_N) q(\mathbf{u}^*, \boldsymbol{\alpha}^* \mid \mathbf{u}^c, \boldsymbol{\alpha}^c) \pi_{\mathbf{x}^* \to \mathbf{x}^c}} \left|\frac{\partial(\boldsymbol{\alpha}^*, \mathbf{u}^*)}{\partial(\boldsymbol{\alpha}^c, \mathbf{u}^c)}\right|, 1\right\} > U \sim \mathrm{Unif}(0,1),$$

   then accept the proposed transition and add the proposed state space configuration to the Markov chain, *i.e.*, $\mathbf{x}^{(r)} = \mathbf{x}^*$ and $\mathcal{T}^{(r)} = \mathcal{T}^*$, otherwise reject and add the current state space configuration to the Markov chain, *i.e.*, $\mathbf{x}^{(r)} = \mathbf{x}^{(r-1)}$ and $\mathcal{T}^{(r)} = \mathcal{T}^{(r-1)}$.

After some burn-in iterations, this provides a dependent sample from the posterior distribution $\pi(\boldsymbol{\alpha}, \mathcal{T} \mid \mathbf{m}_1, \ldots, \mathbf{m}_N)$.

4 Marginal modelling

In our study, we first calculate the time series of monthly maxima and subtract the monthly means to partially remove non-stationarity, assuming the dependence maintains the same throughout all months of the year. For each pollutant and meteorological parameters, we deal with the remaining non-stationary effects by including the time covariate t into the modelling of the location parameter μ. In particular, we consider: the linear trend model, *i.e.*, $\mu_d(t) = \beta_{0;d} + \beta_{1;d} t$, the cycle trend model, *i.e.*, $\mu_d(t) = \beta_{0;d} + \beta_{1;d} \sin\left(\frac{2\pi t}{12}\right) + \beta_{2;d} \cos\left(\frac{2\pi t}{12}\right)$, and the linear and cycle trend model, *i.e.*, $\mu_d(t) = \beta_{0;d} + \beta_{1;d} t + \beta_{2;d} \sin\left(\frac{2\pi t}{12}\right) + \beta_{3;d} \cos\left(\frac{2\pi t}{12}\right)$, where the βs are unknown parameters measuring the effect of the time covariate. We then fit each of these models at each site $d = 1, \ldots, D$ by maximizing the likelihood

$$
\begin{aligned}
L(\mu_d(t), \sigma_d, \xi_d \mid m_{d;1}, \ldots, m_{d;N}) = \sigma_d^{-N} \exp\left[-\sum_{i=1}^{N} \left\{ 1 + \xi_d \left(\frac{m_{i;d} - \mu_d(t)}{\sigma_d} \right) \right\}^{-\frac{1}{\xi_d}} \right] \\
\times \prod_{i=1}^{N} \left\{ 1 + \xi_d \left(\frac{m_{i;d} - \mu_d(t)}{\sigma_d} \right) \right\}^{-1+\frac{1}{\xi_d}},
\end{aligned}
$$

where $1 + \xi_d(\frac{m_i - \mu_d(t)}{\sigma_d}) > 0$, for $i = 1, \ldots, N$, $\xi_d \neq 0$, and the GEV location parameter $\mu_d(t)$ varies across time t. We select the best model for each of the sites under study by minimizing the BIC criterion, defined as BIC$= -2\ell(\mathcal{M}) + p\log(N)$, where p represent the number of parameters in the model $\mathcal{M}$, N is the sample size, and $\ell(\mathcal{M})$ is the maximized log-likelihood for model $\mathcal{M}$. Finally, we transform the margins to the Fréchet scale using our "best" model as follows:

$$
\tilde{m}_{i;d} = -1/\log\{\hat{G}_d(m_{i;d})\}, \quad i = 1, \ldots, N,\ d = 1, \ldots, D,
$$

where $\hat{G}_d$ denotes the fitted GEV distribution (for the "best" model) with estimated parameters $\hat{\mu}_d(t), \hat{\sigma}_d$ and $\hat{\xi}_d$.