UCL Université catholique de Louvain

INSTITUT DE STATISTIQUE
BIOSTATISTIQUE ET
SCIENCES ACTUARIELLES
(ISBA)

ISBA

DISCUSSION PAPER

2017/17

Bayesian Clustering and Dimension Reduction in Multivariate Extremes

Vettori, S., Huser, R., Segers, J. and M. Genton

Bayesian Clustering and Dimension Reduction in Multivariate Extremes

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

May 31, 2017

Abstract

Describing the complex dependence structure of multivariate extremes is particularly challenging and requires very versatile, yet interpretable, models. To tackle this issue we explore two related approaches: clustering and dimension reduction. In particular, we develop a novel statistical algorithm that takes advantage of the inherent hierarchical dependence structure of the max-stable nested logistic distribution and that uses reversible jump Markov chain Monte Carlo techniques to identify homogeneous clusters of variables. Dimension reduction is achieved when clusters are found to be completely independent. We significantly decrease the computational complexity of full likelihood inference by deriving a recursive formula for the nested logistic model likelihood. The algorithm performance is verified through extensive simulation experiments which also consider different likelihood procedures. The new methodology is used to investigate the dependence relationships between extreme concentration of multiple pollutants across a number of sites in California and how these pollutants are related to extreme weather conditions. Overall, we show that our approach allows for the identification of homogeneous clusters of extremes and has valid applications in multivariate data analysis, such as air pollution monitoring, where it can guide policymaking.

Keywords: Air pollution; Clustering; Extreme event; Fast likelihood inference; Reversible jump Markov chain Monte Carlo.

[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 Introduction

Modelling multivariate extreme phenomena requires flexible, yet interpretable, models with sound theoretical underpinning, criteria that are exponentially more difficult to meet as dimensionality increases; see, *e.g.*, the review papers by Davison *et al.* (2012) and Davison and Huser (2015). The analysis of extreme events is also very challenging because of the intrinsic lack of data; extreme datasets consist of only the largest observations in a sample, and therefore are generally characterised by small sample sizes. Clustering and dimension reduction are useful methods applied in many fields to summarise and reduce the complexity of multivariate datasets; however, most standard statistical techniques are generally unsuitable to describe the behaviour of random variables in the tails. Due to these limitations, most applications of multivariate Extreme-Value Theory have focused chiefly on relatively low-dimensional cases. Clustering of extreme events has been mostly carried out to study the effects of temporal dependence within univariate time series; see Chavez-Demoulin and Davison (2012) for a survey of these methods. Recently, however, Haug *et al.* (2015) proposed to perform dimension reduction in the multivariate framework, adapting the classical Principal Components Analysis (PCA) to the context of Extreme-Value Theory. Chautru (2015) combined this approach with statistical learning and data mining methods, introducing a non-parametric technique able to identify possibly overlapping clusters of asymptotically dependent extremes. Cooley and Thibaud (2016) summarised tail dependence in high dimensions by the decomposition of a matrix of pairwise dependence metrics. Moreover, Bernard *et al.* (2013) proposed a new clustering technique for spatial extremes that integrates the F-madogram distance in the K-means algorithm. However, none of these approaches allows for the assessment of uncertainty in the final clustering configuration. In this work, we contribute to this developing research field by proposing an easily interpretable and efficient clustering and dimension reduction technique for multivariate extremes in moderate or large dimensions that takes the clustering uncertainty into account for prediction and relies on the nested logistic distribution (McFadden, 1978; Tawn, 1990; Coles and Tawn, 1991; Stephen-

son, 2003), which is max-stable and, therefore, asymptotically justified for the description of the whole multivariate extremal dependence structure (de Haan and Resnick, 1977; de Haan, 1984). In particular, thanks to its inherent hierarchical dependence structure, the nested logistic model is able to describe the dependence within and between distinct clusters of extreme variables.

Likelihood inference for max-stable distributions is known to be challenging because of its computational burden. Indeed, the full likelihood in large dimensions becomes numerically intractable (Castruccio *et al.*, 2016; Bienvenüe and Robert, 2015). One classical solution that dramatically reduces the computational burden but leads to a loss of efficiency consists in using composite likelihood techniques (Varin and Vidoni, 2005; Varin, 2008; Padoan *et al.*, 2010; Genton *et al.*, 2011; Davison and Gholamrezaee, 2012; Huser and Davison, 2013). Other approaches include a simplified likelihood procedure proposed by Stephenson and Tawn (2005) which leads to more efficient inference both computationally and statistically. In this work, we propose an alternative solution to the full likelihood computational burden when dealing with high dimensions, by deriving a recursive formula that significantly reduces the complexity of the nested logistic likelihood without causing any efficiency loss or introducing any additional bias.

The nested logistic model can also be defined in terms of nested Archimedean copulas or nested Gumbel copulas (Okhrin *et al.*, 2013a; Hofert and Pham, 2013). There exist already a number of papers on inference on such copulas, both parametric and non-parametric, see, e.g., Okhrin *et al.* (2013b), Segers and Uyttendaele (2014) and Górecki *et al.* (2016). Our approach consists in embedding the nested logistic model within a Bayesian framework. Few authors have so far implemented fully-Bayesian inference for the estimation of extremal dependence; see, *e.g.*, Guillotte *et al.* (2011), Ribatet *et al.* (2012), Reich and Shaby (2012), Sabourin *et al.* (2013), Sabourin and Naveau (2014), Shaby (2014) and Thibaud *et al.* (2016). Taking advantage of Bayesian computer-intensive methods, such as the reversible jump Markov chain Monte Carlo (RJ-MCMC) algorithm (Green, 1995), we construct a clustering and dimension reduction algorithm which is able to estimate the model parameters and simultaneously determine the tree

configuration governing the extremal dependence and its uncertainty, directly from the data. An extensive simulation study is conducted in order to evaluate the algorithm's performance, implementing both the Stephenson and Tawn (2005) likelihood and our recursive likelihood formula within the algorithm.

Finally, we apply our algorithm to multivariate air pollution data in order to investigate the dependence relations between extreme concentration of air pollutants and to understand how these relate to extreme weather conditions.

The general framework of the Extreme-Value Theory and the nested logistic model are recalled in Section 2 and the associated likelihood inference methods, including the proposed recursive likelihood formula, are summarised in Section 3. In Section 4 we describe the proposed dimension reduction and clustering algorithm, and in Section 5 we verify its performance through an extensive simulation study. The algorithm is applied to air quality data in Section 6. Finally, these results are discussed in Section 7. Theoretical details are reported in the Appendix and the Supplementary Material.

2 Modelling Extremal Dependence

2.1 General framework

Suppose that $\{\mathbf{Y}_i = (Y_{i;1}, \ldots, Y_{i;D})^\top\}_{i=1}^N$ is a sequence of independent and identically distributed (i.i.d.) copies of the D-dimensional random vector $\mathbf{Y}$, representing D variables of interest having a common distribution function (d.f.) F with margins $F_d(y), d = 1, \ldots, D$, and let $\mathbf{M}_n = (M_{n;1}, \ldots, M_{n;D})^\top = \left(\max_{1\le i\le n} Y_{i;1}, \ldots, \max_{1\le i\le n} Y_{i;D}\right)^\top$ denote the vector of multivariate componentwise maxima. We assume that, as $n \to \infty$, for some sequences of vectors $\mathbf{a}_n = (a_{n;1}, \ldots, a_{n;D})^\top \in \mathbb{R}_+^D$ and $\mathbf{b}_n = (b_{n;1}, \ldots, b_{n;D})^\top \in \mathbb{R}^D$, the renormalised vector of componentwise maxima $\mathbf{M}_n^* = \mathbf{a}_n^{-1}(\mathbf{M}_n - \mathbf{b}_n)$ converges in distribution to the random vector $\mathbf{Z}$, with limiting D-dimensional extreme-value d.f. G and non-degenerate margins; that is, the distribution F belongs to the max-domain of attraction (MDA) of G. Upon marginal standardization,

we may assume unit-Fréchet margins, *i.e.*, $G_d(z_d) = \exp(-1/z_d)$ for $z_d > 0, d = 1, \ldots, D$. The joint distribution of the random vector $\mathbf{Z}$ may be written as

$$\mathrm{P}(\mathbf{Z} \leq \mathbf{z}) = G(\mathbf{z}) = \exp\{-V(\mathbf{z})\}, \qquad V(\mathbf{z}) = \int_{S_D} \max_{1 \leq d \leq D} \left(\frac{\omega_d}{z_d}\right) \mathrm{d}H(\boldsymbol{\omega}), \qquad \mathbf{z} \in \mathbb{R}_+^D, \quad (1)$$

where $V(\mathbf{z})$ is a homogeneous function of order -1, *i.e.*, $V(s\mathbf{z}) = s^{-1}V(\mathbf{z})$, called exponent function, and H is a finite spectral measure on the unit simplex $S_D = \{\boldsymbol{\omega} \in [0,1]^D \colon \sum_{d=1}^D \omega_d = 1\}$ for $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_D)^\top$, satisfying the mean constraints $\int_{S_D} \omega_d \mathrm{d}H(\boldsymbol{\omega}) = 1, d = 1, \ldots, D$ (de Haan and Resnick, 1977; de Haan, 1984).

A useful summary of the dependence strength between multivariate extremes is the extremal coefficient, proposed by Smith (1990) and defined as $\theta_D = V(\mathbf{1}) \in [1, D]$, where $\mathbf{1}$ is a vector of ones. The coefficient θ_D decreases as the dependence strength between the margins increases, with $\theta_D = 1$ and $\theta_D = D$ corresponding to complete dependence and independence respectively.

2.2 The logistic and nested logistic models

Among max-stable distributions, the oldest and simplest one is the logistic model (Gumbel, 1960a,b), characterised by the exponent function

$$V_l(\mathbf{z} \mid \alpha_0) = \left(\sum_{d=1}^D z_d^{-1/\alpha_0}\right)^{\alpha_0}, \quad \alpha_0 \in (0, 1]. \quad (2)$$

The parameter α_0 summarises the dependence strength between the extreme observations $\mathbf{z} = (z_1, \ldots, z_D)^\top \in \mathbb{R}_+^D$. In particular, the cases of complete dependence and independence between the margins correspond to $\alpha_0 \to 0$ and $\alpha_0 = 1$, respectively. Therefore, the logistic model summarises the extremal dependence structure by only one parameter α_0, and its components are exchangeable. The extremal coefficient for the logistic distribution is $\theta_D = D^{\alpha_0}$.

Generalising the idea of the logistic model, McFadden (1978) and Tawn (1990), see also Coles and Tawn (1991), proposed the nested logistic distribution, a more flexible multivariate max-stable model that maintains the logistic model's simplicity and interpretability. The nested

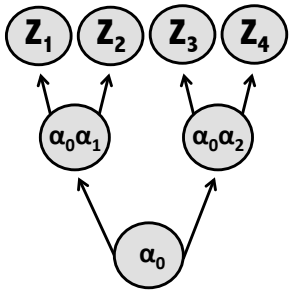

Figure 1: Example of simple tree structure, summarising the extremal dependence of the vector $\mathbf{Z} = (Z_1, Z_2, Z_3, Z_4)^\top$, where the dependence between the variables $(Z_i, Z_j)^\top$, $i \in \{1, 2\}$, $j \in \{3, 4\}$ is summarised by a logistic distribution with parameter α_0, and the dependence between the variables $(Z_1, Z_2)^\top$ and $(Z_3, Z_4)^\top$ is summarised by a logistic distribution with parameter $\alpha_0\alpha_1$ and $\alpha_0\alpha_2$, respectively.

logistic model implies that the vector $\mathbf{z}$, containing the extreme observations, is split into K homogeneous clusters and it describes the dependence within and between these clusters using logistic distributions (2). It is defined by the exponent function

$$V_{nl}(\mathbf{z} \mid \boldsymbol{\alpha}) = V_l\left\{V_l(\mathbf{z}_1 \mid \alpha_0\alpha_1)^{-1}, \ldots, V_l(\mathbf{z}_K \mid \alpha_0\alpha_K)^{-1} \mid \alpha_0\right\} = \left\{\sum_{k=1}^{K}\left(\sum_{i_k=1}^{D_k} z_{k;i_k}^{-\frac{1}{\alpha_0\alpha_k}}\right)^{\alpha_k}\right\}^{\alpha_0}, \quad (3)$$

where $V_l(\mathbf{z}_k \mid \alpha_k)$ is the logistic exponent function in (2) of the sub-vector $\mathbf{z}_k = (z_{k;1}, \ldots, z_{k;D_k})^\top$ comprising extreme observations belonging to the k^{th} cluster of dimension $D_k \in \{1, \ldots, D\}$, $k = 1, \ldots, K$, with $D = \sum_{k=1}^{K} D_k$, and where $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_K)^\top \in (0, 1]^{K+1}$ are the between-cluster and within-cluster dependence parameters. This hierarchical structure can be represented with a tree, as illustrated in Figure 1. In practice, more complex situations may arise with more clusters and, possibly, an arbitrary number of layers. Apart from its interpretability, the nested logistic model is also appealing for its inference properties as described in Section 3.3. The extremal coefficient for the nested logistic distribution is $\theta_D = \left(\sum_{k=1}^{K} D_k^{\alpha_k}\right)^{\alpha_0}$.

3 Full Likelihood Inference

3.1 Classical approach

The classical inference approach consists in splitting the data into N blocks with an equal number of observations n, from which componentwise maxima data $\mathbf{m}_i = (m_{i;1}, \ldots, m_{i;D})^\top$, $i = 1, \ldots, N$,

are extracted. Assuming unit Fréchet marginal distributions, the likelihood function may be expressed as

$$L(\boldsymbol{\alpha} \mid \mathbf{m}_1, \ldots, \mathbf{m}_N) = \prod_{i=1}^{N} \exp\{-V(\mathbf{m}_i \mid \boldsymbol{\alpha})\} \left\{ \sum_{E \in \mathcal{E}} \prod_{S \in E} -\dot{V}_S(\mathbf{m}_i \mid \boldsymbol{\alpha}) \right\}, \tag{4}$$

where $\boldsymbol{\alpha}$ is the vector of unknown dependence parameters, $\mathcal{E}$ denotes the collection of all partitions of $\mathcal{D} = \{1, \ldots, D\}$, $\dot{V}_S$ denotes the partial derivative of the function V in (1) with respect to the variables whose indices lie in $S \subset \mathcal{D}$. The cardinality of $\mathcal{E}$ equals B_D, the Bell number of order D (Graham *et al.*, 1988), meaning that the number of terms to be computed and the storage space required for evaluating the likelihood grow super-exponentially with the dimension D (Castruccio *et al.*, 2016; Huser *et al.*, 2016).

3.2 Stephenson and Tawn likelihood

Stephenson and Tawn (2005) developed a simplified likelihood approach which uses the additional information of the partition combining componentwise maxima that occurred simultaneously. More specifically, for each $i = 1, \ldots, N$, let $E_i \in \mathcal{E}$ denote the partition combining the maxima $\mathbf{m}_i$ with identical occurrence times. For instance, for $D = 3$, $E_1 = \{1, \{2, 3\}\}$ indicates that, in the first block, the maxima $m_{1;2}$ and $m_{1;3}$ occurred simultaneously, but separately from $m_{1;1}$. The Stephenson and Tawn likelihood might be written as

$$L\{\boldsymbol{\alpha} \mid (\mathbf{m}_1, E_1), \ldots, (\mathbf{m}_N, E_N)\} = \prod_{i=1}^{N} \exp\{-V(\mathbf{m}_i \mid \boldsymbol{\alpha})\} \left\{ \prod_{S \in E_i} -\dot{V}_S(\mathbf{m}_i \mid \boldsymbol{\alpha}) \right\}, \tag{5}$$

where $\boldsymbol{\alpha}$ is the vector of unknown parameters and $\dot{V}_S$ denotes the partial derivative of V in (1) with respect to the variables in S. Therefore, using the partitions leads to a more efficient inference, both computationally and statistically, which comes at the price of introducing bias in the parameter estimation, see, e.g., Huser *et al.* (2016); Thibaud *et al.* (2016) and our simulation experiment in Section 5. Wadsworth (2014) proposed a bias reduction technique, which works well in low dependence scenarios, but remains intensive in case of strong dependence for large

dimensions. To further speed up the computation of (5), we derived a recursive formula to compute the partial derivatives of the nested logistic model exponent function. Let $\boldsymbol{m}_{i;k;1:d_k}$ be the sub-vector of the first $1 \leq d_k \leq D_k$ components of the maxima vector belonging to $1 \leq \kappa \leq K$ clusters, with $k = 1, \ldots, \kappa$. The formula may be expressed as

$$\begin{aligned}\frac{\partial^{\sum_{k=1}^{\kappa} d_k} V_{nl}(\mathbf{m}_i \mid \boldsymbol{\alpha})}{\partial \prod_{k=1}^{\kappa} \mathbf{m}_{i;k;1:d_k}} &= \prod_{i_1=1}^{d_1} m_{i;1;i_1}^{-\frac{1}{\alpha_0\alpha_1}-1} \cdots \prod_{i_\kappa=1}^{d_\kappa} m_{i;\kappa;i_\kappa}^{-\frac{1}{\alpha_0\alpha_\kappa}-1} \\ &\times \sum_{i_1=1}^{d_1} \cdots \sum_{i_\kappa=1}^{d_\kappa} \gamma_{i_1;\ldots;i_\kappa}^{(d_1;\ldots;d_\kappa)} V_l(\mathbf{m}_{i;1;1:d_1}^{1/\alpha_0} \mid \alpha_1)^{i_1-\frac{d_1}{\alpha_1}} \cdots V_l(\mathbf{m}_{i;\kappa;1:d_\kappa}^{1/\alpha_0} \mid \alpha_\kappa)^{i_\kappa-\frac{d_\kappa}{\alpha_\kappa}} V_{nl}(\mathbf{m}_i \mid \boldsymbol{\alpha})^{1-\frac{\sum_{k=1}^{\kappa} i_k}{\alpha_0}};\end{aligned} \tag{6}$$

the total complexity for the computation of the recursive coefficients $\gamma_{i_1;\ldots;i_\kappa}^{(d_1;\ldots;d_\kappa)}$ is reduced to $\mathcal{O}\left(\sum_{k=1}^{\kappa} d_1 \cdots d_{k-1} d_k^2\right)$. If all clusters are of equal size, *i.e.*, $d_k = d_1$, for all $k = 2, \ldots, \kappa$, then one has $\mathcal{O}\left(\sum_{k=1}^{\kappa} d_1^{k+1}\right)$. The proof of (6) and the expression for $\gamma_{i_1;\ldots;i_\kappa}^{(d_1;\ldots;d_\kappa)}$ are given in the Supplementary Material.

3.3 Recursive full likelihood formula

Similarly to the recursive full likelihood formula for the logistic distribution that was derived by Shi (1995), we here develop a recursive formula to efficiently compute the full likelihood function (4) for the nested logistic distribution (3), while avoiding the combinatorial explosion of floating point operations. Let $\boldsymbol{m}_{i;k;1:D_k}$ be the sub-vector of maxima belonging to the k^{th} cluster of dimension D_k; the recursive likelihood formula may be written as

$$\begin{aligned}L(\boldsymbol{\alpha} \mid \mathbf{m}_1, \ldots, \mathbf{m}_N) &= \prod_{i=1}^{N} \exp\{-V_{nl}(\mathbf{m}_i \mid \boldsymbol{\alpha})\} \prod_{i_1=1}^{D_1} m_{i;1;i_1}^{-\frac{1}{\alpha_0\alpha_1}-1} \cdots \prod_{i_K=1}^{D_K} m_{i;K;i_K}^{-\frac{1}{\alpha_0\alpha_k}-1} \sum_{i_1=1}^{D_1} \cdots \sum_{i_K=1}^{D_K} \\ &\times \sum_{j=1}^{\sum_{k=1}^{K} i_k} \beta_{i_1;\ldots;i_K;j}^{(D_1;\ldots;D_K)} V_l(\mathbf{m}_{i;1;1:D_1}^{1/\alpha_0} \mid \alpha_1)^{i_1-\frac{D_1}{\alpha_1}} \cdots V_l(\mathbf{m}_{i;K;1:D_K}^{1/\alpha_0} \mid \alpha_k)^{i_K-\frac{D_K}{\alpha_k}} V_{nl}(\mathbf{m}_i \mid \boldsymbol{\alpha})^{j-\frac{\sum_{k=1}^{K} i_k}{\alpha_0}},\end{aligned} \tag{7}$$

where $V_{nl}(\mathbf{m}_i \mid \boldsymbol{\alpha})$ is defined in (3) and $V_l(\mathbf{m}_{i;k;1:D_k}^{1/\alpha_0} \mid \alpha_k) = \left(\sum_{i_k=1}^{D_k} m_{i;i_k}^{-1/\alpha_0\alpha_k}\right)^{\alpha_k}$, $i = 1, \ldots, N$, $k = 1, \ldots, K$. The coefficients $\beta_{j_1;\ldots;j_K;j}^{(D_1;\ldots;D_K)}$ can be computed recursively in explicit form, as in (6), reducing the computational complexity to $\mathcal{O}\left(\sum_{k=1}^{K} (D_1 + \cdots + D_k) D_1 \cdots D_{k-1} D_k^2\right) \ll B_D$; the proof of (7) and the expression for $\beta_{j_1;\ldots;j_K;j}^{(D_1;\ldots;D_K)}$ are given in Appendix A.3. When all the clusters

are of equal size, *i.e.* $D_k = D_1$ for all $k = 2, \ldots, K$, the complexity is $\mathcal{O}\left(K \sum_{k=1}^{K} D_1^{k+2}\right)$, which differs from the computation of (6) by one order only. Therefore, the computational time needed to compute the full likelihood using the recursive formula is polynomial in terms of the number of variables D, but exponential-linear in terms of the number of clusters K, suggesting that in practice it might be convenient to prevent K from being large when D is large. The recursive formula is coded in `C++`, integrated within `R` using the package `Rcpp`. The code is available on the authors' website `https://extstat.kaust.edu.sa/Pages/Sabrina_Vettori.aspx`. Recursive formulas for deeper trees with more layers may be derived similarly, although at a higher computational complexity.

4 Dimension Reduction and Clustering Algorithm

4.1 Dimension reduction using Metropolis-Hastings

To identify independent clusters of variables, we need to allow the dependence parameters in (3) to take values at the boundary of their domain of definition, *i.e.*, $\alpha_p = 1$, $p = 0, \ldots, K$. Using the Metropolis-Hastings Markov chain Monte Carlo (MCMC) algorithm (Hasting, 1970), we sample from the posterior distribution $\pi(\boldsymbol{\alpha} \mid \mathbf{m}_1, \ldots, \mathbf{m}_N)$, with $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_K)^\top$, obtained as a product of uninformative independent prior distributions $\pi(\boldsymbol{\alpha}) = \pi(\alpha_0)\pi(\alpha_1)\cdots\pi(\alpha_K)$ and the likelihood $L(\boldsymbol{\alpha} \mid \mathbf{m}_1, \ldots, \mathbf{m}_N)$ in (5), where the partition notation has been omitted for simplicity, or in (7). At each MCMC iteration, a candidate parameter value α_p^* is proposed for each parameter α_p based on the proposal distribution $q(\alpha_p^* \mid \alpha_p^c)$, where α_p^c denotes the current value of the parameter α_p. Similarly to Coles and Pauli (2002), we define the proposal distribution as a mixture between a uniform distribution centred at the current value of the parameter α_p^c with an interval of length $2\varepsilon_p$, and a point mass at $\alpha_p = 1$, *i.e.*,

$$q(\alpha_p^* \mid \alpha_p^c) = \begin{cases} 0.5\delta_1 + 0.5p(\alpha_p^* \mid \alpha_p^c), & \text{if } \alpha_p^c \neq 1 \text{ and } 1 \in [\alpha_p^c - \varepsilon_p, \alpha_p^c + \varepsilon_p]; \\ p(\alpha_p^* \mid \alpha_p^c), & \text{if } \alpha_p^c = 1 \text{ or } 1 \notin [\alpha_p^c - \varepsilon_p, \alpha_p^c + \varepsilon_p], \end{cases}$$

where δ_1 denotes the Dirac delta function centred at 1 and $p(\alpha_p^* \mid \alpha_p^c)$ is the density of a uniform random variable with boundaries $\left\{\max\left(0.1, \alpha_p^c - \varepsilon_p\right), \min\left(\alpha_p^c + \varepsilon_p, 1\right)\right\}$. The value of α_p^c is then updated based on the acceptance probability $\min\left\{\frac{\pi(\boldsymbol{\alpha}^*|\mathbf{m}_1,\ldots,\mathbf{m}_N)q(\alpha_p^c|\alpha_p^*)}{\pi(\boldsymbol{\alpha}^c|\mathbf{m}_1,\ldots,\mathbf{m}_N)q(\alpha_p^*|\alpha_p^c)}, 1\right\}$. Adaptive methodologies for choosing the proposal tuning parameter ε_p allow for efficient simulations even in high dimensions. We update ε_p at each iteration, ensuring that the acceptance ratio takes values between 20% and 50%, in order to guarantee well-mixing properties of the chain. A description of the algorithm in pseudocode is provided in the Supplementary Material.

4.2 Clustering using reversible jump MCMC

To estimate the unknown extremal dependence structure directly from the data, we explore an extension of the MCMC algorithm, called reversible jump MCMC algorithm (Green, 1995). The reversible jump MCMC allows for the simulation of the posterior distribution on spaces of possibly varying dimensions and therefore it is capable of moving between all possible extremal dependence structures that can be represented within a tree framework. The posterior probability associated with each tree structure is obtained from the relative time such a tree has spent on the chain. Any proposed transition from the current tree, with parameters $\boldsymbol{\alpha}^c$ of dimension $\dim(\boldsymbol{\alpha}^c)$, to the proposed tree, with parameters $\boldsymbol{\alpha}^*$ of dimension $\dim(\boldsymbol{\alpha}^*)$, must be reversible. Therefore, if the transition involves a dimension change, a random auxiliary variable $\mathbf{u}^c$ is generated such that, when moving from the current status $\mathbf{x}^c = (\boldsymbol{\alpha}^c, \mathbf{u}^c)$ to the proposed status $\mathbf{x}^* = (\boldsymbol{\alpha}^*, \mathbf{u}^*)$, the dimension matching condition $\dim(\mathbf{x}^c) = \dim(\mathbf{x}^*)$ is satisfied and the mapping $g_{\mathbf{x}^c \to \mathbf{x}^*}$ is a bijection. The proposed transition is accepted based on the acceptance probability

$$\min\left\{\frac{\pi(\boldsymbol{\alpha}^* \mid \mathbf{m}_1, \ldots, \mathbf{m}_N)}{\pi(\boldsymbol{\alpha}^c \mid \mathbf{m}_1, \ldots, \mathbf{m}_N)} \frac{q(\mathbf{u}^c, \boldsymbol{\alpha}^c \mid \mathbf{u}^*, \boldsymbol{\alpha}^*)}{q(\mathbf{u}^*, \boldsymbol{\alpha}^* \mid \mathbf{u}^c, \boldsymbol{\alpha}^c)} \frac{\pi_{\mathbf{x}^c \to \mathbf{x}^*}}{\pi_{\mathbf{x}^* \to \mathbf{x}^c}} \left|\frac{\partial(\boldsymbol{\alpha}^*, \mathbf{u}^*)}{\partial(\boldsymbol{\alpha}^c, \mathbf{u}^c)}\right|, 1\right\},$$

where $\mathbf{u}^*$ is the auxiliary variable generated to meet the reversibility condition, $\pi(\boldsymbol{\alpha}^* \mid \mathbf{m}_1, \ldots, \mathbf{m}_N)$ indicates the posterior distribution evaluated at the state $\mathbf{x}^*$, $q(\mathbf{u}^*, \boldsymbol{\alpha}^* \mid \mathbf{u}^c, \boldsymbol{\alpha}^c)$ is the proposal distribution for moving from the state $\mathbf{x}^c$ to the state $\mathbf{x}^*$ and $\pi_{\mathbf{x}^c \to \mathbf{x}^*}$ is the probability of choosing

such a move type; $\mathbf{u}^c$, $\pi(\boldsymbol{\alpha}^c \mid \mathbf{m}_1, \ldots, \mathbf{m}_N)$, $q(\mathbf{u}^c, \boldsymbol{\alpha}^c \mid \mathbf{u}^*, \boldsymbol{\alpha}^*)$ and $\pi_{\mathbf{x}^* \to \mathbf{x}^c}$ correspond to the reverse move counterparts and $\left|\frac{\partial(\boldsymbol{\alpha}^*, \mathbf{u}^*)}{\partial(\boldsymbol{\alpha}^c, \mathbf{u}^c)}\right|$ is the Jacobian of the transformation $g_{\mathbf{x}^c \to \mathbf{x}^*}$. In practice, there might be no need to generate any auxiliary variable in one direction or the other. An explanation of the algorithm in pseudocode is provided in the Supplementary Material.

To explore the space of all possible tree structures, we define three move types, described below and represented in Figure 2. Each move type is incorporated into the reversible jump MCMC and chosen with equal probability $\pi_{\mathbf{x}^c \to \mathbf{x}^*} = 1/3$. Given the clustering and dimension reduction properties characterising our procedure, we henceforth refer to this as Clustering and Dimension Reduction (CDR)-MCMC algorithm.

Split Move The split move consists in splitting an existing cluster at random. It is the inverse of the merge move defined below, implying a dimension change transition, for instance, the transition from the state $\mathbf{x}^c = (\alpha_1^c, u^c)^\top$ to $\mathbf{x}^* = (\alpha_1^* = \alpha_1^c + u^c, \alpha_2^* = \alpha_1^c - u^c)^\top$. The proposal distribution for the split move $q(u^c, \alpha_1^c \mid \alpha_1^*, \alpha_2^*)$ is then defined as the density of an auxiliary uniform random variable with boundaries chosen such that $\{\max(0.1, \alpha_1^c - \eta) \leq \alpha_1^*, \alpha_2^* \leq \min(\alpha_1^c + \eta, 1)\}$, with $\eta \in [0, 1]$, an arbitrary constant which, in practice, should be chosen in order to ensure that the chain is well-mixing; in our case we fix $\eta = 0.4$. Here, the Jacobian reduces to $\left|\frac{\partial(\alpha_1^*, \alpha_2^*)}{\partial(\alpha_1^c, u^c)}\right| = 2$.

Merge Move The merge move consists in merging two clusters at random. It is the inverse of the split move, implying a dimension change transition from the state $\mathbf{x}^c = (\alpha_1^c, \alpha_2^c)^\top$ to $\mathbf{x}^* = \left(\alpha_1^* = \frac{\alpha_1^c + \alpha_2^c}{2}, u^* = \frac{\alpha_1^c - \alpha_2^c}{2}\right)^\top$. In this case there is no need to generate any auxiliary uniform random variable and the Jacobian reduces to $\left|\frac{\partial(\alpha_1^*, u^*)}{\partial(\alpha_1^c, \alpha_2^c)}\right| = \frac{1}{2}$.

Swap Move The swap move consists in exchanging some variables from one cluster to another at random. This move type is self-reversible because transitioning from the current state $\mathbf{x}^c = (\alpha_1^c, \alpha_2^c)^\top$ to the proposed state $\mathbf{x}^* = (\alpha_1^*, \alpha_2^*)^\top$ does not involve any parameter dimension change and therefore the Jacobian is simply equal to one.

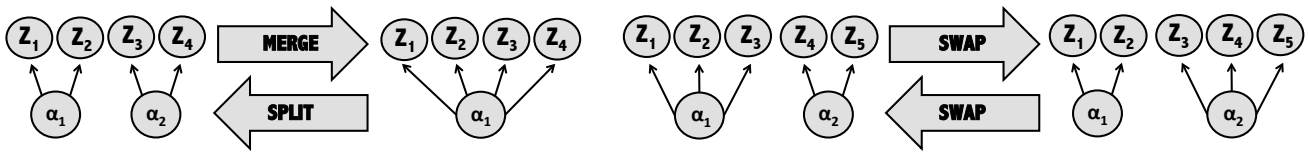

Figure 2: Illustration of the reversible split, merge and swap moves implemented in the CDR-MCMC algorithm.

5 Numerical Experiments

5.1 Simulation settings

The CDR-MCMC algorithm performance is tested on 1000 independent simulated datasets of various dimensions $D = 4, 10, 15$, imposing the dependence structures represented in Figure 3. The data are generated from: the nested logistic distribution (3), as explained by Stephenson (2003); a nested Archimedean copula whose distribution is in the max-domain of attraction (MDA) of the nested logistic distribution; and the Student-t copula with unit Fréchet margins, whose limit max-stable distribution is not the nested logistic model, but rather the extremal t model (Opitz, 2013), thus implying a more severe model misspecification. More specifically, data in the MDA of the nested logistic distribution are sampled using the outer power nested Clayton

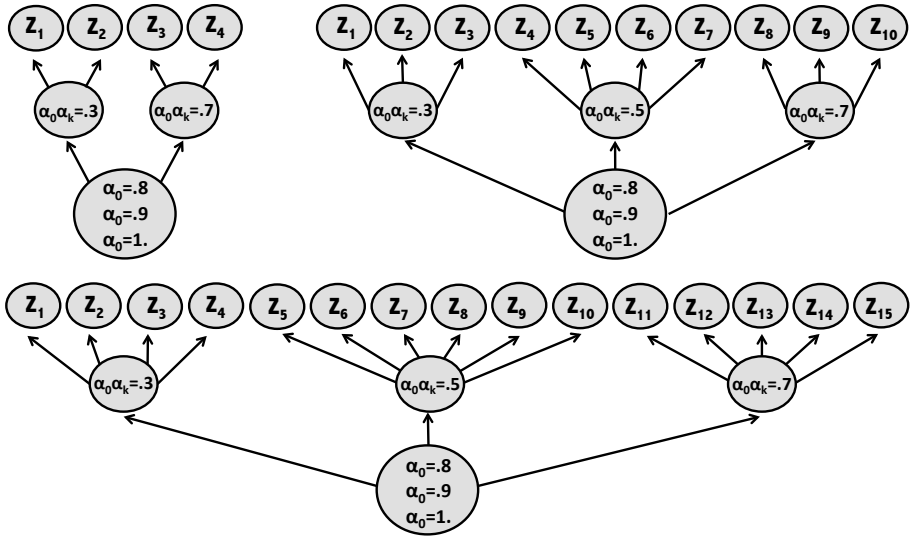

Figure 3: Dependence structure configurations used for the data simulation when $D = 4$ (top left), $D = 10$ (top right) and $D = 15$ (bottom). The coefficients $\alpha_0 = 0.8, 0.9, 1.0$ represent the cases from weak dependence to complete independence between the clusters, and the coefficients $\alpha_0\alpha_k = 0.3, 0.5, 0.7$ represent the cases of strong, mild and weak dependence within the clusters, respectively.

copula with two nesting levels and K child copulas (clusters), *i.e.*,

$$C(\mathbf{u}) = C\{C(\mathbf{u}_1;\varphi_1),\ldots,C(\mathbf{u}_K;\varphi_K);\varphi_0\}, \tag{8}$$

where $\mathbf{u} = (\mathbf{u}_1^\top,\ldots,\mathbf{u}_K^\top)^\top \in [0,1]^D$ and $\varphi_p = (1-t)^{-1/\theta_p}, p = 0,1,\ldots,K$, are Archimedean generators of the Clayton's family (Hofert and Pham, 2013). The parameters $\theta_p \in [1,\infty)$ fulfil the sufficient nesting condition if $\theta_0 \leq \theta_k$, $k = 1,\ldots,K$ and are fixed according to $\theta_0 = 1/\alpha_0$ and $\theta_k = 1/(\alpha_0\alpha_k)$ with $\alpha_0, \alpha_k, k = 1,2,3$, specified in Figure 3. Sampling from such copulas is explained by Hofert (2011). More details about sampling nested Archimedean copulas using `R` can be found in Hofert and Martin (2011). Student-t data are generated from the multivariate Student-t distribution with 10 degrees of freedom, zero mean vector and covariance matrix $\boldsymbol{\Sigma}_{D\times D}(\rho_0)$ as specified below:

$$\boldsymbol{\Sigma}_{4\times4}(\rho_0) = \begin{bmatrix} \mathbf{A}_2(\rho_1) & \rho_0\mathbf{1}_{2\times2} \\ \rho_0\mathbf{1}_{2\times2} & \mathbf{A}_2(\rho_3) \end{bmatrix},$$

$$\boldsymbol{\Sigma}_{10\times10}(\rho_0) = \begin{bmatrix} \mathbf{A}_3(\rho_1) & \rho_0\mathbf{1}_{3\times4} & \rho_0\mathbf{1}_{3\times3} \\ \rho_0\mathbf{1}_{4\times3} & \mathbf{A}_4(\rho_2) & \rho_0\mathbf{1}_{4\times3} \\ \rho_0\mathbf{1}_{3\times3} & \rho_0\mathbf{1}_{3\times4} & \mathbf{A}_3(\rho_3) \end{bmatrix}, \boldsymbol{\Sigma}_{15\times15}(\rho_0) = \begin{bmatrix} \mathbf{A}_4(\rho_1) & \rho_0\mathbf{1}_{4\times6} & \rho_0\mathbf{1}_{4\times5} \\ \rho_0\mathbf{1}_{6\times4} & \mathbf{A}_6(\rho_2) & \rho_0\mathbf{1}_{6\times5} \\ \rho_0\mathbf{1}_{5\times4} & \rho_0\mathbf{1}_{5\times6} & \mathbf{A}_5(\rho_3) \end{bmatrix},$$

where $\mathbf{1}_{n\times m}$ is the $n\times m$ matrix of ones and $\mathbf{A}_n(\rho_k) = (1-\rho_k)\mathbf{I}_n + \rho_k\mathbf{1}_{n\times m}$, $k = 1,2,3$, with $\mathbf{I}_n$ being the $n\times n$ identify matrix and $\rho_k = 0.98, 0.94, 0.86$. The simulation is repeated for different values of $\rho_0 = 0.77, 0.62, 0$. The coefficients ρ_0 and ρ_k, $k = 1,2,3$, are chosen such that the pairwise extremal coefficients θ_2 for the limiting extremal t distribution match the extremal coefficients calculated for the nested logistic model, with respective parameters α_0 and $\alpha_0\alpha_k, k = 1,2,3$, specified in Figure 3, except for $\rho_0 = 0$. Indeed, for $\rho_0 = 0$, $\theta_2 = 1.99$, only approximately 2, which corresponds to the complete independence case of $\alpha_0 = 1$. For each simulation setting under model misspecification we generate datasets of sample sizes $n = 10000, 20000, 40000$ and extract $N = 100, 200, 400$ componentwise maxima, respectively. The simulation results are obtained considering $R = 1500$ algorithm iterations, with $R/5 = 300$ burn-in iterations and fixing the thinning factor to 5. We do not show simulation results for data sampled from the

nested logistic distribution (3) as they lead to the same conclusion as simulation experiments conducted for data simulated from the nested outer power Clayton copula (8).

5.2 Summary of simulation results

The histograms in Figure 4 indicate the percentage of times that each of the tree structure (A, B, C and D) is identified as the mode of the posterior distribution of trees by the CDR-MCMC algorithm, when the recursive likelihood formula (7) is used in each likelihood evaluation, considering data generated from both the copula in (8) and the Student-t distribution. The algorithm is generally able to identify the true data tree structure, represented by tree A (green), or it can identify at least the cluster characterised by strong dependence strength in tree B (yellow). Moreover, when the data are generated from the copula (8), the algorithm performance improves as the value of the α_0 parameter and the sample size increase; meaning that the between-cluster dependence strength plays an important role in identifying the true dependence structure. When data are simulated from the Student-t distribution, the true tree structure is

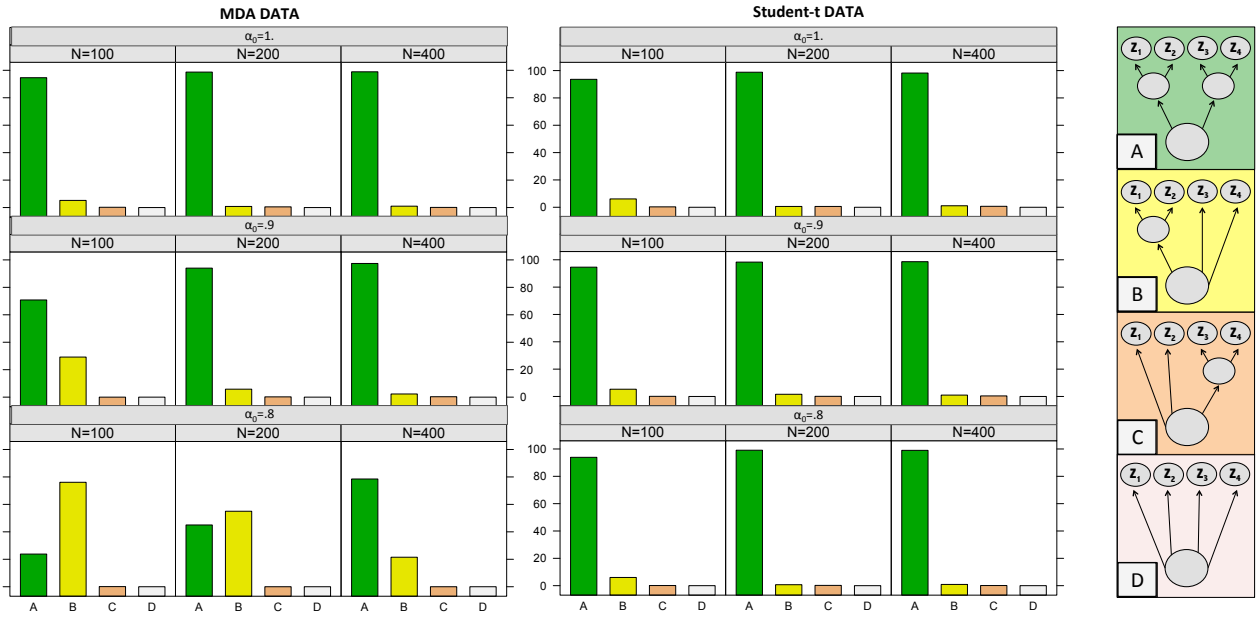

Figure 4: The histograms (left) report the percentage of times a specific tree structure (right) is identified as the mode of the posterior distribution of trees by the CDR-MCMC algorithm after $R = 1500$ iterations over $B = 1000$ replicates for data generated from the copula in (8) and from the Student-t copula with $D = 4$ and tree structure comprising two clusters of variables (see the top-left tree in Figure 3). The recursive formula (7) was used for the likelihood evaluation. The true tree structure is the tree A (green).

mostly identified by the algorithm for any sample sizes. This means that our algorithm works well in misspecified settings.

Figure 5 displays the true positive rates obtained by applying the CDR-MCMC algorithm, when using either the recursive formula (7) or the Stephenson and Tawn likelihood (5), for data simulated from the copula (8) and from the Student-t distribution with $D = 10, 15$. In accordance with the findings for $D = 4$, the algorithm performs reasonably well, identifying all clusters more than 80% of the time for all sample sizes when using the recursive likelihood. The effect of misspecification is more apparent in larger dimensions, but remains quite weak overall. In contrast, the true positive rate is generally lower for all clusters when the Stephenson and Tawn likelihood is implemented, and it slightly decreases as the between-cluster dependence increases.

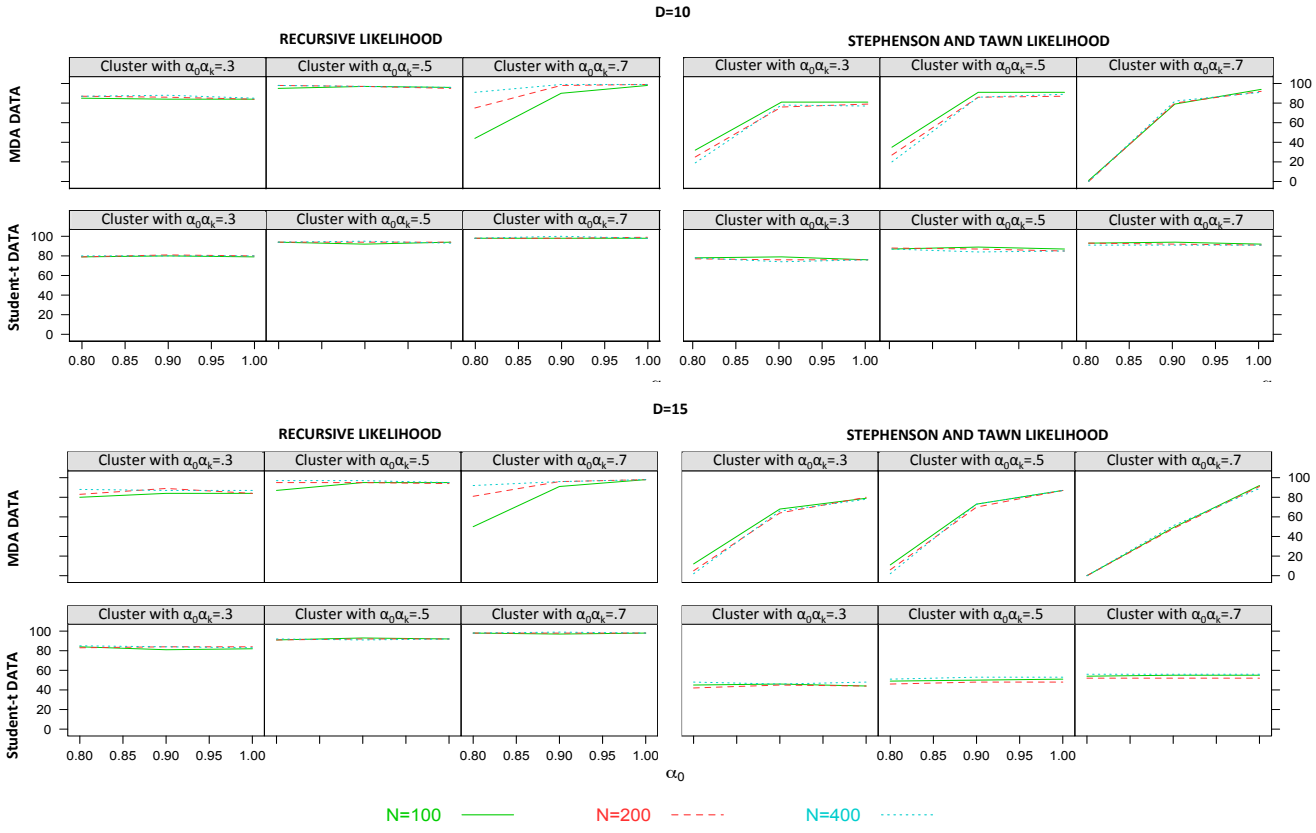

Figure 5: The true positive rate calculated for each cluster is plotted against different values of the parameter α_0 (abscissa) for different sample sizes N (colours) obtained by applying the CDR-MCMC algorithm with $R = 1500$ iterations over $B = 1000$ replicates for data generated from the copula in (8) and from the Student-t copula with $D = 10, 15$ and tree structure comprising three clusters of variables (see the bottom tree in Figure 3) implementing either the recursive formula (7) (left) or the Stephenson and Tawn (5) likelihood (right), respectively.

6 Application to Air Pollution Data

6.1 Air pollutants

Air pollution has various negative effects on human health, ranging from respiratory illnesses to premature death, and causes serious global environmental issues like global warming and ozone depletion. Generally, regional air quality is affected by the topography and weather, but also by the emission sources themselves. Ground level ozone (O_3) is formed by chemical reactions that depend on the mix of volatile organic compounds (VOCs) and nitrogen dioxide (NO_2), in the presence of heat and sunlight (Jacob and Winner, 2009). NO_2, like its precursor nitric oxide (NO), and carbon monoxide (CO) are mostly created by the combustion of fossil fuels in power plants and automobile engines (Cofala *et al.*, 2007). VOCs include a large variety of both natural and artificial chemical species, such as methane and isoprene.

Several organisations, including the United States Environmental Protection Agency (U.S.

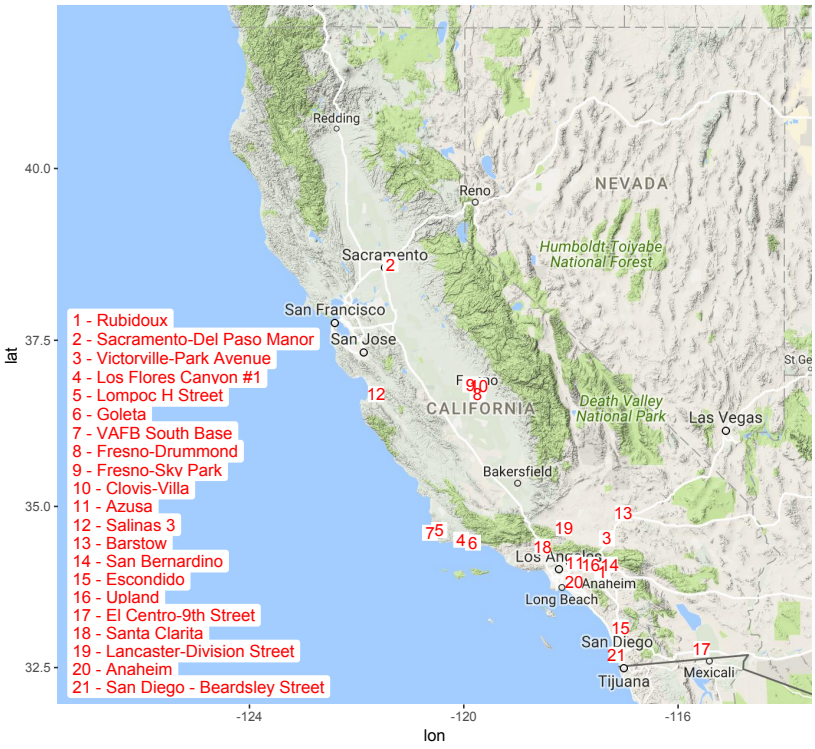

Figure 6: Map of California with the 21 sites under study indicated by numbers.

EPA), have traditionally focused on controlling the emissions of each of the most dangerous air pollutants, also called criteria pollutants, *e.g.*, O_3, NO_2 and CO, separately. However, long-term and peak exposures to multiple air pollutants simultaneously have been demonstrated to cause serious health problems (Dominici *et al.*, 2010; Johns *et al.*, 2012) and so policy-makers worldwide, including the U.S. EPA, are now moving towards a multi-pollutant approach to quantify air pollution risks. There is also growing interest in studying the health effects of the interaction between ozone and temperature, see, *e.g.*, Kahle *et al.* (2015), particularly given that temperatures are expected to rise in the coming decades. In this work, we investigate the dependence relationships between extreme concentrations of air pollutants and extreme weather conditions through max-stable distributions using the CDR-MCMC algorithm described in the previous sections. In particular, we focus on daily observations available from the Air Quality System (AQS) database on the EPA website `http://aqsdr1.epa.gov/aqsweb/aqstmp/airdata/download_files.html` collected at 21 sites across the state of California, which is one of the most populated and polluted areas of the US. The site locations can be visualised in Figure 6. Details on the data preprocessing and marginal estimations are available in the Supplementary Material.

6.2 Multivariate analysis of California air pollution extremes

Considering daily observation of the variables CO, NO, NO_2, O_3 and temperature (T) collected from January 2004 to December 2015, we derive at least 100 multivariate monthly maxima for each site in Figure 6 to which we apply the CDR-MCMC algorithm with $R = 1500$ iterations, burn-in $R/5 = 300$ and thinning factor 5. The most frequent dependence structures identified by the algorithm and the corresponding posterior probabilities obtained for each of the sites under study are represented in Figure 7. In particular, the most common trees that are identified as the mode of the posterior distribution of trees by the CDR-MCMC algorithm are: tree A, for 43% of the sites, that includes only the CO-NO cluster; and tree B, for 33% of the sites, that includes two clusters, respectively grouping the maxima of NO and CO and the maxima of NO_2

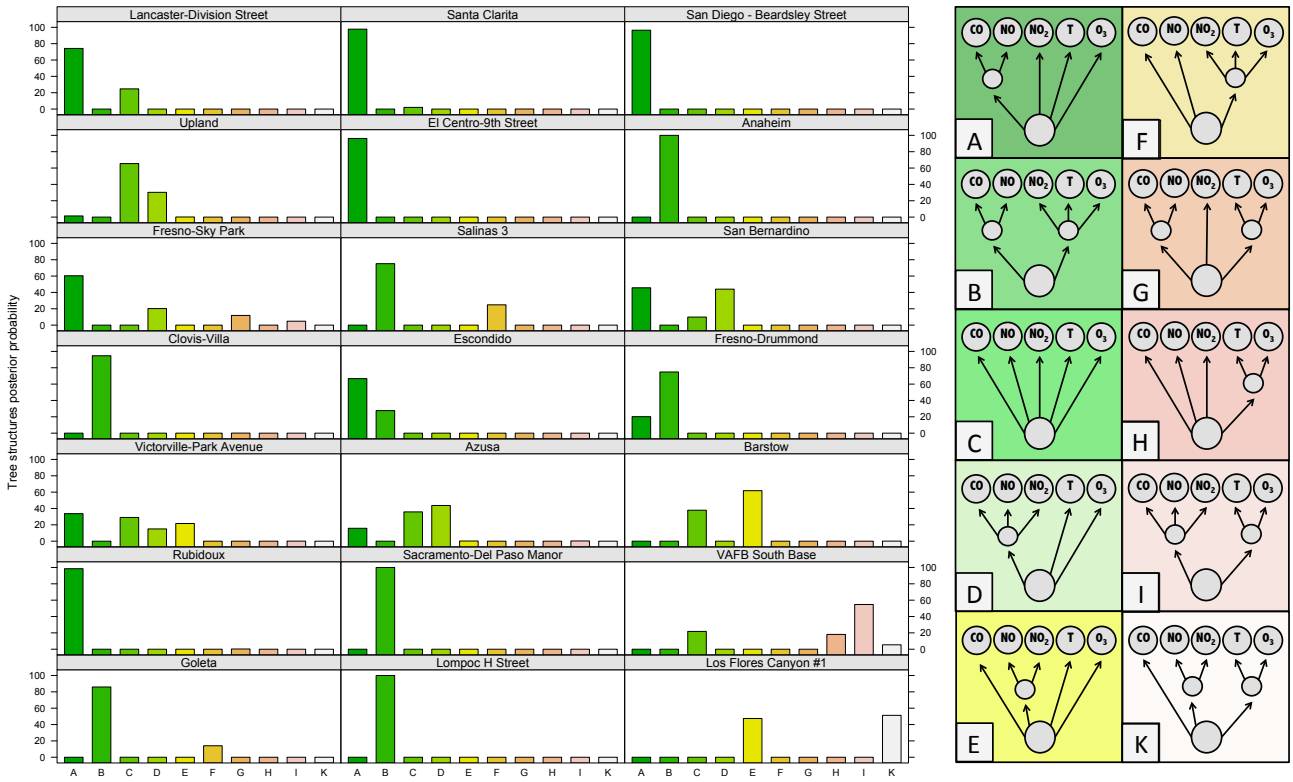

Figure 7: A different letter is associated to each of the most frequent tree structures (right) identified by the CDR-MCMC algorithm after $R = 1500$ iterations, burn-in $R/5 = 300$ and thinning factor 5. The histograms (left) report the posterior probability associated to each tree for each site under study calculated according to the number of times each tree appears in the algorithm chain.

and O_3 with T. Interestingly, the data dependence structure cannot always be represented by one specific tree, for instance the CDR-MCMC algorithm suggests to represent the extremal dependence structure at Victorville-Park Avenue using a mixture of the trees A, C, D and E.

Figure 8 represents the point estimates for the nested logistic model parameters $\boldsymbol{\alpha}$, considering the tree structures A and B in Figure 7. The values of the parameter α_0 represented in the left map are generally close to $\alpha_0 = 0.7$ for the southern sites, whereas α_0 tends to be close to 1 for the northern sites, indicating that most of the identified clusters can be treated independently. As expected, the dependence strength within the clusters is stronger. The maxima of NO and CO are found to be particularly dependent in areas characterised by heavy traffic, especially between the cities of Anaheim and Long Beach and close to Los Angeles and San Diego. The dependence strength between the maxima of NO_2-O_3 and T is generally mild or weak for the inland sites, but increases for the sites near the coast.

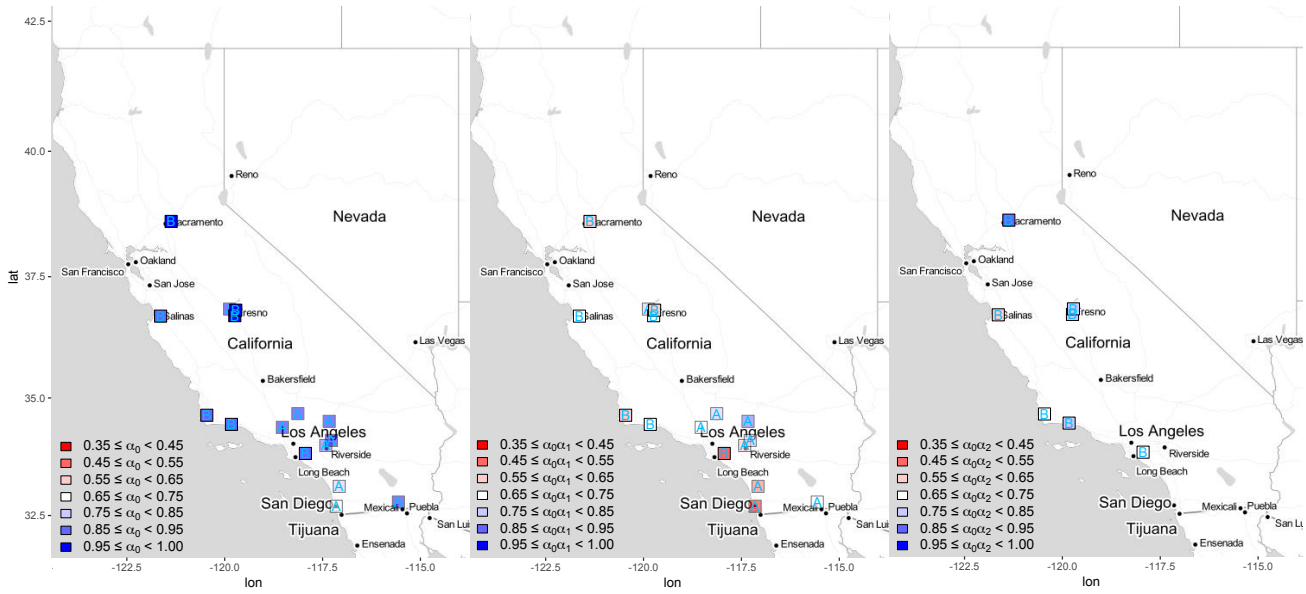

Figure 8: A colour scale is used to represent the point estimates of the nested logistic model (3) parameters computed as the median of the sub-chain corresponding to the most likely tree after burn-in and thinning. The parameter α_0 represents the dependence between the clusters (left map) and the parameters $\alpha_0\alpha_1$, $\alpha_0\alpha_2$ represent the dependence within the clusters CO-NO (central map) and O_3-NO_2-T (right map), respectively.

6.3 Analysis of Clovis-Villa air pollution extremes

We now include in our analysis the monthly maxima of relative humidity (RH), barometric pressure (BP) and wind speed (WS), together with concentration maxima of non-methane organic compounds (NM), which are essentially VOCs without methane. For illustrative purposes, we focus on the site named Clovis-Villa, located in Fresno, for which we were able to derive 54 multivariate monthly maxima, collected from January 2006 to December 2014. The two tree structures identified by the CDR-MCMC algorithm are illustrated in Figure 9. In accordance with previous findings, our algorithm groups the maxima of NO and CO through 98% of the chain, and the maxima of O_3, NO_2 and T, now combined with RH, through 84% of the chain.

The algorithm output is provided in Figure 10. The sub-chains seem stationary from the trace-plots on the left panels and the autocorrelation functions on the right panels hint towards good sub-chain's mixing properties. In particular, the posterior median of the parameter α_0 estimated by the algorithm is exactly 1, indicating that the two clusters of variables, as well as the other single variables, can be treated independently. Therefore, the algorithm allows us to

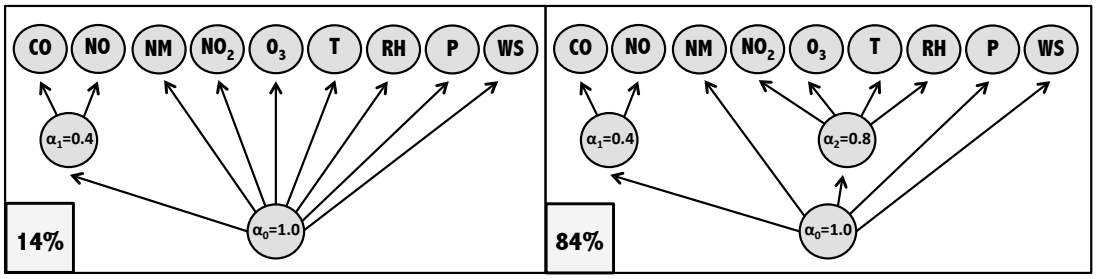

Figure 9: Dependence structures identified by the CDR-MCMC algorithm after $R = 5000$ iterations and the associated posterior probability for Clovis-Villa.

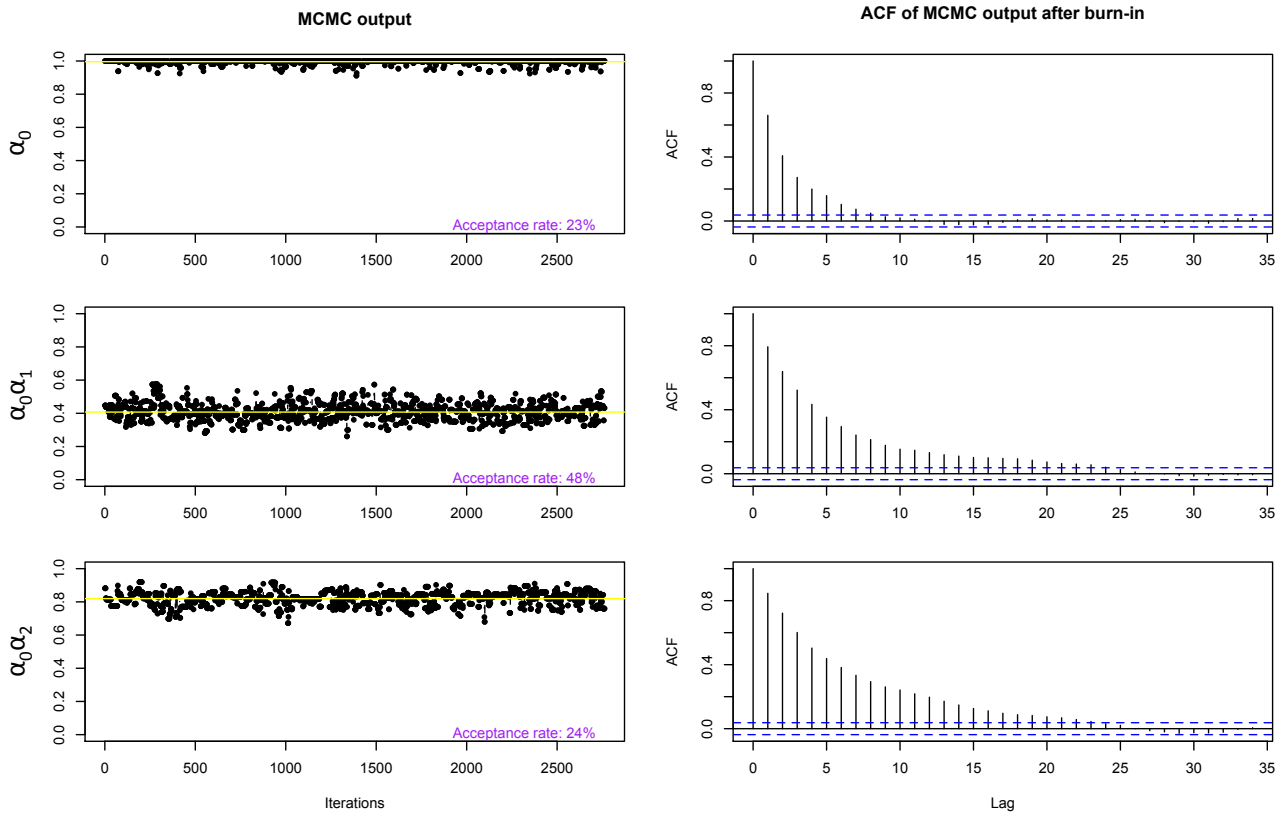

Figure 10: Trace-plots (left) and the autocorrelation functions (right) for the sub-chain corresponding to the tree structure represented in Figure 9, with posterior probability of 84% obtained by applying the CDR-MCMC algorithm after $R = 5000$ iterations, thinned by a factor 5 after a burn-in of $R/5$ iterations. The posterior medians are represented by yellow lines.

significantly reduce the complexity of the data describing their dependence structure by using only three parameters, or even two if $\alpha_0 = 1$, for the nine variables considered here. In contrast, fitting the bivariate logistic model to all possible pairs of variables yields 36 estimated dependence parameters, as shown in Figure 11. The bivariate fits suggest a relation of strong dependence between the maxima of NO and CO and of moderate dependence between the maxima of O_3, NO_2,

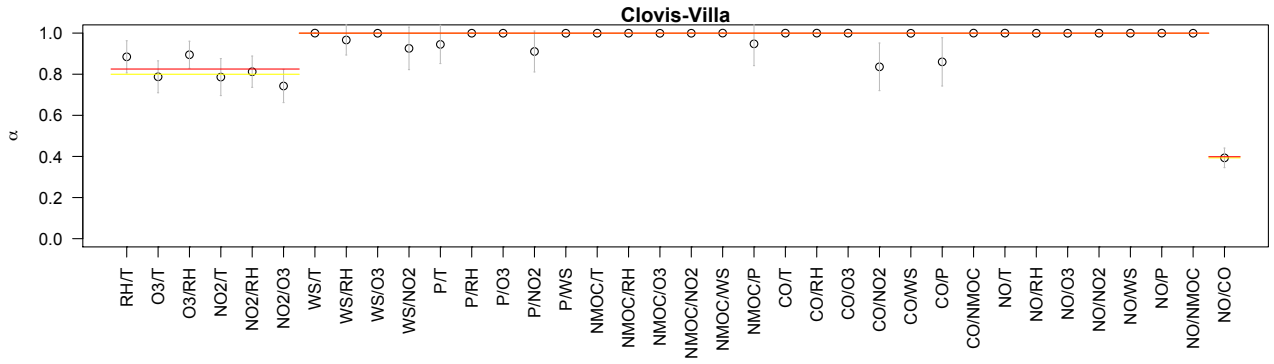

Figure 11: Parameters estimates (circles) and 95% credible intervals (vertical segments) resulting from logistic model bivariate fits for each pair of pollutants and meteorological parameters. The red lines represent the median of the parameter estimates and the yellow lines the posterior medians obtained from the multivariate fit using the CDR-MCMC algorithm after $R = 5000$ iterations and burn-in $R/5$ iterations with thinning factor 5.

T and RH, as indicated by the point estimates obtained from the algorithm output. Interestingly, the joint fits match the bivariate fits almost perfectly, except for a few pairs of variables with high variability.

The Air Quality Index (AQI) is a measure typically used by the U.S. EPA to communicate the level of air pollution to the public. When multiple criteria pollutants are measured at the same location, the EPA reports the largest AQI value as a measure of air quality; for more details see, *e.g.*, Airnow (2016). The AQI is divided into different categories that indicate different levels of health concern. Using our algorithm, we are able to sample from the posterior predictive distribution of the AQI values for the maxima of the criteria pollutants CO, O_3 and NO_2, taking into account the uncertainty associated with the trees summarising the dependence relations between these pollutants, meteorological parameters and other air pollutants. In Figure 12 we display high p-quantiles with probabilities ranging from $p = 0.5$ to $p = 0.996$, considering August 2006 and August 2014 as baselines, computed for the smallest, the average and the largest AQI monthly maxima for CO, O_3 and NO_2. Notice that the values of $p = 1 - 1/12 \approx 0.917$ and $p = 1 - 1/(12 \times 20) \approx 0.996$ roughly correspond to 1 and 20 year-return levels, respectively, under stationary conditions. The AQI categories are represented by different colours. For comparison purposes, we also computed high quantiles for the site named Victorville-Park Avenue. Since

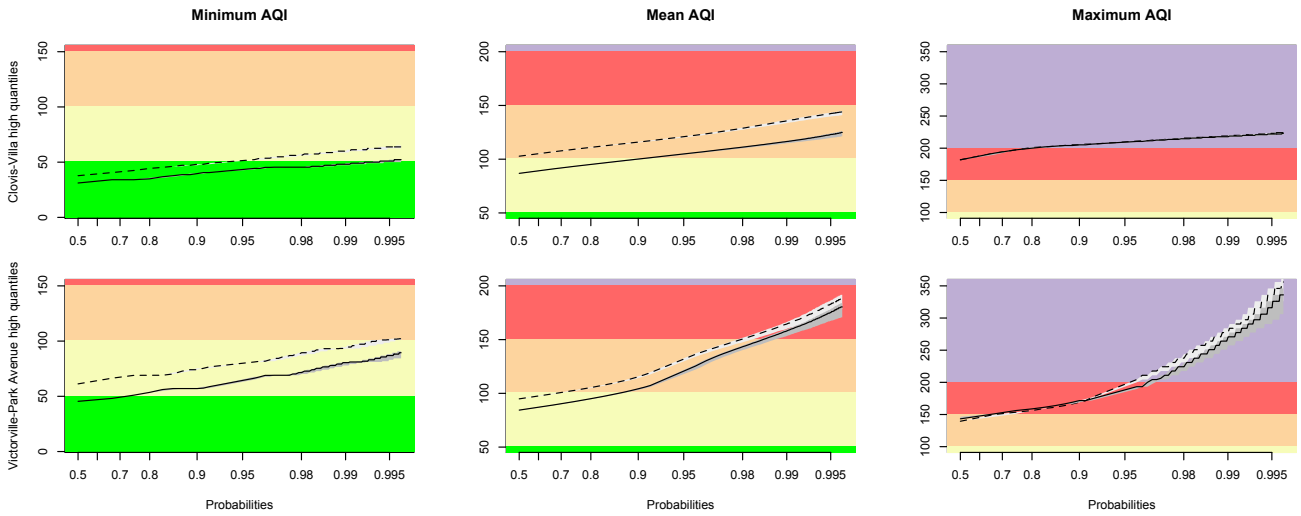

Figure 12: High p-quantiles z_p computed for the minimum (left), the average (center) and the maximum (right) AQI setting August 2006 (dashed lines) or August 2014 (solid lines) as baselines with 95% bootstrap credible bands (grey areas) for CO, O_3 and NO_2 indexes, obtained for the Clovis-Villa (top) and Victorville-Park Avenue (bottom) site applying the CDR-MCMC algorithm after $R = 50000$ iterations and burn-in $R/5$ iterations with thinning factor 5. AQI categories: 0-50 satisfactory (green); 51-100 acceptable (yellow); 101-150 unhealthy for sensitive groups (orange); 151-200 unhealthy (red); >200 very unhealthy (purple). Probabilities are displayed on a Gumbel scale, i.e., z_p is plotted against $-\log\{-\log(p)\}$.

we estimate the dependence structure separately from the margins, the 95% bootstrap credible intervals solely reflect the model uncertainty associated with the clustering, illustrating the ability of the CDR-MCMC algorithm to realistically account for the model uncertainty in predictions. In practice, it might be better to adopt a fully Bayesian approach aggregating the margins and dependence structure uncertainties. Interestingly, the 95% bootstrap credible bands are very narrow in the case of the Clovis-Villa site. Indeed, the two plausible dependence structures identified by the CDR-MCMC algorithm only differ for the cluster formed by the maxima of O_3, NO_2, T and RH, as illustrated in Figure 9. On the other hand, as shown in Figure 7, the site Victorville-Park Avenue is characterised by a higher model uncertainty, which reflects on much wider credible bands for the high quantiles. The high p-quantiles projections calculated based on August 2006 are generally larger than the high quantiles based on August 2014. For 2014, the minimum AQI exceeds the satisfactory level of 50 with probability about 0.5% at Clovis-Villa and around 20% at Victorville-Park Avenue, i.e., once every 15-20 years and 3 times per year on average, respectively, under stationarity. Moreover, the average AQI generally lies within the

unhealthy category only for sensitive groups of people in Clovis-Villa whereas at Victorville-Park Avenue it is expected to exceed this category with probability about 1.5%, so roughly once every 5 years on average. The maximum AQI high quantiles generally lie within the very unhealthy category, indicating that at least one of the criteria pollutants under study exceeds this most critical threshold with probability about 15% at Clovis-Villa and around 2.5% at Victorville-Park Avenue, and therefore 2 times per year and once every 3 or 4 years on average, respectively.

7 Discussion

We proposed a novel clustering and dimension reduction algorithm for multivariate extremes based on the hierarchical dependence structure of the nested logistic model and Bayesian computational tools. Our algorithm has the advantage of allowing the identification of model parameters and of performing model selection simultaneously, providing a Bayesian assessment of the model uncertainty. Numerical experiments show that the algorithm is mostly able to identify the true dependence structure from the data, even when the nested logistic model is misspecified.

The CDR-MCMC algorithm was applied to air pollution concentration maxima collected at 21 sites across California. A particularly strong dependence relation between the maxima of CO and NO was found for most of the sites located between or within big cities, possibly because both pollutants are released by motor vehicles. Moreover, many of the sites close to coastal areas present a strong to mild dependence relation between the maxima of O_3 and NO_2 and high temperatures. This is expected as O_3 generally forms in chemical reactions that depend on the presence of NO_2 and heat. Fitting the bivariate logistic model to all possible pairs of variables for the Clovis-Villa site in Fresno yields similar conclusions, verifying that, despite its simplicity, the nested logistic model was flexible enough to provide a good estimation of the data dependence structure. The AQI high quantiles indicate that air quality is generally expected to exceed the healthy threshold within the next two decades under stationarity. Further efforts to reduce emissions seem necessary in order to protect public health, especially given that longer-

term climatic changes projections predict rising temperatures. Our method could be used for obtaining air quality measures that take into account the extremes of multiple pollutants and the public health risks associated with their exposure.

When fitting the nested logistic model or the nested Gumbel copula, its within-cluster exchangeability may be seen as a limitation. However, embedding the nested logistic distribution within a Bayesian framework has the great advantage of describing the data dependence structure using a mixture of trees, allowing therefore to capture complex dependence relationship between variables. A similar procedure might be applied to threshold exceedances instead of maxima by implementing a censored recursive likelihood function, which may lead to further computational improvements. It would also be interesting to explore spatial extensions of our approach.

References

Airnow (2016) Air Quality Index (AQI) Basics. *https://airnow.gov/index.cfm?action=aqibasics.aqi*

Bernard, E., Naveau, P., Vrac, M. and Mestre, O. (2013) Clustering of maxima: Spatial dependencies among heavy rainfall in france. *Journal of Climate* **26**(20), 7929–7937.

Bienvenüe, A. and Robert, C. (2015) Likelihood inference for multivariate extreme value distributions whose spectral vectors have known conditional distributions. *Scandinavian Journal of Statistics*. To appear.

Castruccio, S., Huser, R. and Genton, M. G. (2016) High-order composite likelihood inference for max-stable distributions and processes. *Journal of Computational and Graphical Statistics* **25**(4), 1212–1229.

Chautru, E. (2015) Dimension reduction in multivariate extreme value analysis. *Electronic Journal of Statistics* **9**, 383–418.

Chavez-Demoulin, V. and Davison, A. C. (2012) Modelling time series extremes. *Revstat Statistical Journal* **10**(1), 109–133.

Cofala, J., Amann, M., Klimont, Z., Kupiainen, K. and Hoglund-Isaksson, L. (2007) Scenarios of global anthropogenic emissions of air pollutants and methane until 2030. *Atmospheric Environment* **41**(38), 8486–8499.

Coles, B. S. and Pauli, F. (2002) Models and inference for uncertainty in extremal dependence. *Biometrika* **89**(1), 183–196.

Coles, S. G. and Tawn, J. A. (1991) Modelling extreme multivariate events. *Journal of the Royal Statistical Society. Series B* **53**(2), 377–392.

Cooley, D. and Thibaud, E. (2016) Principal component decomposition and completely positive decomposition of dependence for multivariate extremes. *https://arxiv.org/abs/1612.07190* Submitted.

Davison, A. and Huser, R. (2015) Statistics of extremes. *Annual Review of Statistics and Its Application* **2**, 203–235.

Davison, A. C. and Gholamrezaee, M. M. (2012) Geostatistics of extremes. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **468**, 581–608.

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

Dominici, F., Peng, R. D., Barr, C. D. and Bell, M. L. (2010) Protecting human health from air pollution: shifting from a single-pollutant to a multi-pollutant approach. *Epidemiology* **21**(2), 187–194.

Genton, M. G., Ma, Y. and Sang, H. (2011) On the likelihood function of Gaussian max-stable processes. *Biometrika* **98**(2), 481–488.

Górecki, J., Hofert, M. and Holeňa, M. (2016) An approach to structure determination and estimation of hierarchical Archimedean copulas and its application to Bayesian classification. *Journal of Intelligent Information Systems* **46**(1), 21–59.

Graham, R., Knuth, D. E. and Patashnik, O. (1988) *Concrete Mathematics*. Reading MA: Addison-Wesley.

Green, P. J. (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, 711–732.

Guillotte, S., Perron, F. and Segers, J. (2011) Non-parametric bayesian inference on bivariate extremes. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **73**(3), 377–406.

Gumbel, E. J. (1960a) Distributions des valeurs extrêmes en plusieurs dimensions. *Publications de l'Institute de Statistique de l'Université de Paris* **9**(294), 171–173.

Gumbel, E. J. (1960b) Bivariate exponential distributions. *Journal of the American Statistical Association* **55**, 698–707.

de Haan, L. (1984) A spectral representation for max-stable processes. *The Annals of Probability* **12**(4), 1194–1204.

de Haan, L. and Resnick, S. I. (1977) Limit theory for multivariate sample extremes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **40**, 317–337.

Hasting, W. K. (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109.

Haug, S., Claudia, K. and Kuhn, G. (2015) Copula structure analysis based on extreme dependence. *Statistics and Its Interface* **8**(1), 93–107.

Hofert, M. (2011) Efficiently sampling nested Archimedean copulas. *Computational Statistics and Data Analysis* **55**(1), 57–70.

Hofert, M. and Martin, M. (2011) Nested archimedean copulas meet R : The nacopula package. *Journal of Statistical Software* **39**(9), 1–20.

Hofert, M. and Pham, D. (2013) Densities of nested Archimedean copulas. *Journal of Multivariate Analysis* **118**, 37–52.

Huser, R. and Davison, A. (2013) Composite likelihood estimation for the Brown-Resnick process. *Biometrika* **100**, 511–518.

Huser, R., Davison, A. C. and Genton, M. G. (2016) Likelihood estimators for multivariate extremes. *Extremes* **19**(1), 79–103.

Jacob, D. and Winner, D. (2009) Effect of climate change on air quality. *Atmospheric Environment* **43**(1), 51–63.

Johns, D. O., Wichers Stanek, L., Walker, K., Benromdhane, S., Hubbell, B., Ross, M., Devlin, R. B., Costa, D. L. and Greenbaum, D. S. (2012) Practical advancement of multipollutant scientific and risk assessment approaches for ambient air pollution. *Environ Health Perspect* **120**(9).

Kahle, J. J., Neas, L. M., Devlin, R. B., Case, M. W., Schmitt, M. T., Madden, M. C. and Diaz-Sanchez, D. (2015) Interaction effects of temperature and ozone on lung function and markers of systemic inflammation, coagulation, and fibrinolysis: A crossover study of healthy young volunteers. *Environ Health Perspect* **123**(4).

McFadden, D. L. (1978) Modeling the choice of residental location. In *Spatial Interaction Theory and Planning Models (eds A. Karlquist et al.)*, pp. 75–96. North-Holland, Amsterdam.

Okhrin, O., Okhrin, Y. and Schmid, W. (2013a) Properties of Hierarchical Archimedean Copulas. *Statistics and Risk Modelling with Applications in Finance and Insurance* **30**(1), 21–54.

Okhrin, O., Okhrin, Y. and Schmid, W. (2013b) On the structure and estimation of hierarchical archimedean copulas. *Journal of Econometrics* **173**(2), 189 – 204.

Opitz, T. (2013) Extremal t processes: Elliptical domain of attraction and a spectral representation. *Journal of Multivariate Analysis* **122**, 409–413.

Padoan, S. A., Ribatet, M. and Sisson, S. A. (2010) Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association* **105**(489), 263–277.

Reich, B. J. and Shaby, B. A. (2012) A hierarchical max-stable spatial model for extreme precipitation. *Annals of Applied Statistics* **6**(4), 1430–1451.

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

Sabourin, A. and Naveau, P. (2014) Bayesian Dirichlet mixture model for multivariate extremes: A re-parametrization. *Computational Statistics and Data Analysis* **71**, 542 – 567.

Sabourin, A., Naveau, P. and Fougères, A.-L. (2013) Bayesian model averaging for multivariate extremes. *Extremes* **16**(3), 325–350.

Segers, J. and Uyttendaele, N. (2014) Nonparametric estimation of the tree structure of a nested archimedean copula. *Computational Statistics and Data Analysis* **72**, 190 – 204.

Shaby, B. A. (2014) The open-faced sandwich adjustment for MCMC using estimating functions. *Journal of Computational and Graphical Statistics* **23**, 853–876.

Shi, D. (1995) Fisher information for a multivariate extreme value distribution. *Biometrika* **82**, 644–649.

Smith, R. L. (1990) Max-stable processes and spatial extremes. *Unpublished manuscript* Unpublished manuscript. University of North Carolina, Chapel Hill, U.S.

Stephenson, A. (2003) Simulating multivariate extreme value distributions of logistic type. *Extremes* **6**(1), 49–59.

Stephenson, B. A. and Tawn, J. (2005) Exploiting occurence times in likelihood inference for componetwise maxima. *Biometrika* **92**(1), 213–227.

Tawn, J. A. (1990) Modelling multivariate extreme value distributions. *Biometrika* **77**(2), 245–253.

Thibaud, E., J. Aalto, D. S. C., Davison, A. C. and Heikkinen, J. (2016) Bayesian inference for the Brown-Resnick process, with an application to extreme low temperatures. *Annals of Applied Statistics.* **10**(4), 2303–2324.

Varin, B. C. and Vidoni, P. (2005) A note on composite likelihood inference and model selection. *Biometrika* **92**(3), 519–528.

Varin, C. (2008) On composite marginal likelihoods. *AStA Advances in Statistical Analysis* **92**(1), 1–28.

Wadsworth, J. L. (2014) On the occurrence times of componentwise maxima and bias in likelihood inference for multivariate max-stable distributions. *Biometrika* **102**(3), 705–711.

A Recursive formulas for the nested logistic distribution

A.1 Notation

For simplicity, we denote the distribution of the nested logistic model 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, one has omitted the function arguments. Moreover, we use the following vector notation:

$$\begin{aligned}
\boldsymbol{z}_k &= (z_{k;1}, \ldots, z_{k;D_k})^\top && \textit{All variables in cluster } k = 1, \ldots, K; \\
\boldsymbol{z}_{k,1:d_k} &= (z_{k;1}, \ldots, z_{k;d_k})^\top && \textit{sub-vector of the first } d_k \textit{ variables in cluster } k = 1, \ldots, K; \\
\boldsymbol{i}_{1:\kappa} &= (i_1, \ldots, i_\kappa)^\top && \textit{Vector of } \kappa \textit{ indices.}
\end{aligned}$$

A.2 Preliminary results

From the definition in Section A.1, one deduces the following derivatives:

$$\frac{\partial V_k}{\partial z_{k;i_k}} = \alpha_k \left\{ \sum_{i_k=1}^{D_k} z_{k;i_k}^{-1/(\alpha_0\alpha_k)} \right\}^{\alpha_k - 1} \times \frac{-1}{\alpha_0\alpha_k} z_{k;i_k}^{-1/(\alpha_0\alpha_k)-1} = -\frac{1}{\alpha_0} z_{k;i_k}^{-1/(\alpha_0\alpha_k)-1} V_k^{1-1/\alpha_k}; \quad (9)$$

$$\frac{\partial V}{\partial z_{k;i_k}} = \alpha_0 \left(\sum_{k=1}^{K} V_k \right)^{\alpha_0 - 1} \frac{\partial V_k}{\partial z_{k;i_k}} = -z_{k;i_k}^{-1/(\alpha_0\alpha_k)-1} V_k^{1-1/\alpha_k} V^{1-1/\alpha_0}; \quad (10)$$

$$\frac{\partial G}{\partial z_{k;i_k}} = -\frac{\partial V}{\partial z_{k;i_k}} \exp(-V) = z_{k;i_k}^{-1/(\alpha_0\alpha_k)-1} G V_k^{1-1/\alpha_k} V^{1-1/\alpha_0}. \quad (11)$$

A.3 Partial derivatives of G

The distribution G is a function of $D = \sum_{k=1}^{K} D_k$ variables, namely $\boldsymbol{z}_1 = (z_{1;1}, \ldots, z_{1;D_1})^\top, \ldots, \boldsymbol{z}_K = (z_{K;1}, \ldots, z_{K;D_K})^\top$. The partial derivative of G 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

$$\frac{\partial^{\sum_{k=1}^{\kappa} d_k} G}{\partial \prod_{k=1}^{\kappa} \mathbf{z}_{k;1:d_k}} = G \prod_{i_1=1}^{d_1} z_{1;i_1}^{-\frac{1}{\alpha_0\alpha_1}-1} \cdots \prod_{i_\kappa=1}^{d_\kappa} z_{\kappa;i_\kappa}^{-\frac{1}{\alpha_0\alpha_\kappa}-1} \sum_{i_1=1}^{d_1} \cdots \sum_{i_\kappa=1}^{d_\kappa} \sum_{j=1}^{\sum_{k=1}^{\kappa} i_k} \beta_{i_1;\ldots;i_\kappa;j}^{(d_1;\ldots;d_\kappa)} V_1^{i_1 - \frac{d_1}{\alpha_1}} \cdots V_\kappa^{i_\kappa - \frac{d_\kappa}{\alpha_k}} V^{j - \frac{\sum_{k=1}^{\kappa} i_k}{\alpha_0}} \quad (12)$$

where the coefficients $\beta_{i_1;\ldots;i_\kappa;j}^{(d_1;\ldots;d_\kappa)}$ can be computed recursively as demonstrated below. By specialising the expression in (12) to $\kappa = K$ clusters and $d_1 = D_1, \ldots, d_K = D_K$ variables, one obtains the density, or full likelihood, for one replicate.

Proof Equation (12) 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 $\beta_{i_1;\ldots;i_\kappa;j}^{(d_1;\ldots;d_\kappa)}$. More precisely, we demonstrate the following four steps:

1. (12) holds for $\kappa = 1$ and $d_1 = 1$;
2. If (12) 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 (12) holds for $\kappa \in \{1, \ldots, K - 1\}$, then it also holds for $\kappa \mapsto \kappa + 1$ with $d_{\kappa+1} = 1$;
4. If (12) 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 (11) one has

$$\frac{\partial G}{\partial z_{1;1}} = z_{1;1}^{-1/(\alpha_0\alpha_1)-1} G V_1^{1-1/\alpha_1} V^{1-1/\alpha_0},$$

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

Step 2. Assuming that (12) holds for $\kappa = 1$ and $d_1 \in \{1, \dots, D_1 - 1\}$, and using (9), (10) and (11), one obtains

$$\frac{\partial^{d_1+1} G}{\partial \mathbf{z}_{1;1:d_1+1}} = \prod_{i_1=1}^{d_1} z_{1;i_1}^{-\frac{1}{\alpha_0\alpha_1}-1} \left\{ \frac{\partial G}{z_{1;d_1+1}} \sum_{i_1=1}^{d_1}\sum_{j=1}^{i_1} \beta_{i_1;j}^{(d_1)} V_1^{i_1-\frac{d_1}{\alpha_1}} V^{j-\frac{i_1}{\alpha_0}} + G \sum_{i_1=1}^{d_1}\sum_{j=1}^{i_1} \beta_{i_1;j}^{(d_1)} \frac{\partial V_1^{i_1-\frac{d_1}{\alpha_1}}}{z_{1;d_1+1}} V^{j-\frac{i_1}{\alpha_0}} \right.$$
$$\left. + G \sum_{i_1=1}^{d_1}\sum_{j=1}^{i_1} \beta_{i_1;j}^{(d_1)} V_1^{i_1-\frac{d_1}{\alpha_1}} \frac{\partial V^{j-\frac{i_1}{\alpha_0}}}{z_{1;d_1+1}} \right\}$$

$$= G \prod_{i_1=1}^{d_1+1} z_{1;i_1}^{-\frac{1}{\alpha_0\alpha_1}-1} \left\{ \sum_{i_1=1}^{d_1}\sum_{j=1}^{i_1} \beta_{i_1;j}^{(d_1)} V_1^{(i_1+1)-\frac{(d_1+1)}{\alpha_1}} V^{(l+1)-\frac{(i_1+1)}{\alpha_0}} \right.$$
$$- \sum_{i_1=1}^{d_1}\sum_{j=1}^{i_1} \beta_{i_1;j}^{(d_1)} \left(i_1 - \frac{d_1}{\alpha_1}\right) \left(\frac{1}{\alpha_0}\right) V_1^{i_1-\frac{d_1+1}{\alpha_1}} V^{j-\frac{i_1}{\alpha_0}} - \sum_{i_1=1}^{d_1}\sum_{j=1}^{i_1} \beta_{i_1;j}^{(d_1)} \left(l - \frac{i_1}{\alpha_0}\right) V_1^{(i_1+1)-\frac{d_1+1}{\alpha_1}} V^{j-\frac{(i_1+1)}{\alpha_0}} \Bigg\}$$
$$= G \prod_{i_1=1}^{d_1+1} z_{1;i_1}^{-\frac{1}{\alpha_0\alpha_1}-1} \left\{ \sum_{i_1=2}^{d_1+1}\sum_{l=2}^{i_1} \beta_{i_1-1;j-1}^{(d_1)} V_1^{i_1-\frac{(d_1+1)}{\alpha_1}} V^{j-\frac{i_1}{\alpha_0}} - \sum_{i_1=1}^{d_1}\sum_{j=1}^{i_1} \beta_{i_1;j}^{(d_1)} \left(i_1 - \frac{d_1}{\alpha_1}\right) \left(\frac{1}{\alpha_0}\right) V_1^{i_1-\frac{d_1+1}{\alpha_1}} V^{j-\frac{i_1}{\alpha_0}} \right.$$
$$\left. - \sum_{i_1=2}^{d_1+1}\sum_{j=1}^{i_1} \beta_{i_1-1;j}^{(d_1)} \left(l - \frac{i_1-1}{\alpha_0}\right) V_1^{i_1-\frac{d_1+1}{\alpha_1}} V^{j-\frac{i_1}{\alpha_0}} \right\} = G \prod_{i_1=1}^{d_1+1} z_{1;i_1}^{-\frac{1}{\alpha_0\alpha_1}-1} \sum_{i_1=1}^{d_1+1}\sum_{j=1}^{i_1} \beta_{i_1;j}^{(d_1+1)} V_1^{i_1-\frac{d_1+1}{\alpha_1}} V^{j-\frac{i_1}{\alpha_0}},$$

where

$$\beta_{i_1;j}^{(d_1+1)} = \beta_{i_1-1;j-1}^{(d_1)} - \frac{1}{\alpha_0}\left(i_1 - \frac{d_1}{\alpha_1}\right)\beta_{i_1;j}^{(d_1)} - \left(l - \frac{i_1-1}{\alpha_0}\right)\beta_{i_1-1;j}^{(d_1)}, \quad 1 \le j \le i_1 \le d_1 + 1, \tag{13}$$

with

$$\beta_{i_1;j}^{(d_1)} = 0, \qquad \text{for all} \quad i_1 \notin \{1, \dots, d_1\}, \quad \text{or} \quad j \notin \{1, \dots, i_1\}. \tag{14}$$

Hence, (12) holds by induction for $\kappa = 1$ and any $1 \le d_1 \le D_1$, and the recursive formula to compute coefficients $\beta_{i_1;j}^{(d_1)}$ is given by (13) and (14) with the initial condition $\beta_{1;1}^{(1)} = 1$.

Step 3. Assuming that (12) holds for $\kappa \in \{1,\dots,K-1\}$, one obtains

$$\begin{aligned}
&\frac{\partial^{1+\sum_{k=1}^{\kappa} d_k} G}{\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} \left\{ \frac{\partial G}{\partial z_{\kappa+1;1}} \sum_{i_1=1}^{d_1} \cdots \sum_{i_\kappa=1}^{d_\kappa} V_1^{i_1-\frac{d_1}{\alpha_1}} \cdots V_k^{i_\kappa-\frac{d_\kappa}{\alpha_k}} \right. \\
&\left. \times \sum_{j=1}^{i_1+\cdots+i_\kappa} \beta_{i_1;\dots;i_\kappa;j}^{(d_1;\dots;d_\kappa)} V^{j-\frac{i_1+\cdots+i_\kappa}{\alpha_0}} + G \sum_{i_1=1}^{d_1} \cdots \sum_{i_\kappa=1}^{d_\kappa} \sum_{j=1}^{i_1+\cdots+i_\kappa} \beta_{i_1;\dots;i_\kappa;j}^{(d_1;\dots;d_\kappa)} V_1^{i_1-\frac{d_1}{\alpha_1}} \cdots V_k^{i_\kappa-\frac{d_\kappa}{\alpha_k}} \frac{\partial V^{j-\frac{i_1+\cdots+i_\kappa}{\alpha_0}}}{\partial z_{\kappa+1;1}} \right\} \\
&= G \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} V_1^{i_1-\frac{d_1}{\alpha_1}} \cdots V_k^{i_\kappa-\frac{d_\kappa}{\alpha_k}} \left\{ \sum_{l=2}^{i_1+\cdots+i_\kappa+1} \beta_{i_1;\dots;i_\kappa;j-1}^{(d_1;\dots;d_\kappa)} \right. \\
&\left. \times V_{\kappa+1}^{1-\frac{1}{\alpha_{\kappa+1}}} V^{j-\frac{i_1+\cdots+i_\kappa+1}{\alpha_0}} - \sum_{j=1}^{i_1+\cdots+i_\kappa} \beta_{i_1;\dots;i_\kappa;j}^{(d_1;\dots;d_\kappa)} \left(l - \frac{i_1+\cdots+i_\kappa}{\alpha_0} \right) V_{\kappa+1}^{1-\frac{1}{\alpha_{\kappa+1}}} V^{j-\frac{i_1+\cdots+i_\kappa+1}{\alpha_0}} \right\} \\
&= G \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} \sum_{j=1}^{i_1+\cdots+i_\kappa+1} \beta_{i_1;\dots;i_\kappa;1;j}^{(d_1;\dots;d_\kappa;1)} V_1^{i_1-\frac{d_1}{\alpha_1}} \cdots V_k^{i_\kappa-\frac{d_\kappa}{\alpha_k}} V_{\kappa+1}^{1-\frac{1}{\alpha_{\kappa+1}}} V^{j-\frac{i_1+\cdots+i_\kappa+1}{\alpha_0}},
\end{aligned}$$

where

$$\beta_{i_1;\dots;i_\kappa;1;j}^{(d_1;\dots;d_\kappa;1)} = \beta_{i_1;\dots;i_\kappa;j-1}^{(d_1;\dots;d_\kappa)} - \left(l - \frac{i_1+\cdots+i_\kappa}{\alpha_0} \right) \beta_{i_1;\dots;i_\kappa;j}^{(d_1;\dots;d_\kappa)}, 1 \le j \le \sum_{k=1}^{\kappa} i_k + 1, 1 \le i_k \le d_k,\ k = 1,\dots,\kappa, \tag{15}$$

with

$$\beta_{i_1;\dots;i_\kappa;1;j}^{(d_1;\dots;d_\kappa;1)} = 0, \quad \text{for all} \quad i_k \notin \{1,\dots,d_k\},\ k = 1,\dots,\kappa, \quad \text{or} \quad l \notin \{1,\dots,i_1+\cdots+i_\kappa\}. \tag{16}$$

Hence, (12) holds by induction for $\kappa \mapsto \kappa+1$ with $d_{\kappa+1} = 1$ and the recursive formula to compute coefficients $\beta_{i_1;\dots;i_\kappa,1,l}^{(d_1;\dots;d_\kappa,1)}$ is given by (15) and (16).

Step 4. Assuming that (12) holds for $\kappa \in \{1,\dots,K\}$ and $d_\kappa \in \{1,\dots,D_\kappa - 1\}$, one obtain

$$\frac{\partial^{1+\sum_{k=1}^{\kappa} d_k} G}{\partial \prod_{k=1}^{\kappa} \mathbf{z}_{k;1:d_k} \partial z_{\kappa;d_\kappa+1}} = \prod_{i_1=1}^{d_1} z_{1;i_1}^{-\frac{1}{\alpha_0\alpha_1}-1} \cdots \prod_{i_\kappa=1}^{d_\kappa} z_{\kappa;i_\kappa}^{-\frac{1}{\alpha_0\alpha_\kappa}-1} \left\{ \frac{\partial G}{\partial z_{\kappa;d_\kappa+1}} \sum_{i_1=1}^{d_1} \cdots \sum_{i_\kappa=1}^{d_\kappa} \sum_{j=1}^{i_1+\cdots+i_\kappa} \beta_{i_1;\dots;i_\kappa;j}^{(d_1;\dots;d_\kappa)} \right.$$

$$
\begin{aligned}
&\times V_1^{i_1-\frac{d_1}{\alpha_1}}\cdots V_\kappa^{i_\kappa-\frac{d_\kappa}{\alpha_k}}V^{j-\frac{i_1+\cdots+i_\kappa}{\alpha_0}}+G\sum_{i_1=1}^{d_1}\cdots\sum_{i_\kappa=1}^{d_\kappa}\sum_{j=1}^{i_1+\cdots+i_\kappa}\beta_{i_1;\ldots;i_\kappa;j}^{(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}}V^{j-\frac{i_1+\cdots+i_\kappa}{\alpha_0}}\\
&+G\sum_{i_1=1}^{d_1}\cdots\sum_{i_\kappa=1}^{d_\kappa}\sum_{j=1}^{i_1+\cdots+i_\kappa}\beta_{i_1;\ldots;i_\kappa;j}^{(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^{j-\frac{i_1+\cdots+i_\kappa}{\alpha_0}}}{\partial z_{\kappa;d_\kappa+1}}\Bigg\}\\
&=G\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}\Bigg\{\sum_{i_1=1}^{d_1}\cdots\sum_{i_\kappa=2}^{d_\kappa+1}\sum_{l=2}^{i_1+\cdots+i_\kappa}\beta_{i_1;\ldots;i_\kappa-1;j-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}}V^{j-\frac{i_1+\cdots+i_\kappa}{\alpha_0}}\\
&-\sum_{i_1=1}^{d_1}\cdots\sum_{i_\kappa=1}^{d_\kappa}\sum_{j=1}^{i_1+\cdots+i_\kappa}\beta_{i_1;\ldots;i_\kappa;j}^{(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^{j-\frac{i_1+\cdots+i_\kappa}{\alpha_0}}\\
&-\sum_{i_1=1}^{d_1}\cdots\sum_{i_\kappa=2}^{d_\kappa+1}\sum_{j=1}^{i_1+\cdots+i_\kappa-1}\beta_{i_1;\ldots;i_\kappa-1;j}^{(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(l-\frac{i_1+\cdots+i_\kappa-1}{\alpha_0}\right)V^{j-\frac{i_1+\cdots+i_\kappa}{\alpha_0}}\Bigg\}\\
&=G\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+1}\sum_{j=1}^{i_1+\cdots+i_\kappa}\beta_{i_1;\ldots;i_\kappa;j}^{(d_1;\ldots;d_\kappa+1)}V_1^{i_1-\frac{d_1}{\alpha_1}}\cdots V_k^{i_\kappa-\frac{d_\kappa+1}{\alpha_k}}V^{j-\frac{i_1+\cdots+i_\kappa}{\alpha_0}}
\end{aligned}
$$

where

$$
\beta_{i_1;\ldots;i_\kappa;j}^{(d_1;\ldots;d_\kappa+1)}=\beta_{i_1;\ldots;i_\kappa-1;j-1}^{(d_1;\ldots;d_\kappa)}-\frac{1}{\alpha_0}\left(i_\kappa-\frac{d_\kappa}{\alpha_\kappa}\right)\beta_{i_1;\ldots;i_\kappa;j}^{(d_1;\ldots;d_\kappa)}-\left(l-\frac{i_1+\cdots+i_\kappa-1}{\alpha_0}\right)\beta_{i_1;\ldots;i_\kappa-1;j}^{(d_1;\ldots;d_\kappa)},\quad(17)
$$

if $1\le j\le 1+\sum_{k=1}^{\kappa}i_k, i_k\le d_k,\ k=1,\ldots,\kappa$, with

$$
\beta_{i_1;\ldots;i_\kappa;j}^{(d_1;\ldots;d_\kappa+1)}=0,\quad\text{for all}\quad i_k\notin\{1,\ldots,d_k\},\ k=1,\ldots,\kappa,\quad\text{or}\quad j\notin\{1,\ldots,\sum_{k=1}^{\kappa}i_k\}.\qquad(18)
$$

Hence, (12) holds by induction for $\in\{1,\ldots,K\}$ with $d_\kappa\mapsto d_\kappa+1$, and the recursive formula to compute coefficients $\beta_{i_1;\ldots;i_\kappa;j}^{(d_1;\ldots;d_\kappa+1)}$ is given by (17) and (18).

A.4 Complexity

If $\kappa=1$, the number of coefficients $\beta_{i_1;j}^{(d_1)}$ to be computed recursively in (12) is

$$
\sum_{h=1}^{d_1}\left(\sum_{i_1=1}^{h}\sum_{j=1}^{i_1}1\right)=\sum_{h=1}^{d_1}\frac{h(h+1)}{2}=\frac{1}{2}\left(\sum_{h=1}^{d_1}h^2+\sum_{h=1}^{d_1}h\right)=\frac{d_1(d_1+1)(d_1+2)}{6},
$$

which implies that the complexity is $\mathcal{O}(d_1^3)$. For $1 \leq \kappa \leq K$ clusters of size $d_1, \ldots, d_\kappa$, the number of coefficients $\beta_{i_1;\ldots;i_\kappa;j}^{(d_1;\ldots;d_\kappa)}$ to be computed in (12) is

$$\begin{aligned}
&\sum_{h=1}^{d_\kappa} \left(\sum_{i_1=1}^{d_1} \cdots \sum_{i_{\kappa-1}=1}^{d_{\kappa-1}} \sum_{i_\kappa=1}^{h} \sum_{j=1}^{i_1+\cdots+i_\kappa} 1 \right) = \sum_{h=1}^{d_\kappa} \sum_{i_\kappa=1}^{h} \sum_{i_1=1}^{d_1} \cdots \sum_{i_{\kappa-1}=1}^{d_{\kappa-1}} (i_1 + \cdots + i_\kappa) \\
&= \sum_{h=1}^{d_\kappa} \sum_{i_\kappa=1}^{h} \left\{ \frac{d_1(d_1+1)}{2} d_1 \cdots d_{\kappa-1} + \cdots + d_1 \cdots d_{\kappa-2} \frac{d_{\kappa-1}(d_{\kappa-1}+1)}{2} + d_1 \cdots d_{\kappa-1} i_\kappa \right\} \\
&= \left\{ \frac{d_1(d_1+1)}{2} d_2 \cdots d_{\kappa-1} \frac{d_\kappa(d_\kappa+1)}{2} \right\} + \cdots + \left\{ d_1 \cdots d_{\kappa-2} \frac{d_{\kappa-1}(d_{\kappa-1}+1)}{2} \frac{d_\kappa(d_\kappa+1)}{2} \right\} \\
&\quad + \left\{ d_1 \cdots d_{\kappa-1} \frac{d_\kappa(d_\kappa+1)(d_\kappa+2)}{6} \right\},
\end{aligned}$$

which implies that the total complexity for the computation of the coefficients $\beta_{i_1;\ldots;i_\kappa;j}^{(d_1;\ldots;d_\kappa)}$ in (12) is

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

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