MODELLING MULTIVARIATE EXTREME VALUE DISTRIBUTIONS VIA MARKOV TREES

Shuang Hu, Zuoxiang Peng, Johan Segers

LIDAM Discussion Paper ISBA
2022 / 21

UCLouvain

ISBA | LIDAM
Louvain Institute of Data Analysis and Modeling in economics and statistics

ISBA

Voie du Roman Pays 20 - L1.04.01

B-1348 Louvain-la-Neuve

Email : lidam-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/isba/publication.html

Modelling multivariate extreme value distributions via Markov trees*

Shuang Hu† Zuoxiang Peng‡ Johan Segers§

August 5, 2022

Abstract

Multivariate extreme value distributions are a common choice for modelling multivariate extremes. In high dimensions, however, the construction of flexible and parsimonious models is challenging. We propose to combine bivariate extreme value distributions into a Markov random field with respect to a tree. Although in general not an extreme value distribution itself, this Markov tree is attracted by a multivariate extreme value distribution. The latter serves as a tree-based approximation to an unknown extreme value distribution with the given bivariate distributions as margins. Given data, we learn an appropriate tree structure by Prim's algorithm with estimated pairwise upper tail dependence coefficients or Kendall's tau values as edge weights. The distributions of pairs of connected variables can be fitted in various ways. The resulting tree-structured extreme value distribution allows for inference on rare event probabilities, as illustrated on river discharge data from the upper Danube basin.

Keywords. Kendall's tau; Markov tree; Multivariate extreme value distribution; Prim's algorithm; probabilistic graphical model; rare event; tail dependence

1 Introduction

The modelling of tail dependence and extreme events has become a statistical problem of great interest. Through asymptotic theory, extreme value theory postulates models for sample extremes (extreme value distributions) and excesses over high thresholds (generalized Pareto distributions), both univariate and multivariate. Fields of application include meteorology and climatology (heat waves, wind storms, floods), finance (stock market crashes), reinsurance (large claims), and machine learning (anomaly detection). Text-book treatments can be found in Coles (2001), Beirlant et al. (2004) and de Haan & Ferreira (2006), while Davison & Huser (2015) provides a more recent survey of the field.

Multivariate extreme value distributions arise as the possible limiting distributions of normalized component-wise maxima of independent copies of a random vector. Their construction, however, is not straightforward because of their complicated dependence structure. Non-parametric inference on multivariate extreme value distributions is usually performed in terms of dependence functions such as the angular or spectral measure (Einmahl et al., 2001; Einmahl & Segers, 2009), the Pickands dependence function (Deheuvels, 1991) and the stable tail dependence function (Huang, 1992; Drees & Huang, 1998). Early reviews of parametric models can be found in Coles & Tawn (1991), Kotz & Nadarajah (2000, Chapter 3), and Beirlant et al. (2004, Chapter 9). Well-known parametric models include the

*Running title: *Modelling multivariate extremes*

†Southwest University and UCLouvain. E-mail: hs1995@email.swu.edu.cn. Corresponding author

‡Southwest University. E-mail: pzx@swu.edu.cn

§UCLouvain. E-mail: johan.segers@uclouvain.be

logistic and negative logistic families (Gumbel, 1960; Tawn, 1990), the Hüsler–Reiss model (Hüsler & Reiss, 1989), mixtures of Dirichlet distributions (Boldi & Davison, 2007), the pairwise beta distribution (Cooley et al., 2010), max-linear factor models (Einmahl et al., 2012), the extremal t process (Davison et al., 2012), and models based on Bernstein polynomials (Hanson et al., 2017). Some generic construction principles yielding known and new examples are proposed in Ballani & Schlather (2011) and Kiriliouk et al. (2019). Although these methods perform quite well and have reasonable efficiency in low and moderate dimensions, their application in higher dimensions remains challenging stemming from the potentially complex dependence structure of multivariate extreme value distributions, computational challenges, and the need to balance flexibility with parsimony.

Graphical models represent relatively simple probabilistic structures and are a popular way to model multivariate distributions in a sparse way. Lately, the intersection of extreme value theory and graphical models has attracted quite some attention. Gissibl & Klüppelberg (2018) introduced max-linear models on directed acyclic graphs, whose dependence structures were further investigated in Gissibl et al. (2018). Segers (2020) studied the limit theory of extremes of regularly varying Markov trees, with applications to structured Hüsler–Reiss distributions in Asenova et al. (2021) and with extensions to Markov random fields on block graphs in Asenova & Segers (2021). In parallel, Engelke & Hitz (2020) proposed a way to factor the density of multivariate generalized Pareto distribution on graphs through a variation on the Hammersley–Clifford theorem related to a kind of redefined conditional independence, followed by Engelke & Volgushev (2020), who studied a data-driven methodology of identifying the underlying graphical structures for such a factorization. An illuminating review of sparse models for multivariate extremes is given in Engelke & Ivanovs (2021).

The goal of this paper is to approximate a potentially unknown multivariate extreme value distribution by another multivariate extreme value distribution constructed from a Markov random field with respect to a tree, or Markov tree in short. As special cases of graphical models, Markov trees have simple and tractable structures. The conditional independence relations they satisfy are equivalent to a certain factorization of their joint distribution and eventually lead to an effective dimension reduction. The application of Markov trees in the modelling of multivariate probability distribution goes back to the seminal paper by Chow & Liu (1968), who approximate a d-dimensional discrete distribution by a product of pairwise distributions linked up by a tree. It was demonstrated that in many cases, this method provides a reasonable approximation to the full distribution, capturing at least the marginal distributions and some important features of the joint dependence structure. The idea remains attractive in modern practice. In Horváth et al. (2020), for instance, such Markov tree approximations are used in the context of anomaly detection in machine learning.

A Markov tree involves a set of conditional independence relations between random variables. However, Papastathopoulos & Strokorb (2016) have shown that for a max-stable random vector with positive continuous density, conditional independence of two random variables given the other ones implies unconditional independence of the two variables. Except for trivial cases, a random vector with a multivariate extreme value distribution and a continuous density can therefore not satisfy any conditional independence relations. Hence, the method of Chow & Liu (1968) cannot be applied directly to multivariate extreme value distributions. Still, Engelke & Volgushev (2020) and Segers (2020) demonstrated that conditional independence relations can be incorporated into distributions of multivariate extremes through the lens of excesses over high thresholds. Based on these findings, we propose a method based on graphical models that consists of merging bivariate extreme value distributions along a tree.

Given a multivariate extreme value distribution and a tree on the set of variables, we

construct a Markov tree by letting each connected pair of variables have the same distribution as the corresponding bivariate margin of the extreme value distribution. The Markov tree constructed in this way is shown to belong to the domain of attraction of another extreme value distribution. This second multivariate extreme value distribution serves as a tree-based approximation to the first one, fully determined by the tree structure and certain bivariate margins. Explicit expressions of such a *tree-structured multivariate extreme value distribution* (see Definition 2.2 below) are derived from those of joint tails of Markov trees.

In practice, the true extreme value distribution is unknown and needs to be estimated from the data. To do so, we implement a three-step procedure: (1) Select the maximum dependence tree which has the maximal sum of edge weights, taken as the pairwise estimates of the upper tail dependence coefficient λ or Kendall's τ; (2) Estimate the dependence structures of all the pairs of adjacent variables on the maximum dependence tree; (3) Measure the goodness-of-fit of the constructed model by comparing the distributions of nonadjacent pairs of variables in the original extreme value law on the one hand and in its tree-based approximation on the other hand. It is worth noting that, in the second step, different extreme value parametric families can be assumed for different adjacent pairs and different estimation methods can be applied for the parameters. All in all, the three steps combined provide a flexible and computationally feasible way to fit a flexible class of tree-based multivariate extreme value distributions.

Some theoretical results corresponding to the modelling procedure are also obtained. If the original multivariate extreme value distribution has a tree-based dependence structure, then with probability tending to one, the set of maximum dependence trees weighted by the upper tail dependence coefficients contains the true tree structure. Moreover, the joint estimator of the parameters of all the edge-wise bivariate dependence structures is still asymptotically normal, provided all the edge-wise estimators have a certain asymptotic expansion.

We illustrate the procedure to the average daily discharges of rivers in the upper Danube basin, estimating the probability of extreme flooding in at least one of several gauging stations. Specifically, for $j = 1, \ldots, d$, let ξ_j be the discharge at location j and let u_j denote a given water level at station j. The probability of interest is

$$\mathbb{P}\left(\xi_1 > u_1 \text{ or } \ldots \text{ or } \xi_d > u_d\right). \tag{1.1}$$

Under the assumption that $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_d)$ belongs to the domain of attraction of a multivariate extreme value distribution, the above probability is estimated based on a tree-structured approximation of the attractor.

The paper is organized as follows. In Section 2, we give some preliminaries related to multivariate extreme value distributions and Markov trees. The model construction and its properties are discussed in Section 3. The statistical procedure is detailed in Section 4. A simulation study is conducted in Section 5 and a data analysis of the discharges of rivers in the upper Danube basin is given in Section 6. All proofs as well as some additional illustrations related to the simulations and the case study are postponed to the Appendix in Section A.

2 Preliminaries

2.1 Multivariate extreme value theory

Put $V = \{1, \ldots, d\}$ for some integer $d \geq 2$. Let $\boldsymbol{\xi}^i = (\xi_{i,v}, v \in V)$, $i = 1, \ldots, n$, be an independent random sample from $\boldsymbol{\xi} = (\xi_v, v \in V)$ with univariate marginal distribution functions F_v for $v \in V$. We say $\boldsymbol{\xi}$ is in the *(maximum) domain of attraction* of H, notation

$\boldsymbol{\xi} \in D(H)$, if there exist sequences $\boldsymbol{a}_n = (a_{n,v}, v \in V) \in (0,\infty)^d$ and $\boldsymbol{b}_n = (b_{n,v}, v \in V) \in \mathbb{R}^d$ such that

$$\lim_{n\to\infty} \mathbb{P}\left(\max_{i=1,\dots,n} \xi_{i,v} \le a_{n,v} x_v + b_{n,v},\, v \in V \right) = H(\boldsymbol{x}), \qquad \boldsymbol{x} = (x_v, v \in V) \in \mathbb{R}^d, \tag{2.1}$$

for a limiting distribution H with non-degenerate margins. The distribution H is called a multivariate *extreme value* or *max-stable* distribution. An extreme value distribution is called *simple* if its margins are unit-Fréchet, that is, $H_v(x) = \exp(-1/x)$ for $x \in (0,\infty)$ and $v \in V$. By monotone marginal transformations, we can convert any extreme value distribution into a simple one without changing its copula. This allows us to deal with the marginal distributions and the dependence structure separately. To focus on the latter, we often consider simple multivariate extreme value distributions. For more background on multivariate extreme value distributions, see for instance Resnick (1987, Chapter 5) and Beirlant et al. (2004, Chapter 8).

A multivariate extreme value distribution function H with general margins can be represented by

$$H(\boldsymbol{x}) = \exp\left\{ -\ell\left(-\log H_v(x_v),\, v \in V \right) \right\}, \tag{2.2}$$

for $\boldsymbol{x} \in \mathbb{R}^d$ such that $H_v(x_v) > 0$ for all $v \in V$, and where $\ell : [0,\infty)^d \to [0,\infty)$ is the so-called *stable tail dependence function* of H and $\boldsymbol{\xi}$, from whose distribution it can be recovered via

$$\ell(\boldsymbol{y}) = \lim_{t\to\infty} t\left\{ 1 - \mathbb{P}\left(F_v(\xi_v) \le 1 - t^{-1} y_v,\, v \in V \right) \right\}, \qquad \boldsymbol{y} \in [0,\infty)^d, \tag{2.3}$$

see Huang (1992) or Drees & Huang (1998). The dependence structure of an extreme value distribution is completely described by the stable tail dependence function. In the bivariate case, $\boldsymbol{\xi} = (\xi_1, \xi_2)$, this dependence function is closely related to the *upper tail dependence coefficient* defined by

$$\lambda = \lim_{t\to\infty} t\, \mathbb{P}\left(1 - F_1(\xi_1) \le t^{-1}, 1 - F_2(\xi_2) \le t^{-1} \right), \tag{2.4}$$

which captures the extremal dependence along the main diagonal of the positive plane. The identity $\lambda = 2 - \ell(1,1)$ is clear by the definitions.

2.2 Regularly varying Markov trees

A *graph* is a pair (V, E), where V is a finite set of *nodes* and $E \subseteq \{(a,b) \in V \times V : a \neq b\}$ is a set of paired nodes called *edges*. A graph is called *undirected* if $(a,b) \in E$ implies $(b,a) \in E$. A *path* from a to b, with notation $[a \rightsquigarrow b]$, is a collection $\{(u_0,u_1),(u_1,u_2),\dots,(u_{n-1},u_n)\}$ of distinct and directed edges such that $u_0 = a$ and $u_n = b$. An undirected graph is *connected* if for any pair of distinct nodes a and b, there is a path from a to b. A *cycle* is a path whose start and end nodes are the same. A *tree* is an undirected, connected graph without cycles. For any distinct nodes a and b on a tree, there is a unique path from a to b.

Let $\mathcal{T} = (V, E)$ be a tree and let $\boldsymbol{Y} = (Y_v, v \in V)$ be a d-variate random vector indexed by the nodes of $\mathcal{T}$. For disjoint and nonempty subsets A, B and C of V, we say that C *separates A from B* in $\mathcal{T}$ if any path from a node in A to a node in B passes through at least one node in C. We call $\boldsymbol{Y}$ a *Markov tree* on $\mathcal{T}$ if it satisfies the *global Markov property*, i.e., whenever C separates A from B, the conditional independence relation

$$\boldsymbol{Y}_A \perp \boldsymbol{Y}_B \mid \boldsymbol{Y}_C \tag{2.5}$$

holds, where $\boldsymbol{Y}_A = (Y_a, a \in A)$ for $A \subseteq V$ and we write $\boldsymbol{Y}_V$ as $\boldsymbol{Y}$ when $A = V$.

We first recall some results corresponding to Markov trees and multivariate extreme value distributions. The following condition is adapted from Segers (2020) and ensures the weak

convergence of the conditional distribution of the tree given that it exceeds a high threshold at a fixed variable. Condition 2.1(i) is an assumption about the high level transformation. Condition 2.1(ii) controls the probability of extremes caused by non-extremes.

Condition 2.1. For a non-negative Markov tree $\boldsymbol{Y} = (Y_v, v \in V)$ on the tree $\mathcal{T} = (V, E)$:

(i) for every $e = (a, b) \in E$, there exists a version of the conditional distribution of Y_b given Y_a and a probability measure μ_e on $[0, \infty)$ such that for every $y \in [0, \infty)$ in which the distribution function $y \mapsto \mu_e([0, y])$ is continuous, we have

$$\mathbb{P}(y_a^{-1} Y_b \leq x \mid Y_a = y_a) \to \mu_e([0, x]), \qquad y_a \to \infty; \tag{2.6}$$

(ii) for every edge $e = (a, b) \in E$ such that $\mu_e(\{0\}) > 0$, we have

$$\forall \eta > 0, \quad \lim_{\delta \downarrow 0} \limsup_{y_a \to \infty} \sup_{\varepsilon \in [0, \delta]} \mathbb{P}(y_a^{-1} Y_b > \eta \mid Y_a = \varepsilon y_a) = 0.$$

Assume $\boldsymbol{Y} = (Y_v, v \in V)$ is a non-negative Markov tree on $\mathcal{T} = (V, E)$ satisfying Condition 2.1. If for every $u \in V$ we have

$$t\, \mathbb{P}(Y_u > t) \to 1, \qquad t \to \infty, \tag{2.7}$$

then by Theorem 2 and Corollary 5 in Segers (2020), we have

$$\mathbb{P}(t^{-1} \boldsymbol{Y} \leq \boldsymbol{x} \mid Y_u > t) \to \mathbb{P}(\boldsymbol{W}_u \leq \boldsymbol{x}), \qquad t \to \infty, \tag{2.8}$$

in all continuity points $\boldsymbol{x} \in [0, \infty)^d$ of the distribution function of $\boldsymbol{W}_u$. The weak limit $\boldsymbol{W}_u$ is equal in distribution to $W \boldsymbol{\Theta}_u$, where W is a unit-Pareto random variable that is independent of the random vector $\boldsymbol{\Theta}_u = (\Theta_{u,v}, v \in V)$ called *tail tree* and given by $\Theta_{u,u} = 1$ together with

$$\Theta_{u,v} = \prod_{e \in [u \rightsquigarrow v]} M_e, \qquad v \in V \setminus \{u\},$$

in terms of independent random variables M_e, for $e \in E$, with marginal distributions μ_e in (2.6). In fact, the random vector $\boldsymbol{\Theta}_u$ is the weak limit of the conditional distribution of $t^{-1} \boldsymbol{Y}$ given that $Y_u = t$ as $t \to \infty$. The random vector $\boldsymbol{Y}$ is called a *regularly varying Markov tree*. The distributions μ_e of the random variables M_e can be arbitrary except for the fact that always $\mathbb{E}[M_e] \leq 1$, see the proof of Proposition 2.5.

By Theorem 2 in Segers (2020), the convergence in (2.8) implies that $\boldsymbol{Y}$ is in the domain of attraction of a simple multivariate extreme value distribution, i.e., for independent random samples $\boldsymbol{Y}^i = (Y_{i,v}, v \in V)$, $i = 1, \ldots, n$, generated from $\boldsymbol{Y}$, the convergence (2.1) holds with $a_{n,v} = n$ and $b_{n,v} = 0$ for $v \in V$. To stress the importance of such extreme value distributions, which emerge as the weak limits of normalized component-wise maxima of independent copies of Markov trees, we state the following definition.

Definition 2.2. A d-variate extreme value distribution H is said to *have a dependence structure linked to the tree* $\mathcal{T} = (V, E)$ if, up to increasing marginal transformations, it contains in its domain of attraction a non-negative regularly varying Markov tree $\boldsymbol{Y} = (Y_v, v \in V)$ on $\mathcal{T}$ satisfying Condition 2.1 and Eq. (2.7). We call H a *tree-structured extreme value distribution*.

The stable tail dependence function and upper tail dependence coefficients of a tree-structured extreme value distribution H are given by the coming propositions.

Proposition 2.3. *Let H be a tree-structured extreme value distribution as in Definition 2.2 with respect to $\mathcal{T} = (V, E)$. Let $(M_e, e \in E)$ be independent random variables with marginal distributions μ_e in (2.6). The stable tail dependence function of H is*

$$\ell(\boldsymbol{y}) = \sum_{i=1}^{d} \mathbb{E}\left\{ \max_{j=i,\ldots,d} \left(y_{v_j} \prod_{e \in [v_i \rightsquigarrow v_j]} M_e \right) - \max_{j=i+1,\ldots,d} \left(y_{v_j} \prod_{e \in [v_i \rightsquigarrow v_j]} M_e \right) \right\}, \tag{2.9}$$

where $v_1, \ldots, v_d$ is an arbitrary permutation of $1, \ldots, d$. The maximum over the empty set is defined as zero and M_{uv} is interpreted as one if $u = v$. If, in addition, we have $\mathbb{E}(M_{ab}) = 1$ for all $(a, b) \in E$, then

$$\ell(\boldsymbol{y}) = \mathbb{E}\left\{ \max_{j=1,\ldots,d} \left(y_{v_j} \prod_{e \in [v_1 \rightsquigarrow v_j]} M_e \right) \right\} = \mathbb{E}\left\{ \max_{v=1,\ldots,d} \left(y_v \prod_{e \in [1 \rightsquigarrow v]} M_e \right) \right\}.$$

Proposition 2.4. *Let the random vector $\boldsymbol{X} = (X_v, v \in V)$ have distribution H as in Proposition 2.3. For distinct $a, b \in V$, the stable tail dependence function of (X_a, X_b) is*

$$\ell_{ab}(x, y) = y_a + y_b - \mathbb{E}\left\{ \min\left(y_a, y_b \prod_{e \in [a \rightsquigarrow b]} M_e \right) \right\}, \qquad x, y \geq 0,$$

while its upper tail dependence coefficient is

$$\lambda_{ab} = \mathbb{E}\left\{ \min\left(1, \prod_{e \in [a \rightsquigarrow b]} M_e \right) \right\}. \tag{2.10}$$

Proposition 2.5. *Let $\boldsymbol{X}$ be the random vector in Proposition 2.4 and let λ_{uv} be the upper tail dependence coefficient of (X_u, X_v). For any distinct nodes $a, b, u \in V$ such that u is on the path from a to b, we have*

$$\lambda_{au}\lambda_{ub} \leq \lambda_{ab} \leq \min(\lambda_{au}, \lambda_{ub}). \tag{2.11}$$

A sufficient condition for the inequality on the right-hand side to be strict is that, for any $e \in E$ and any $\varepsilon > 0$, we have $\mathbb{P}(1 - \varepsilon < M_e < 1) > 0$ and $\mathbb{P}(1 < M_e < 1 + \varepsilon) > 0$.

The inequality on the right-hand side of (2.11) can be an equality, even in non-trivial situations. Let for instance $\varepsilon_1, Y_2, \varepsilon_3$ be independent unit-Fréchet random variables and define $Y_j = \max(Y_2, \varepsilon_j)/2$ for $j \in \{1, 3\}$. Then Y_1 and Y_3 are conditionally independent given Y_2 and (Y_1, Y_2, Y_3) is a regularly varying Markov tree with respect to the tree 1–2–3. Still, the tail dependence coefficients are $\lambda_{12} = \lambda_{23} = \lambda_{13} = 1/2$. Note that $\mathbb{P}(M_{12} = 0) = \mathbb{P}(M_{12} = 2) = 1/2$ while $\mathbb{P}(M_{23} = 1/2) = 1$, so that indeed $\mathbb{E}\{\min(1, M_{12}M_{23})\} = \mathbb{E}(M_{12}M_{23}) = 1/2$ and $\mathbb{E}\{\min(1, M_{12})\} = 1/2$.

The left-hand side inequality of (2.11) can be an equality too. For instance, if (a, u) and (u, b) are edges and if M_{au} and M_{ub} are bounded by 1 almost surely, then $\lambda_{ab} = \mathbb{E}\{\min(1, M_{au}M_{ub})\} = \mathbb{E}(M_{au}M_{ub}) = \mathbb{E}(M_{au})\,\mathbb{E}(M_{ub}) = \mathbb{E}\{\min(1, M_{au})\}\,\mathbb{E}\{\min(1, M_{ub})\} = \lambda_{au}\lambda_{ub}$.

Remark 2.6. For any pair $a, b \in V$ of distinct nodes, Proposition 2.5 and a recursive argument imply

$$\prod_{e \in [a \rightsquigarrow b]} \lambda_e \leq \lambda_{ab} \leq \min_{e \in [a \rightsquigarrow b]} \lambda_e.$$

3 Model definition and properties

3.1 Model construction

Let G be a simple d-variate extreme value distribution to be approximated by a tree-structured extreme value distribution on a tree $\mathcal{T} = (V, E)$. As before, write $V = \{1, \ldots, d\}$. Assume $\boldsymbol{Z} = (Z_v, v \in V)$ is a random vector with distribution G having stable tail dependence function ℓ. We construct a Markov tree, denoted by $\boldsymbol{Y} = (Y_v, v \in V)$, on $\mathcal{T}$ by the following definition.

Definition 3.1. Given are a d-variate simple extreme value distribution G and a tree $\mathcal{T} = (V, E)$. We define the Markov tree $\boldsymbol{Y} = (Y_v, v \in V)$ by the following two properties:

(1) the random vector $\boldsymbol{Y}$ satisfies the global Markov property (2.5) with respect to $\mathcal{T}$;

(2) for each pair $(a, b) \in E$, the distribution of (Y_a, Y_b) is the same as the one of (Z_a, Z_b).

By the construction, the distribution of (Y_a, Y_b) for $(a, b) \in E$ is bivariate extreme value with unit-Fréchet margins. Thus Condition 2.1 is satisfied and $\boldsymbol{Y}$ is a regularly varying Markov tree (Segers, 2020, Example 3). Hence the conclusions in Propositions 2.3–2.5 hold for $\boldsymbol{Y}$. Let G_M denote the limiting extreme value distribution of scaled component-wise maxima of independent random samples from $\boldsymbol{Y}$. The representation of G_M follows directly from Eq. (2.2) and Proposition 2.3.

Corollary 3.2. *The Markov tree $\boldsymbol{Y}$ in Definition 3.1 is in the domain of attraction of a simple d-variate extreme value distribution G_M with dependence structure linked to the tree $\mathcal{T}$. More precisely, for $\boldsymbol{x} = (x_v, v \in V) \in (0, \infty]^d$, we have*

$$G_M(\boldsymbol{x}) = \exp\left\{-\ell^M(1/x_v, v \in V)\right\}$$

with stable tail dependence function ℓ^M given by the right-hand side in (2.9).

For a given extreme value distribution G, every different tree $\mathcal{T}$ on V leads to a different tree-structured approximation G_M. We will come back to the choice of $\mathcal{T}$ in Subsection 4.1.

The distribution G_M is determined by the increments M_e, see Corollary 3.2, whose laws μ_e can be calculated through the limit in Condition 2.1(i) if G is known. Alternatively, the distribution of M_e for $e = (a, b)$ can also be obtained from the stable tail dependence function of (Z_a, Z_b) analogously as in Proposition 3.9 in Asenova et al. (2021): we have

$$\mathbb{P}(M_{ab} \le x) = \frac{\mathrm{d}\,\ell_{ab}(x, 1)}{\mathrm{d}\,x}, \qquad 0 \le x < \infty, \tag{3.1}$$

where ℓ_{ab} is the stable tail dependence function of (Z_a, Z_b) and where the derivative is taken from the right (which always exists by convexity of ℓ_{ab}).

The stable tail dependence functions ℓ and ℓ^M of G and its tree-based approximation G_M, respectively, are different in general. Still, they share the same margins for pairs of adjacent nodes in $\mathcal{T}$, i.e., $\ell_{ab} = \ell^M_{ab}$ for $(a, b) \in E$. The difference between ℓ and ℓ^M becomes visible when studying pairs (a, b) not in E or when considering higher-order margins.

3.2 Measuring the approximation error

Given the extreme value distribution G, it is theoretically possible to get an explicit expression of G_M in Corollary 3.2. But how well can G be approximated by G_M? This question is important because, like in Chow & Liu (1968), the Markov assumption is not believed

to be satisfied exactly, especially not in high dimensions. Still, it may provide a reasonable approximation, and the question is how large or small the approximation error really is.

By construction, the adjacent pairs on the tree have precisely the same dependence structure as the corresponding pairs of the original max-table distribution. But for pairs of random variables which are nonadjacent on the tree, the tail dependence is in general distinct since the Markov tree has enclosed conditional independence in the model and this may influence the dependence structure of the limiting tree-based extreme value distribution. Therefore, differences can be expected to arise in the nonadjacent pairs on the tree. For $a, b \in V$, let λ_{ab} and λ_{ab}^M denote the upper tail dependence coefficients of the variables with indices a and b in G and in G_M, respectively. For the reason just explained, we focus on $a, b \in V$ with $a \neq b$ such that $(a, b) \notin E$ and quantify the approximation error by

$$D_{\mathcal{T}} = \sum_{a,b \in V, a \neq b, (a,b) \notin E} \left| \lambda_{ab}^M - \lambda_{ab} \right|. \tag{3.2}$$

Note that $D_{\mathcal{T}}$ also equals $\sum_{a,b \in V} |\lambda_{ab}^M - \lambda_{ab}|$ since $\lambda_{ab}^M = \lambda_{ab}$ for $(a, b) \in E$ or when $a = b$. The quantity $D_{\mathcal{T}}$ can be interpreted as the cumulative difference in bivariate tail dependence between the two extreme value distributions.

3.3 Examples

We explain the construction through two examples. Throughout the rest of this paper, the symbols G, G_M, $\boldsymbol{Y}$ and $\boldsymbol{Z}$ are defined in the same way as in Subsection 3.1. Further, $\boldsymbol{Z}_M = (Z_v^M, v \in V)$ denotes a random vector with distribution function G_M. For nonempty $U \subseteq V$, let the stable tail dependence functions of $\boldsymbol{Z}_U = (Z_v, v \in U)$ and $\boldsymbol{Z}_U^M = (Z_v^M, v \in U)$ be denoted by ℓ_U and ℓ_U^M, respectively, where we write ℓ_U as ℓ_{ab} when $U = \{a, b\}$ and where the subscript U is omitted when $U = V$.

Example 3.3 (Hüsler–Reiss model)**.** Assume $\boldsymbol{\Gamma} = (\gamma_{uv})_{u,v \in V}$ is a $d \times d$ dimensional symmetric, conditionally negative definite matrix with elements $\gamma_{uv} \in [0, \infty)$ satisfying $\gamma_{uu} = 0$ for $u \in V$, i.e., $\boldsymbol{a}^\top \boldsymbol{\Gamma} \boldsymbol{a} < 0$ for any nonzero d-dimensional column vector $\boldsymbol{a} = (a_v, v \in V)$ such that $\sum_{v \in V} a_v = 0$. For $\boldsymbol{x} \in (0, \infty)^d$, let $G(\boldsymbol{x})$ be a Hüsler–Reiss distribution (Hüsler & Reiss, 1989) defined by

$$G(\boldsymbol{x}) = \exp\left\{ - \sum_{u \in V} x_u^{-1} \Phi_{d-1}\left(\log \frac{x_v}{x_u} + \frac{\gamma_{uv}}{2}, v \in V \setminus \{u\}; \boldsymbol{\Sigma}_{V,u} \right) \right\},$$

where $\boldsymbol{\Sigma}_{V,u} = (\sigma_{ab}^2)_{a,b \in V \setminus \{u\}}$ is the $(d-1) \times (d-1)$ dimensional square matrix with elements $\sigma_{ab}^2 = (\gamma_{au} + \gamma_{bu} - \gamma_{ab})/2$. The function $\Phi_p(\boldsymbol{x}; \boldsymbol{\Sigma})$ is the p-dimensional normal cumulative distribution function with zero mean and covariance matrix $\boldsymbol{\Sigma}$, with subscript omitted when $p = 1$. For nonempty $U \subseteq V$, the stable tail dependence function of $(Z_v, v \in U)$ is

$$\ell_U(\boldsymbol{x}_U) = \sum_{u \in U} x_u \Phi_{|U \setminus \{u\}|}\left(\log \frac{x_u}{x_v} + \frac{\gamma_{uv}}{2}, v \in V \setminus \{u\}; \boldsymbol{\Sigma}_{U,u} \right), \qquad \boldsymbol{x}_U = (x_u, u \in U) \in [0, \infty)^{|U|},$$

where $|U|$ denotes the number of elements in U. Consequently, we have

$$\lambda_{ab} = 2\left\{1 - \Phi\left(\sqrt{\gamma_{ab}}/2\right)\right\}, \qquad a, b \in V.$$

From Proposition 2.1 of Asenova et al. (2021), G_M is also a Hüsler–Reiss extreme value distribution parameterized by the variogram matrix $\boldsymbol{\Gamma}_M = (\gamma_{uv}^M)_{u,v \in V}$ with elements

$$\gamma_{uv}^M = \sum_{e \in [u \rightsquigarrow v]} \gamma_e. \tag{3.3}$$

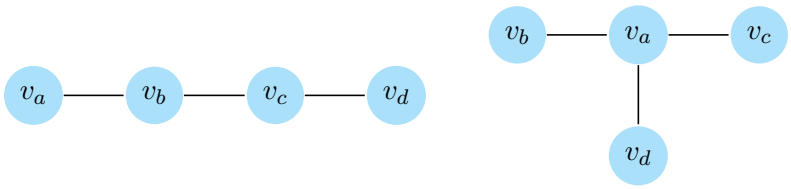

Fig. 1 Structures $\mathcal{T}^{(a)}$ (left) and $\mathcal{T}^{(b)}$ (right) for a tree with 4 nodes.

Thus $\lambda_{ab}^M = \lambda_{ab}$ for $(a,b) \in E$ and

$$\lambda_{ab}^M = 2\left\{1 - \Phi\left(2^{-1}\sqrt{\gamma_{ab}^M}\right)\right\}, \qquad (a,b) \notin E.$$

Consider the 4-dimensional Hüsler–Reiss extreme value distributions with variogram matrices

$$\boldsymbol{\Gamma}_1 = \begin{pmatrix} 0 & 4 & 4 & 4 \\ 4 & 0 & 8 & 8 \\ 4 & 8 & 0 & 8 \\ 4 & 8 & 8 & 0 \end{pmatrix}, \qquad \boldsymbol{\Gamma}_2 = \begin{pmatrix} 0 & 4 & 8 & 16 \\ 4 & 0 & 4 & 8 \\ 8 & 4 & 0 & 4 \\ 16 & 8 & 4 & 0 \end{pmatrix}. \tag{3.4}$$

For a tree having 4 nodes, two types of tree structures, $\mathcal{T}^{(a)}$ and $\mathcal{T}^{(b)}$ in Fig. 1, can be used to create Markov trees.

Given a tree, we can construct Markov trees and obtain the variogram matrix of G_M through (3.3). Table 1 presents the value of $D_{\mathcal{T}}$ defined in (3.2) of the constructed models corresponding to different tree structures. As was expected, the value of $D_{\mathcal{T}}$ is zero if G is a tree-structured extreme value distribution itself and the model G_M is constructed based on the real tree structure of G: this is the case for the Hüsler–Reiss distribution with variogram matrix $\boldsymbol{\Gamma}_1$.

Example 3.4 (Asymmetric logistic model)**.** We also consider a special case of the asymmetric logistic model of Tawn (1990). For a parameter vector $\boldsymbol{\Psi} = (\psi_v, v \in V) \in [0,1]^d$, let

$$G(\boldsymbol{x}) = \exp\left\{-\sum_{u\in V}\frac{1-\psi_u}{x_u} - \max_{u\in V}\left(\frac{\psi_u}{x_u}\right)\right\}, \qquad \boldsymbol{x} \in (0,\infty]^d.$$

It appears as the distribution function of the max-linear model $\boldsymbol{Z} = (Z_v, v \in V)$ with

$$Z_v = \max\left\{\psi_v\varepsilon, (1-\psi_v)\varepsilon_v\right\}, \tag{3.5}$$

where ε and ε_v, for $v \in V$, are independent unit-Fréchet random variables. For $U \subseteq V$, the stable tail dependence function of $(Z_v, v \in U)$ is

$$\ell_U(\boldsymbol{x}_U) = \sum_{u\in U}(1-\psi_u)x_u + \max_{u\in U}(\psi_u x_u),$$

which implies that

$$\lambda_{ab} = \min(\psi_a, \psi_b), \qquad a,b \in V, \quad a \neq b.$$

By (3.1), the distribution of M_{ab} for $(a,b) \in E$ is discrete with at most two atoms:

$$\mathbb{P}(M_{ab} = 0) = 1 - \psi_a, \qquad \mathbb{P}\left(M_{ab} = \frac{\psi_b}{\psi_a}\right) = \psi_a. \tag{3.6}$$

Write $V_i = \{i, \ldots, d\}$ and $e = (e_p, e_s)$ for $e \in E$. By Corollary 3.2 and the maximum–minimums identity, the stable tail dependence function of G_M is

$$\ell^M(\boldsymbol{y}) = \sum_{i=1}^{d} \psi_i^{-1} \sum_{\varnothing \neq U \subseteq V_i} (-1)^{(|U|+1)} \left(\prod_{e \in \bigcup_{j \in U} [i \rightsquigarrow j]} \psi_{e_p} \right) \left\{ \min_{j \in U} (y_j \psi_j) \right\}$$
$$- \sum_{i=2}^{d} \psi_{i-1}^{-1} \sum_{\varnothing \neq U \subseteq V_i} (-1)^{(|U|+1)} \left(\prod_{e \in \bigcup_{j \in U} [(i-1) \rightsquigarrow j]} \psi_{e_p} \right) \left\{ \min_{j \in U} (y_j \psi_j) \right\}; \tag{3.7}$$

see Appendix A.1 for details. Moreover, for $a, b \in V$, $a \neq b$ and $(a, b) \notin E$, it follows from Proposition 2.4 and the distribution of M_e for $e \in E$ in (3.6) that

$$\lambda_{ab}^M = \mathbb{E} \left\{ \min \left(1, \prod_{e \in [a \rightsquigarrow b]} M_e \right) \right\} = \min \left(\prod_{e \in [a \rightsquigarrow b]} \psi_{e_p}, \prod_{e \in [a \rightsquigarrow b]} \psi_{e_s} \right). \tag{3.8}$$

For the 4-variate asymmetric logistic models with parameter vectors $\boldsymbol{\Psi}_1 = (0.8, 0.7, 0.4, 0.2)$ and $\boldsymbol{\Psi}_2 = (0.5, 0.4, 0.3, 0.2)$, the numeric results of $D_{\mathcal{T}}$ in (3.2) for different trees are given in Table 1.

Table 1: The values of $D_{\mathcal{T}}$ in (3.2) and $S_{\mathcal{T}_\lambda}$ in (4.1) for the 4-variate Hüsler–Reiss model in Example 3.3 and the 4-variate asymmetric logistic model in Example 3.4 based on trees with structures $\mathcal{T}^{(a)}$ and $\mathcal{T}^{(b)}$ in Fig. 1, where (i, j, k, l) in the second column denotes the tree $\mathcal{T}$ with $v_a = i$, $v_b = j$, $v_c = k$ and $v_d = l$ in tree structure $\mathcal{T}^{(a)}$ or $\mathcal{T}^{(b)}$.

Structure	Trees	Hüsler–Reiss				Asymmetric logistic			
		$\boldsymbol{\Gamma}_1$		$\boldsymbol{\Gamma}_2$		$\boldsymbol{\Psi}_1$		$\boldsymbol{\Psi}_2$	
		$S_{\mathcal{T}_\lambda}$	$D_{\mathcal{T}}$	$S_{\mathcal{T}_\lambda}$	$D_{\mathcal{T}}$	$S_{\mathcal{T}_\lambda}$	$D_{\mathcal{T}}$	$S_{\mathcal{T}_\lambda}$	$D_{\mathcal{T}}$
$\mathcal{T}^{(a)}$	$(1,2,3,4)$	0.632	0.638	0.952	0.038	1.3	0.376	0.9	0.490
	$(1,2,4,3)$	0.632	0.638	0.792	0.384	1.1	0.716	0.8	0.630
	$(1,3,2,4)$	0.632	0.638	0.632	0.488	1.0	0.568	0.8	0.560
	$(1,3,4,2)$	0.632	0.638	0.632	0.564	0.8	1.076	0.7	0.750
	$(1,4,2,3)$	0.632	0.638	0.520	0.686	0.8	0.908	0.7	0.700
	$(1,4,3,2)$	0.632	0.638	0.680	0.435	0.8	1.076	0.7	0.750
	$(2,1,3,4)$	0.792	0.346	0.792	0.384	1.3	0.344	0.9	0.466
	$(2,1,4,3)$	0.792	0.346	0.680	0.567	1.1	0.704	0.8	0.616
	$(2,3,1,4)$	0.792	0.346	0.520	0.686	1.0	0.548	0.8	0.540
	$(2,4,1,3)$	0.792	0.346	0.360	0.919	0.8	0.888	0.7	0.680
	$(3,2,1,4)$	0.792	0.346	0.680	0.435	1.3	0.296	0.9	0.450
	$(3,1,2,4)$	0.792	0.346	0.632	0.564	1.3	0.284	0.9	0.446
$\mathcal{T}^{(b)}$	$(1,2,3,4)$	0.952	0	0.520	0.669	1.3	0.160	0.9	0.350
	$(2,3,4,1)$	0.632	0.580	0.792	0.272	1.3	0.240	0.9	0.420
	$(3,4,1,2)$	0.632	0.580	0.792	0.272	1.0	0.660	0.8	0.560
	$(4,1,2,3)$	0.632	0.580	0.520	0.669	0.6	1.200	0.6	0.800

4 Estimation and modelling

In this section, we assume the data are drawn from a distribution that, up to increasing marginal transformations, is attracted by a simple extreme value distribution G. Given such

a set of data, we propose a three-step procedure to construct a tree-structured dependence model for extremes. The three steps were given in the introduction and are detailed below.

4.1 Structure learning for the Markov tree

To construct the model, a tree structure is required. So the first crucial step is to choose an appropriate tree structure from all possible candidates. In the framework of structure learning for graphical models, a commonly used way is to give each edge a weight in the complete graph and then choose the tree with maximum or minimum sum of weights along its edges; see for instance Chow & Liu (1968), Horváth et al. (2020) and Engelke & Volgushev (2020). We follow this method and try to find the tree structure which can retain the dependence of G as much as possible.

An intuitive way to define the edge weight of a pair of variables is by the value of a bivariate dependence measure. Here we consider the upper tail dependence coefficient λ defined in (2.4) as well as Kendall's τ. Another possible edge weight, advocated in Engelke & Volgushev (2020), is the extremal variogram. In that paper, the upper tail dependence coefficient is used as edge weight as well but under the name 'extremal correlation'. Note that, in contrast to Engelke & Volgushev (2020), we do not assume that G is tree-structured.

The upper tail dependence coefficient of a random vector is the same as the one of the extreme value distribution to which it is attracted, and this provides a motivation for using this coefficient when given data sampled from a distribution in the domain of attraction of an extreme value distribution. For Kendall's tau, however, this argument does not hold: its value can differ considerably between the data-generating distribution and the extreme value distribution to which it is attracted. Nevertheless, the method based on Kendall's tau performs well in the simulations and in the case study in Sections 5 and 6, and this is why we include it here too.

Let ω_{ab} denote the weight associated to $(a,b) \in V \times V$, with $\omega \in \{\lambda, \tau\}$. We define the set of *maximum (tail) dependence trees* by

$$\boldsymbol{T}_\omega^\star = \arg\max_{\mathcal{T}} S_{\mathcal{T}_\omega} \quad \text{where } S_{\mathcal{T}_\omega} = \sum_{(a,b)\in E} \omega_{ab} \text{ for } \mathcal{T} = (V,E), \tag{4.1}$$

the maximum being taken over all possible trees over the set V, and with λ_{ab} and τ_{ab} the tail dependence coefficient and Kendall's tau, respectively, of the data-generating distribution for the pair of variables indexed by $(a,b) \in V^2$. The notation $\mathcal{T}_\omega^\star$ will be used for a generic member of the set $\boldsymbol{T}_\omega^\star$, which will often be a singleton.

If G has a dependence structure linked to a tree $\mathcal{T}$, then the latter belongs to the set of maximum dependence trees $\boldsymbol{T}_\lambda^\star$ weighted by the upper tail dependence coefficients. If, moreover, the latter set is a singleton, then $\mathcal{T}$ can be identified as the unique maximum dependence tree. A sufficient condition is that the inequality on the right-hand side of Proposition 2.5 is strict.

Proposition 4.1. *Let G be an extreme value distribution with dependence structure linked to a tree $\mathcal{T} = (V,E)$ as in Definition 2.2. Then the tree $\mathcal{T}$ belongs to $\boldsymbol{T}_\lambda^\star$ in (4.1) with edge weights equal to the bivariate upper tail dependence coefficients of G, that is, $\sum_{(a,b)\in E} \lambda_{ab} \geq \sum_{(a',b')\in E'} \lambda_{a'b'}$ for any tree $\mathcal{T}' = (V,E')$. The set $\boldsymbol{T}_\lambda^\star$ is a singleton and thus equal to $\{\mathcal{T}\}$ if for every triple of distinct nodes $a,u,b \in V$ with u on the path between a and b we have $\lambda_{ab} < \min(\lambda_{au}, \lambda_{ub})$.*

By way of example, consider the 4-variate Hüsler–Reiss distribution G with variogram matrix $\boldsymbol{\Gamma}_1$ in (3.4). This G is structured along the tree $\mathcal{T}^{(b)}$ on the right-hand side of Fig. 1 with $(a,b,c,d) = (1,2,3,4)$. According to Table 1, the value of $S_{\mathcal{T}_\lambda}$ is maximized

uniquely at the true tree, with value 0.952. In general, however, the collection of trees $\boldsymbol{T}_\lambda^\star$ in Proposition 4.1 need not be singleton, not even if $0 < \lambda_{ab} < 1$ for all $(a,b) \in V \times V$ with $a \neq b$. For a counter-example, see the trivariate regularly varying Markov tree after Proposition 2.5, for which all pairwise tail dependence coefficients are equal to 1/2 and thus all trees have the same sum of edge weights.

In practice, the true extreme value distribution G is unknown and the edge weights must be estimated from the data. Assume we observe independent copies $\boldsymbol{\xi}^i = (\xi_{i,v}, v \in V)$, $i = 1, \ldots, n$, of the d-variate random vector $\boldsymbol{\xi} = (\xi_v, v \in V)$ with marginal distributions F_v for $v \in V$ such that the standardized vector $(1/(1 - F_v(\xi_v)), v \in V)$ belongs to the max-domain of attraction of G. Equivalently, we have relation (2.3) for $\boldsymbol{\xi}$ with ℓ equal to the stable tail dependence function of the random vector $\boldsymbol{Z}$ with distribution function G. Note that $\boldsymbol{\xi}$ does not necessarily belong to the domain of attraction of a multivariate extreme value distribution since the margins are not required to be attracted by univariate extreme value distributions. From Corollary 2.3, the Markov tree $\boldsymbol{Y}$ constructed in Definition 3.1 has the same stable tail dependence function as the random vector $\boldsymbol{Z}_M$ with distribution G_M in Corollary 3.2 since $\boldsymbol{Y} \in D(G_M)$. Moreover, for all $(a,b) \in E$, the pairs (Y_a, Y_b) and (Z_a, Z_b) are equal in distribution by construction. Consequently, the stable dependence functions of all adjacent pairs on $\mathcal{T}$ of $\boldsymbol{\xi}$, $\boldsymbol{Z}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}_M$ are identical. The same statement holds also for the upper tail dependence coefficients. Note that the increasing component-wise transformations $x_v \mapsto 1/(1 - F_v(x_v))$ have no effect on λ and τ. Thus we can estimate the edge weights directly from samples drawn from $\boldsymbol{\xi}$.

For $(a,b) \in E$, we take account of the empirical estimator of τ_{ab} for (ξ_a, ξ_b) given by

$$\hat{\tau}_{ab} = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} \left\{ \operatorname{sgn}(\xi_{i,a} - \xi_{j,a}) \operatorname{sgn}(\xi_{i,b} - \xi_{j,b}) \right\},$$

which, by classical theory of U-statistics, is consistent and asymptotically normal as $n \to \infty$; see, e.g., van der Vaart (1998, Theorem 12.3 and Example 12.5). Let $\mathbb{1}(\cdot)$ denote the indicator function. For the upper tail dependence coefficient defined in (2.4), we consider the empirical estimator of λ_{ab} for $a, b \in V$ defined by

$$\hat{\lambda}_{ab} = \frac{1}{k_\lambda} \sum_{i=1}^{n} \mathbb{1} \left\{ \hat{F}_{na}(\xi_{i,a}) \ge 1 - \frac{k_\lambda}{n}, \, \hat{F}_{nb}(\xi_{i,b}) \ge 1 - \frac{k_\lambda}{n} \right\}, \tag{4.2}$$

where $\hat{F}_{na}(x) = n^{-1} \sum_{j=1}^{n} \mathbb{1}(\xi_{j,a} \le x)$ is the empirical distribution function of ξ_a and $k_\lambda = k_\lambda(n)$ is an intermediate sequence such that $k_\lambda \to \infty$ and $k_\lambda / n \to 0$ as $n \to \infty$. Its consistency and asymptotic normality were shown, for example, in Schmidt & Stadtmüller (2006). Consequently, for $\hat{\omega} \in \{\hat{\lambda}, \hat{\tau}\}$, we define

$$\hat{\boldsymbol{T}}_\omega^\star = \arg\max_{\mathcal{T}} \hat{S}_{\mathcal{T}_\omega} = \arg\max_{\mathcal{T}=(V,E)} \left(\sum_{(a,b) \in E} \hat{\omega}_{ab} \right) \tag{4.3}$$

to be the collection of maximum dependence trees with respect to the estimated edge weights $\hat{\omega}_{ab}$. A generic tree in the set $\hat{\boldsymbol{T}}_\omega^\star$ will be denoted by $\hat{\mathcal{T}}_\omega^\star$.

Proposition 4.2. *Let $\boldsymbol{T}_\lambda^\star$ be the set of maximum dependence trees in* (4.1) *weighted by the tail dependence coefficients and let $\hat{\boldsymbol{T}}_\lambda^\star$ in* (4.3) *be its sample equivalent based on the empirical tail dependence coefficients in* (4.2)*. If $k_\lambda = k_\lambda(n)$ is an intermediate sequence such that $k_\lambda \to \infty$ and $k_\lambda / n \to 0$ as $n \to \infty$, then*

$$\mathbb{P}\left(\hat{\boldsymbol{T}}_\lambda^\star \subseteq \boldsymbol{T}_\lambda^\star \right) \to 1, \qquad n \to \infty.$$

Moreover, if $\boldsymbol{T}_\lambda^\star$ is a singleton, with unique element $\mathcal{T}_\lambda^\star$, then, with probability tending to one, $\hat{\boldsymbol{T}}_\lambda^\star$ is a singleton too and its unique element $\hat{\mathcal{T}}_\lambda^\star$ satisfies

$$\mathbb{P}\left(\hat{\mathcal{T}}_\lambda^\star = \mathcal{T}_\lambda^\star\right) \to 1, \qquad n \to \infty.$$

By the consistency of the empirical estimator of Kendall's τ (van der Vaart, 1998, Example 12.5), a similar result as in Proposition 4.2 holds for $\hat{\boldsymbol{T}}_\tau^\star$ as well. The assertion and the proof are completely similar to those for λ and are omitted for brevity.

Once all edge weights $\hat{\omega}_{ab}$ have been computed, Prim's algorithm is used to find a global optimizer $\hat{\mathcal{T}}_\omega^\star$ in (4.3), the solution being unique if all edge weights are distinct. The time complexity of this procedure to find a tree with d nodes is $\mathcal{O}(d^2)$, see Prim (1957) and Papadimitriou & Steiglitz (1998, Theorem 12.2). The algorithm is given in Appendix A.2.

4.2 Estimation of pairwise dependence structures

Once a tree structure (V, E) is specified, we are able to construct a Markov tree according to Definition 3.1. For simplicity, we assume henceforth that the tree is fixed and non-random. In practice, the tree is inferred from the data, but under the conditions of Proposition 4.2, it is with probability tending to one equal to a fixed tree.

Since G_M in Corollary 3.2 is determined by the tree and the distributions of M_e with $e \in E$, it is sufficient to estimate the bivariate distributions of the adjacent pairs. Ultimately, this comes down to the estimation of bivariate dependence structures, as the univariate marginal distributions are standardized. We will estimate the bivariate stable tail dependence function ℓ_{ab} of each adjacent pair (Z_a, Z_b) with $(a, b) \in E$, which also equals the stable tail dependence function ℓ_{ab}^M of (Y_a, Y_b) and of (Z_a^M, Z_b^M) as discussed in Subsection 4.1.

The estimation of the stable tail dependence function dates back to Huang (1992) and Drees & Huang (1998) and has been extensively investigated in the literature, see for instance Peng & Qi (2008), Einmahl et al. (2012), Beirlant et al. (2016), Einmahl et al. (2016), Kiriliouk et al. (2018), and the references therein.

In our framework, for each $e = (a, b) \in E$, we assume that the stable tail dependence function ℓ_e of (Z_a, Z_b) belongs to a parametric family $\{(x, y) \mapsto \ell(x, y; \boldsymbol{\beta}_e), \boldsymbol{\beta}_e \in \boldsymbol{B}_e\}$, where the parameter $\boldsymbol{\beta}_e = (\beta_{e,1}, \ldots, \beta_{e,p_e})^\top$ belongs to some parameter space $\boldsymbol{B}_e \subseteq \mathbb{R}^{p_e}$ of dimension $p_e \geq 1$. The stable tail dependence functions for different pairs are allowed to belong to different parametric families. Assume that the true parameter $\boldsymbol{\beta}_e^0$ of ℓ_e belongs to the interior $\boldsymbol{B}_e^o$ of $\boldsymbol{B}_e$. Let $\boldsymbol{\beta}_0 = \left(\boldsymbol{\beta}_e^0, e \in E\right)$ be the $(\sum_{e \in E} p_e) \times 1$ column vector of true parameters. It is to be noted that Definition 2.2 of a tree-structured extreme-value distribution does not pose any constraints whatsoever on the combination of parametric families and parameter values. This flexibility greatly facilitates model construction and statistical inference.

For each $e = (a, b) \in E$, we estimate the parameter $\boldsymbol{\beta}_e^0$ using only the data in components a and b. Define

$$\hat{\boldsymbol{\beta}}_n = (\hat{\boldsymbol{\beta}}_{n,e}, e \in E)$$

to be the estimator of $\boldsymbol{\beta}_0$, where $\hat{\boldsymbol{\beta}}_{n,e}$ is any consistent estimator of $\boldsymbol{\beta}_e^0$ for $e \in E$. Since the convergence in probability of a sequence of vectors is equivalent to convergence of every one of the component sequences separately (van der Vaart, 1998, Theorem 2.7), the estimator $\hat{\boldsymbol{\beta}}_n$ is jointly consistent provided the estimators of $\boldsymbol{\beta}_e^0$ for $e \in E$ are all consistent. The analogous statement for convergence in distribution is, however, not true. Still, we show that the joint estimator is asymptotically normal if the edge-wise estimators have a certain asymptotic expansion in terms of the empirical stable tail dependence function, which is recalled below. The expansion is shared by the estimators proposed in Einmahl et al. (2008), Einmahl et al. (2012) and Einmahl et al. (2018), as we will show as well.

Given an independent random sample $\boldsymbol{\xi}^i = (\xi_{i,v}, v \in V)$, $i = 1, \ldots, n$, let R_v^i denote the rank of $\xi_{i,v}$ among $\xi_{1,v}, \ldots, \xi_{n,v}$ for $v \in V$. The empirical stable tail dependence function $\hat{\ell}_{n,k}$ is

$$\hat{\ell}_{n,k}(\boldsymbol{x}) = \frac{1}{k} \sum_{i=1}^{n} \mathbb{1}\left(R_1^i > n + \frac{1}{2} - kx_1 \text{ or } \ldots \text{ or } R_d^i > n + \frac{1}{2} - kx_d \right), \qquad \boldsymbol{x} \in [0, \infty)^d,$$

where $k = k(n)$ is an intermediate sequence such that $k \to \infty$ and $k/n \to 0$ as $n \to \infty$ (Huang, 1992; Drees & Huang, 1998). The asymptotic distribution of $\hat{\ell}_{n,k}$ was studied in Einmahl et al. (2012) and Bücher et al. (2014) under the following assumption.

Assumption 4.3. Let $k = k(n)$ be an intermediate sequence such that $k \to \infty$ and $k/n \to 0$ as $n \to \infty$. There exists $\zeta > 0$ such that the following two asymptotic relations hold:

(C1) $t^{-1} \mathbb{P}[1 - F_1(\xi_1) \le tx_1 \text{ or } \ldots \text{ or } 1 - F_d(\xi_d) \le tx_d] - \ell(\boldsymbol{x}) = O(t^{\zeta})$ as $t \downarrow 0$, uniformly in $\boldsymbol{x} \in \Delta_{d-1} = \left\{ \boldsymbol{w} \in [0,1]^d : w_1 + \cdots + w_d = 1 \right\}$;

(C2) $k = o\left(n^{2\zeta/(1+2\zeta)} \right)$ as $n \to \infty$.

Let W be a mean-zero Gaussian process on $[0, \infty)^d$ with continuous trajectories and covariance function

$$\mathbb{E}[W(\boldsymbol{x}) W(\boldsymbol{y})] = \ell(\boldsymbol{x}) + \ell(\boldsymbol{y}) - \ell(\boldsymbol{x} \vee \boldsymbol{y}), \qquad \boldsymbol{x}, \boldsymbol{y} \in [0, \infty)^d,$$

where $\boldsymbol{x} \vee \boldsymbol{y} = (\max\{x_j, y_j\})_{j=1}^d$. Further, let $\dot{\ell}_j(x)$ denote the right-hand partial derivative of ℓ with respect to the j-th coordinate, for $j \in \{1, \ldots, d\}$. Since ℓ is convex, this one-sided partial derivative always exists and is continuous in points where the ordinary two-sided partial derivative exists. Consider the stochastic process α on $[0, \infty)^d$ defined by

$$\alpha(\boldsymbol{x}) = W(\boldsymbol{x}) - \sum_{j=1}^{d} \dot{\ell}_j(\boldsymbol{x}) W(x_j \boldsymbol{e}_j), \qquad \boldsymbol{x} \in [0, \infty)^d, \tag{4.4}$$

where $\boldsymbol{e}_j$ denotes the j-th canonical unit vector in $\mathbb{R}^d$. In Einmahl et al. (2012, Theorem 4.6), the asymptotic distribution of $\sqrt{k}(\hat{\ell}_{n,k} - \ell)$ is established with respect to the topology of uniform convergence on $[0,1]^d$. An additional condition needed on ℓ is that $\dot{\ell}_j$ is continuous in $\boldsymbol{x}$ whenever $x_j > 0$. For certain models, however, this condition fails; a counter-example is the asymmetric logistic distribution in Example 3.4. A more general result is given in Bücher et al. (2014, Theorem 5.1), who establish weak convergence of $\sqrt{k}(\hat{\ell}_{n,k} - \ell)$ in a slightly weaker topology but without additional differentiability conditions on ℓ. The topology is based on epi- and hypographs of functions, whence the name hypi-topology. What concerns us here is that the topology is still strong enough to yield weak convergence in certain L^p-spaces and thus of certain integrals of the empirical stable tail dependence function. This will be crucial when studying parameter estimates based on $\hat{\ell}_{n,k}$.

Theorem 4.4. *Let $\mu_1, \ldots, \mu_q$ be finite Borel measures on $[0,1]^d$ such that for all $j \in \{1, \ldots, d\}$ and $r \in \{1, \ldots, q\}$, the set of points $\boldsymbol{x}$ with $x_j > 0$ in which $\dot{\ell}_j$ is not continuous is a μ_r-null set. Let $\psi_r \in L^2(\mu_r)$ for all $r \in \{1, \ldots, q\}$. Then as $n \to \infty$, we have*

$$\left(\int_{[0,1]^d} \sqrt{k} \left(\hat{\ell}_{n,k}(\boldsymbol{x}) - \ell(\boldsymbol{x}) \right) \psi_r(\boldsymbol{x}) \, \mu_r(\mathrm{d}\boldsymbol{x}) \right)_{r=1}^{q} \xrightarrow{d} \left(\int_{[0,1]^d} \alpha(\boldsymbol{x}) \, \psi_r(\boldsymbol{x}) \, \mu_r(\mathrm{d}\boldsymbol{x}) \right)_{r=1}^{q}. \tag{4.5}$$

The limit is centered multivariate normal with covariance matrix $\boldsymbol{\Sigma} = (\sigma_{rs})_{r,s=1}^q$ having entries

$$\sigma_{rs} = \int_{[0,1]^d} \int_{[0,1]^d} \mathbb{E}[\alpha(\boldsymbol{x}) \alpha(\boldsymbol{y})] \, \psi_r(\boldsymbol{x}) \, \psi_s(\boldsymbol{y}) \, \mu_r(\mathrm{d}\boldsymbol{x}) \, \mu_s(\mathrm{d}\boldsymbol{y}).$$

Remark 4.5. The measures μ_r in Theorem 4.4 can be discrete, in which case (4.5) states the asymptotic normality of the finite-dimensional distributions of $\sqrt{k}\{\hat{\ell}_{n,k}(\boldsymbol{x}) - \ell(\boldsymbol{x})\}$, at least in points $\boldsymbol{x} \in [0,1]^d$ such that for every $j \in \{1,\dots,d\}$ we have either $x_j = 0$ or $\dot{\ell}_j$ is continuous in $\boldsymbol{x}$.

In the bivariate case, the non-parametric estimator $\hat{\ell}_{n,k,e}(x_a, x_b)$ of $\ell_e(x_a, x_b)$ for $e = (a,b) \in E$ becomes

$$\hat{\ell}_{n,k,e}(x_a, x_b) = \frac{1}{k}\sum_{i=1}^{n} \mathbb{1}\left(R_a^i > n + \frac{1}{2} - kx_a \text{ or } R_b^i > n + \frac{1}{2} - kx_b\right) = \hat{\ell}_{n,k}(x_a\boldsymbol{e}_a + x_b\boldsymbol{e}_b).$$

As the formula indicates, it arises from $\hat{\ell}_{n,k}$ by setting $x_j = 0$ whenever $j \neq a$ and $j \neq b$.

Assumption 4.6. There exist an intermediate sequence $k = k(n)$ and, for every $e = (a,b) \in E$, a finite Borel measure ν_e on $[0,1]^2$ and a vector of functions $\boldsymbol{\psi}_e = (\psi_{e,1}, \dots, \psi_{e,p_e})^\top : [0,1]^2 \to \mathbb{R}^{p_e}$ in $L^2(\nu_e)$ such that the following two properties hold:

(i) As $n \to \infty$, we have

$$\sqrt{k}\left(\hat{\boldsymbol{\beta}}_{n,e} - \boldsymbol{\beta}_e^0\right) = \int_{[0,1]^2} \sqrt{k}\left(\hat{\ell}_{n,k,e}(x_a, x_b) - \ell_e(x_a, x_b)\right) \boldsymbol{\psi}_e(x_a, x_b)\, \nu_e(\mathrm{d}(x_a, x_b)) + o_p(\mathbf{1}). \tag{4.6}$$

(ii) For $c \in \{a, b\}$, the set of points (x_a, x_b) such that $x_c > 0$ and such that the first-order partial derivative of ℓ_e with respect to x_c does not exist is a ν_e-null set.

Corollary 4.7. *Under Assumptions 4.3 and 4.6, for the same intermediate sequence $k(n)$, we have*

$$\sqrt{k}\left(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\right) \xrightarrow{d} \left(\int_{[0,1]^2} \alpha(x_a\boldsymbol{e}_a + x_b\boldsymbol{e}_b)\, \boldsymbol{\psi}_e(x_a, x_b)\, \nu_e(\mathrm{d}(x_a, x_b))\right)_{e=(a,b)\in E}, \qquad n \to \infty,$$

with α as in (4.4). The limit vector is centered multivariate normal with covariance matrix $\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_{ee'})_{e,e'\in E}$ whose block indexed by $e = (a,b)$ and $e' = (a',b')$ in E is of dimension $p_e \times p_{e'}$ and is given by

$$\boldsymbol{\Sigma}_{ee'} = \int_{[0,1]^2}\int_{[0,1]^2} \mathbb{E}\left[\alpha(x_a\boldsymbol{e}_a + x_b\boldsymbol{e}_b)\, \alpha(x_{a'}\boldsymbol{e}_{a'} + x_{b'}\boldsymbol{e}_{b'})\right] \cdot \boldsymbol{\psi}_e(x_a, x_b)\, \boldsymbol{\psi}_e(x_{a'}, x_{b'})^\top\, \nu_e(\mathrm{d}(x_a, x_b))\, \nu_{e'}(\mathrm{d}(x_{a'}, x_{b'})).$$

Corollary 4.7 allows for a combination of different estimators for the edge parameters $\boldsymbol{\beta}_e$, as long as each of them satisfies Assumption 4.6. In the next examples, we discuss three such estimators.

Example 4.8 (M-estimator)**.** The idea behind the M-estimator proposed in Einmahl et al. (2012) is that of matching weighted integrals of the empirical stable tail dependence function with their model-based values. For integer $q_e \geq p_e$, let $\boldsymbol{g}_e = (g_{e,1}, \dots, g_{e,q_e})^\top : [0,1]^2 \to \mathbb{R}^{q_e}$ be a column vector of Lebesgue square-integrable functions such that the map $\boldsymbol{\varphi}_e : \boldsymbol{B}_e \to \mathbb{R}^{q_e}$ defined by

$$\boldsymbol{\varphi}_e(\boldsymbol{\beta}) = \int_{[0,1]^2} \ell_e(x_a, x_b; \boldsymbol{\beta})\, \boldsymbol{g}_e(x_a, x_b)\, \mathrm{d}\, x_a\, \mathrm{d}\, x_b$$

is a homeomorphism between $\boldsymbol{B}_e^o$ and $\boldsymbol{\varphi}_e(\boldsymbol{B}_e^o)$. For $e = (a,b) \in E$, the M-estimator $\hat{\boldsymbol{\beta}}_{n,k,e}^{\mathrm{M}}$ of $\boldsymbol{\beta}_e^0$ is a minimizer of the criterion function

$$Q_{n,k,e}(\boldsymbol{\beta}) = \sum_{m=1}^{q_e}\left[\int_{[0,1]^2} \left\{\hat{\ell}_{n,k,e}(x_a, x_b) - \ell_e(x_a, x_b; \boldsymbol{\beta})\right\} g_{e,m}(x_a, x_b)\, \mathrm{d}\, x_a\, \mathrm{d}\, x_b\right]^2.$$

Suppose that $\boldsymbol{\varphi_e}$ is twice continuously differentiable and that the $q_e \times p_e$ Jacobian matrix $\nabla\boldsymbol{\varphi}_e(\boldsymbol{\beta}_e^0)$ has rank p_e. Following the proof of Theorem 4.2 in Einmahl et al. (2012), it can be shown that expansion (4.6) holds with ν_e the two-dimensional Lebesgue measure and with

$$\boldsymbol{\psi}_e(x_a, x_b) = \left\{\nabla\boldsymbol{\varphi}_e(\boldsymbol{\beta}_e^0)^\top \nabla\boldsymbol{\varphi}_e(\boldsymbol{\beta}_e^0)\right\}^{-1} \nabla\boldsymbol{\varphi}_e(\boldsymbol{\beta}_e^0)^\top \boldsymbol{g}_e(x_a, x_b). \tag{4.7}$$

By convexity, the function $\dot{\ell}_e$ is differentiable Lebesgue almost everywhere. Corollary 4.7 thus applies with ν_e and $\boldsymbol{\psi}_e$ as stated.

Example 4.9 (Method of moments)**.** The moments estimator of Einmahl et al. (2008) arises as the special case $q_e = p_e$ of the M-estimator of Einmahl et al. (2012) discussed in Example 4.8 above. Rather than as a minimizer of a criterion function, the estimator $\hat{\boldsymbol{\beta}}_{n,k,e}^{\mathrm{MM}}$ is defined as the solution of the equation

$$\int_{[0,1]^2} \hat{\ell}_{n,k,e}(x_a, x_b)\, \boldsymbol{g}_e(x_a, x_b)\, \mathrm{d}\, x_a\, \mathrm{d}\, x_b = \boldsymbol{\varphi}_e\left(\hat{\boldsymbol{\beta}}_{n,k,e}^{\mathrm{MM}}\right),$$

The Jacobian matrix $\nabla\boldsymbol{\varphi}_e(\boldsymbol{\beta}_e^0)$ being an invertible square $p_e \times p_e$ matrix, the function $\boldsymbol{\psi}_e$ in (4.7) simplifies to

$$\boldsymbol{\psi}_e(x_a, x_b) = \left\{\nabla\boldsymbol{\varphi}_e(\boldsymbol{\beta}_e^0)\right\}^{-1} \boldsymbol{g}_e(x_a, x_b).$$

Example 4.10 (Weighted least squares estimators)**.** The weighted least squares estimator studied in Einmahl et al. (2018) avoids the integration of stable tail dependence functions, facilitating computations. Let $q_e \geq p_e$ and let $\boldsymbol{c}_e^1, \ldots, \boldsymbol{c}_e^{q_e}$ be q_e points in $(0,1]^2$. For $\boldsymbol{\beta}_e \in \boldsymbol{B}_e$, set

$$\hat{\boldsymbol{L}}_{n,k,e} = \left(\hat{\ell}_{n,k,e}\left(\boldsymbol{c}_e^m\right), m = 1, \ldots, q_e\right), \qquad \boldsymbol{L}_e(\boldsymbol{\beta}_e) = \left(\ell_e\left(\boldsymbol{c}_e^m; \boldsymbol{\beta}_e\right), m = 1, \ldots, q_e\right).$$

Assume $\boldsymbol{L}_e$ is a homeomorphism from $\boldsymbol{B}_e$ to $\boldsymbol{L}_e(\boldsymbol{B}_e)$. Let $\boldsymbol{\Omega}_e(\boldsymbol{\beta}_e)$ be a $q_e \times q_e$ dimensional symmetric, positive definite matrix, twice continuously differentiable as a function of $\boldsymbol{\beta}_e$, and with a smallest eigenvalue that remains bounded away from zero uniformly in $\boldsymbol{\beta}_e$. The weighted least squares estimator $\hat{\boldsymbol{\beta}}_{n,k,e}^{\mathrm{WLS}}$ of $\boldsymbol{\beta}_e^0$ is defined as

$$\hat{\boldsymbol{\beta}}_{n,k,e}^{\mathrm{WLS}} = \operatorname*{arg\,min}_{\boldsymbol{\beta}_e \in \boldsymbol{B}_e} \left(\hat{\boldsymbol{L}}_{n,k,e} - \boldsymbol{L}_e(\boldsymbol{\beta}_e)\right)^\top \boldsymbol{\Omega}_e(\boldsymbol{\beta}_e) \left(\hat{\boldsymbol{L}}_{n,k,e} - \boldsymbol{L}_e(\boldsymbol{\beta}_e)\right).$$

In the numerical examples in Sections 5 and 6, $\boldsymbol{\Omega}_e(\boldsymbol{\beta}_e)$ is set to be the $q_e \times q_e$ identity matrix.

Assume that $\boldsymbol{L}_e$ is twice continuously differentiable on a neighbourhood of $\boldsymbol{\beta}_e^0$ and that the $q_e \times p_e$ Jacobian matrix $\nabla\boldsymbol{L}_e(\boldsymbol{\beta}_e^0)$ has rank p_e. Define the $p_e \times q_e$-dimensional matrix

$$\boldsymbol{A}_e = \left\{\nabla\boldsymbol{L}_e(\boldsymbol{\beta}_e^0)^\top \boldsymbol{\Omega}_e(\boldsymbol{\beta}_e^0) \nabla\boldsymbol{L}_e(\boldsymbol{\beta}_e^0)\right\}^{-1} \nabla\boldsymbol{L}_e(\boldsymbol{\beta}_e^0)^\top \boldsymbol{\Omega}_e(\boldsymbol{\beta}_e^0).$$

Assume that the first-order partial derivatives of ℓ_e exist in all q_e points $\boldsymbol{c}_e^m$. Then the vector $\sqrt{k}\left(\hat{\boldsymbol{L}}_{n,k,e} - \boldsymbol{L}_e(\boldsymbol{\beta}_e^0)\right)$ is asymptotic normal by Theorem 4.4 and Remark 4.5. By Theorem 2 in Einmahl et al. (2018), we have

$$\sqrt{k}\left(\hat{\boldsymbol{\beta}}_{n,k,e}^{\mathrm{WLS}} - \boldsymbol{\beta}_e^0\right) = \boldsymbol{A}_e \sqrt{k}\left(\hat{\boldsymbol{L}}_{n,k,e} - \boldsymbol{L}_e(\boldsymbol{\beta}_e^0)\right) + o_p(\mathbf{1}), \qquad n \to \infty.$$

This expansion is of the form (4.6) with ν_e the counting measure on the points $\boldsymbol{c}_e^1, \ldots, \boldsymbol{c}_e^{q_e}$ and with $\boldsymbol{\psi}_e = (\psi_{e,1}, \ldots, \psi_{e,p_e})^\top$ having component functions

$$\psi_{e,j}(x_a, x_b) = (\boldsymbol{A}_e)_{j,m} \quad \text{if} \quad (x_a, x_b) = \boldsymbol{c}_e^m.$$

Corollary 4.7 thus applies.

4.3 Measuring the goodness-of-fit

As soon as the bivariate margins of all the adjacent pairs on $\boldsymbol{Y}$ are estimated, an estimate of distribution function G_M can be obtained from Corollary 3.2 by replacing the unknown quantities in there by their estimates. Likewise, a model-based estimate $\hat{\lambda}_{ab}^M$ of the upper tail dependence coefficient λ_{ab}^M can be obtained from (2.10). Thus we take

$$\hat{D}_{\mathcal{T}} = \sum_{a,b \in V, (a,b) \notin E} |\hat{\lambda}_{ab}^M - \hat{\lambda}_{ab}| \tag{4.8}$$

as a consistent estimator of $D_{\mathcal{T}}$ defined in (3.2) to measure the goodness-of-fit, where $\hat{\lambda}_{ab}$ is the empirical estimator of λ_{ab} given in (4.2).

5 Simulation study

To evaluate the performance of the proposed approach, we perform our modelling procedure on random samples from the Hüsler–Reiss and asymmetric logistic extreme value distributions in Examples 3.3 and 3.4 respectively. To achieve a more realistic situation, we add some lighter-tailed noise. Specifically, starting from random vectors $\boldsymbol{Z}^i$, for $i = 1, \ldots, n$, drawn independently from a simple Hüsler–Reiss or asymmetric logistic distribution, we construct the samples from a random vector $\boldsymbol{\xi}$ by

$$\boldsymbol{\xi}^i = \boldsymbol{Z}^i + \boldsymbol{\varepsilon}^i, \qquad i = 1, \ldots, n, \tag{5.1}$$

where $\boldsymbol{\varepsilon}^i$, for $i = 1, \ldots, n$, are independent of $\boldsymbol{Z}^i$ and are themselves independent and identically distributed random vectors with independent Fréchet(2) distributed entries, i.e., $\mathbb{P}(\varepsilon_v^i \leq x) = \exp(-1/x^2)$ for $x > 0$ and $v \in V$.

The measure $\hat{D}_{\mathcal{T}}$ in (4.8) is calculated to assess the model's goodness-of-fit. The resulting tree-based models are applied to the estimation of the rare event probabilities in (1.1). To do that, we need take account of both the dependence structure and the margins of the sample. Thus it is more convenient to work with the assumption that a general random vector $\boldsymbol{\xi}$ is in the domain of attraction of a multivariate extreme value distribution H with margins H_v for $v \in V$. This is equivalent to the assumption $(1/(1-F_v(\xi_v)), v \in V) \in D(G)$ with G the simple multivariate extreme value distribution that relates to H through $G(\boldsymbol{x}) = H(H_v^{\leftarrow}(e^{-1/x_v}), v \in V)$ (the superscript arrow denoting the functional inverse), together with the condition that F_v is in the domain of attraction of the univariate extreme value distribution H_v for $v \in V$ respectively. Then by Eq. (8.81) in Beirlant et al. (2004), we have

$$\begin{aligned} \mathbb{P}\left(\xi_1 > u_1 \text{ or } \ldots \text{ or } \xi_d > u_d\right) &\approx 1 - G(-1/\log F_v(u_v),\, v \in V) \\ &\approx 1 - G_M(-1/\log F_v(u_v),\, v \in V) \end{aligned} \tag{5.2}$$

for thresholds u_v such that all probabilities $F_v(u_v)$ are close to unity.

For the random vector $\boldsymbol{\xi}$ with copies constructed in (5.1), we know that $\boldsymbol{\xi} \in D(G)$ with G equal to the common distribution of $\boldsymbol{Z}^i$, $i = 1, \ldots, n$. Moreover, the marginal distributions F_v, $v \in V$, belong to the domain of attraction of the unit-Fréchet distribution G_v, so that $-\log F_v(u_v) \approx 1/u_v$ provided u_v is sufficiently large. Therefore, the rare event probability on the left-hand side of (5.2) can be approximated by $1 - G_M(\boldsymbol{u})$ by plugging the marginal approximations into the quantity on the right-hand side. We will show boxplots of the logarithm of the relative approximation error

$$\mathrm{AE} = \log \frac{1 - \hat{G}_M(\boldsymbol{u})}{1 - F(\boldsymbol{u})}, \tag{5.3}$$

where $\hat{G}_M$ is the estimate of G_M based on the three-step procedure in Section 4 and F is the joint cumulative distribution function of $\boldsymbol{\xi}$. For simplicity, three components are considered simultaneously. In each experiment, we take u_j, for $j = 1, 2, 3$, as the 0.999 empirical marginal quantiles and $u_j = \infty$ for $j = 4, \ldots, d$. All simulations are done in `R`. Realizations of samples of the Hüsler–Reiss distribution are simulated using the package `graphicalExtremes` (Engelke et al., 2019). Samples of the asymmetric logistic distribution are generated through (3.5). The true probability $1 - F(\boldsymbol{u})$ is calculated by numerical integration as implemented in the package `cubature` (Narasimhan et al., 2019) based on the convolution of $\boldsymbol{Z}$ and $\boldsymbol{\varepsilon}$: $F(\boldsymbol{u}) = \int_{[\mathbf{0},\boldsymbol{u}]} G(u_1 - x_1, \ldots, u_d - x_d) \prod_{v \in V} \left(2x_v^{-3} \exp(-1/x_v^2)\right) \mathrm{d}\,\boldsymbol{x}$.

5.1 Hüsler–Reiss distribution

The Hüsler–Reiss distribution may have a dependence structure linked to a tree, see Example 3.3. Random samples can be generated from either a general Hüsler–Reiss distribution or a structured one. Both cases are considered and experiments are based on 300 repetitions with samples of size $n = 1000$. The procedure is detailed in Appendix A.2.

For random samples generated from the 10-dimensional Hüsler–Reiss distribution with variogram matrix $\boldsymbol{\Gamma}_3$ in Table 4 and dependence structure linked to the tree given in Fig. 5 of Appendix A.2, the proportion of wrongly estimated edges

$$1 - \frac{|E_{\hat{\mathcal{T}}_\omega^\star} \cap E_{\mathcal{T}}|}{d-1}$$

for $\omega \in \{\lambda, \tau\}$ is recorded. For Kendall's tau, all trees $\hat{\mathcal{T}}_\tau^\star$ found in the 300 replications are the same as the true one, while for the upper tail dependence coefficient, there are about 23.2% of replications in which $\hat{\mathcal{T}}_\lambda^\star$ does not agree with the true tree in one out of nine edges and 1.3% of replications where two out of nine edges are wrong. This is partly caused by the selected $k_\lambda = 0.1n$ for the estimation of λ. Fig. 6 of Appendix A.2 shows an interesting tendency in the influence of k_λ on the number of replications where $\hat{\mathcal{T}}_\lambda^\star$ does not coincide with the true tree: the best proportion k_λ/n does not appear in the region close to zero but around 0.5, even if the estimation bias is large for such k_λ. This provides a practical argument for using Kendall's tau to learn the tree structure, even though τ is not a measure of tail dependence.

Recall that the limiting distribution G_M is also a Hüsler–Reiss distribution with variogram matrix $\boldsymbol{\Gamma}_M$ which can be recovered from the estimated parameters of bivariate stable tail dependence functions and the selected tree structure, see Example 3.3. A Hüsler–Reiss distribution is totally determined by its variogram matrix. Therefore, the distance between the variogram matrix $\hat{\boldsymbol{\Gamma}}_M = (\hat{\gamma}_{ij}^M)_{i,j \in V}$ of the tree-based model $\hat{G}_M$ and the true variogram matrix $\boldsymbol{\Gamma} = (\gamma_{ij})_{i,j \in V}$ can also be used to measure the approximation error. We compare the variogram matrices through the Frobenius norm of their difference. Boxplots of the relative distance of $\hat{\boldsymbol{\Gamma}}_M$ to $\boldsymbol{\Gamma}$ given by

$$\frac{\|\hat{\boldsymbol{\Gamma}}_M - \boldsymbol{\Gamma}\|_F}{\|\boldsymbol{\Gamma}\|_F} = \frac{\left[\sum_{i=1}^d \sum_{j=1}^d (\hat{\gamma}_{ij}^M - \gamma_{ij})^2\right]^{1/2}}{\left[\sum_{i=1}^d \sum_{j=1}^d \gamma_{ij}^2\right]^{1/2}} \tag{5.4}$$

based on 300 replications are given in Fig. 2. The parameters of the stable tail dependence functions are estimated using the moments estimator, the M-estimator and the weighted least squares estimator discussed in Subsection 4.2. The moments estimator and M-estimator turned out to have quite similar performances, which is why we only show the results for the M-estimator and weighted least squares estimator. The two types of edge weights seem comparable in the sense of approximation distance even though λ is not as good as τ in

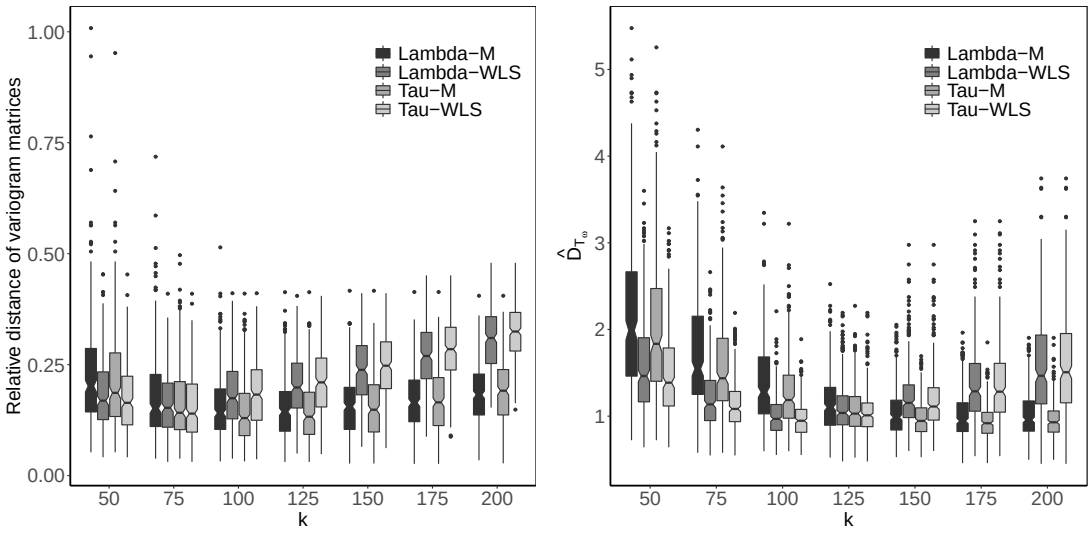

Fig. 2 Boxplots of $\frac{\|\hat{\mathbf{\Gamma}}_M - \mathbf{\Gamma}\|_F}{\|\mathbf{\Gamma}\|_F}$ (left) and $\hat{D}_{\mathcal{T}_\omega}$ in (4.8) (right) for 10-dimensional Hüsler–Reiss distribution with variogram matrix $\mathbf{\Gamma}_3$ based on $\hat{\mathcal{T}}_\lambda^\star$ (Lambda) and $\hat{\mathcal{T}}_\tau^\star$ (Tau), where bivariate stable tail dependence functions are fitted based on M-estimator (M) and weighted least squares estimator (WLS) at $k = 50, 75, \ldots, 200$ (300 samples of size $n = 1000$)

recovering the true tree structure. Compared with the M-estimator, the weighted least squares estimator is more sensitive to the choice of k. Boxplots of $\hat{D}_{\mathcal{T}}$ basically correspond to those of the relative distance of $\hat{\mathbf{\Gamma}}_M$ to $\mathbf{\Gamma}$, supporting the use of $\hat{D}_{\mathcal{T}}$ as a goodness-of-fit measure.

The model based on $\hat{\mathcal{T}}_\tau^\star$ is used to estimate the probability in (1.1). From Fig. 3 we see that the proposed models underestimates the probability in (1.1). This is mainly caused by the underestimation of the variogram matrix.

For samples generated from the 10-dimensional Hüsler–Reiss distribution with variogram matrix $\mathbf{\Gamma}_4$ which is not necessarily tree-structured, the simulation results have similar patterns as those in the first case but with larger errors. The variogram matrix $\mathbf{\Gamma}_4$ and boxplots can be found in Table 4 and Fig. 7–8 in Appendix A.2.

5.2 Asymmetric logistic distribution

In reality, we do not know the true distribution G. The bivariate distribution family we use to model adjacent pairs may be different from the true one. To see the estimation error in this scenario, we generate samples from the 5-dimensional asymmetric logistic distribution in Example 3.4 (a special max-linear model) with randomly generated parameter $\mathbf{\Psi}_3 = (0.763, 0.835, 0.602, 0.747, 0.859)$. Still, we model the adjacent pairs on the constructed Markov tree by bivariate Hüsler–Reiss distributions.

According to Fig. 4, the two types of edges weights, λ and τ, still exhibit a similar performance. Larger k leads to a higher accuracy for both estimators in terms of $\hat{D}_{\mathcal{T}_\omega}$, while the weighted least squares estimator behaves better than the M-estimator for the chosen k, as can also be seen from the approximation error on the right-hand panel in the figure. However, compared with the experiments for the Hüsler–Reiss distribution above, the approximation errors are much larger.

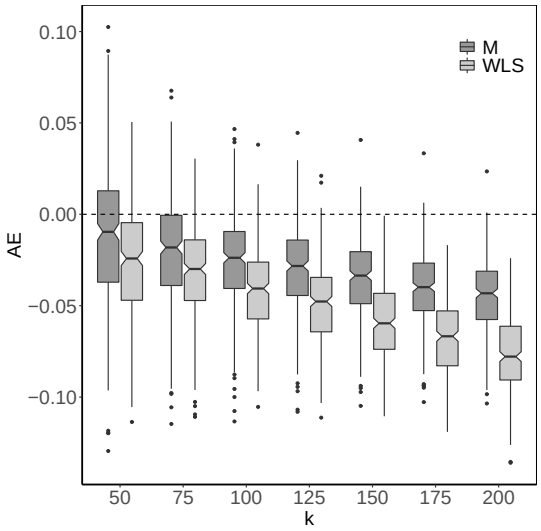

Fig. 3 Boxplot of the approximation errors of the rare event probability in (5.3) based on $\hat{\mathcal{T}}_\tau^\star$ for 10-dimensional Hüsler–Reiss distribution with variogram matrix $\mathbf{\Gamma}_3$, where bivariate stable tail dependence functions are fitted using the M-estimator (M) and the weighted least squares estimator (WLS) at $k = 50, 75, \ldots, 200$, the levels u_1, u_2, u_3 are taken as the 0.999 empirical marginal quantiles and $u_j = \infty$ for $j = 4, \ldots, d$ (300 samples of size $n = 1000$)

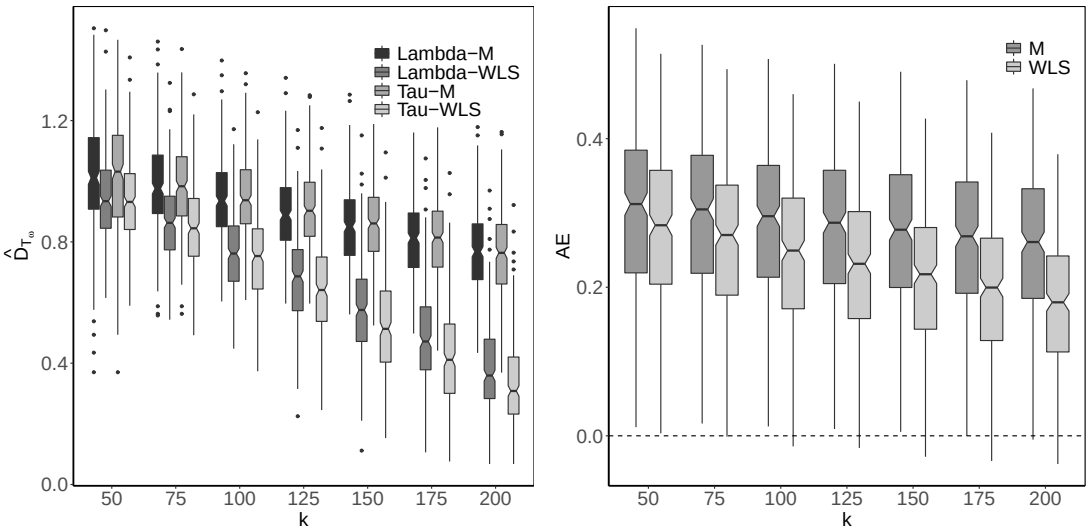

Fig. 4 Boxplots of $\hat{D}_{\mathcal{T}_\omega}$ (left) in (4.8) based on $\hat{\mathcal{T}}_\lambda^\star$ (Lambda) and $\hat{\mathcal{T}}_\tau^\star$ (Tau) and the approximation errors (right) of the rare event probability in (5.3) based on $\hat{\mathcal{T}}_\tau^\star$ for 5-dimensional asymmetric logistic distribution with parameter vector $\mathbf{\Psi}_3$ in Section 5.2, where bivariate Hüsler–Reiss stable tail dependence functions are fitted with M-estimator (M) and weighted least squares estimator (WLS) at $k = 50, 75, \ldots, 200$ (300 samples of size $n = 1000$)

6 Application

In this section we consider the average daily discharges recorded at 31 gauging stations in the upper Danube basin covering parts of Germany, Austria and Switzerland. The topographic map can be found in Figure 1 of Asadi et al. (2015) and the data are available in the supplementary materials of that paper. The series at individual stations have lengths from 54 to 113 years, with 51 years of data for all stations from 1960 to 2010. The average daily discharges were measured in m^3/s, ranging from 20 to 1400. Following Asadi et al. (2015) and Engelke & Hitz (2020), we only consider the daily discharges in the summer months (June, July and August) to get rid of the seasonality. The information for summer discharges at each station is given in Table 5 of Appendix A.2.

For $i = 1, \dots, n$, denote the daily mean discharges at the station j on day i by $\xi_{i,j}$ and assume the distribution function of $\boldsymbol{\xi}^i = (\xi_{i,1}, \dots, \xi_{i,d})$ belongs to the domain of attraction of a multivariate extreme value distribution function. We are concerned about the probability that there will be a flood above a certain (high) threshold for at least one site within the considered stations, precisely, the probability in (1.1). In view of (5.2), the modelling can be split into two parts: fitting the marginal distributions and estimating the simple tree-based multivariate extreme distribution function G_M.

Extreme discharges often occur in clusters. To remove the clusters and create a sequence of independent and identically distributed variables, we follow the procedure in Asadi et al. (2015) for each of the 51 summer periods. It ranks the data within each series and starts from the day with the highest rank of the period across all series (only one series when extracting univariate events) and takes a window of 9 days. An event is formed by taking the largest observation within this window. Then the data in this window are deleted and the process is repeated to form a new event. All events are found when no window of 9 consecutive days remains.

6.1 Marginal fitting

Under the assumption that the distribution of $\boldsymbol{\xi}^i$ belongs to the domain of attraction of a multivariate extreme value distribution, the marginal distributions are also attracted by univariate extreme value distributions. In line with asymptotic theory, we model high-threshold excesses by the generalized Pareto distribution to obtain the margins F_j for $j = 1, \dots, d$. Since the extremal behavior from any available earlier data does not change relative to the 51 common years, see Asadi et al. (2015), all the available data for each station are used to do the marginal fitting.

For $j = 1, \dots, d$ and $0 < p < 1$, let $q_{j,p}$ be the empirical p-quantile of $\xi_{i,j}$ and put $\mathcal{I}_j = \{i \in \{1, \dots, n\} : \xi_{i,j} > q_{j,p}\}$. For each station j, let $\hat{\sigma}_j > 0$ and $\hat{\vartheta}_j \in \mathbb{R}$ be the maximum likelihood estimates of the scale and shape parameters of the generalized Pareto distribution fitted to the excesses $\xi_{i,j} - q_{j,p}$ for $i \in \mathcal{I}_j$. Let N_j and n_j denote the sample size at station j and the number of elements in $\mathcal{I}_j$, respectively. Then the tail distribution function $1 - F_j$ is estimated by

$$1 - \hat{F}_j(x) = \frac{n_j}{N_j} \left\{ 1 - \mathrm{GPD}(x - q_{j,p}; \hat{\sigma}_j, \hat{\vartheta}_j) \right\}, \tag{6.1}$$

where $\mathrm{GPD}(\cdot; \sigma, \vartheta)$ denotes the cumulative distribution function of the generalized Pareto distribution with scale parameter $\sigma > 0$ and shape parameter $\vartheta \in \mathbb{R}$.

In pursuit of a good fit, an appropriate threshold $q_{j,p}$ has to be found. By inspection of the mean residual life plot, it was decided to set the threshold for margin j equal to $q_{j,p}$ for $p = 0.9$. The number of declustered excesses used for the estimation at each station as well as the corresponding thresholds $q_{j,p}$ and maximum likelihood estimates of $\hat{\sigma}_j$ and $\hat{\vartheta}_j$ of fitted

generalized Pareto distribution for $j = 1, \ldots, 31$ are given in Tables 6 and 7 of Appendix A.3 respectively.

6.2 Estimation of the dependence structure

In this subsection, we approximate the dependence structure between extremes of the 31 gauging stations by a multivariate extreme value distributions G_M related to a Markov tree. After declustering, there are $N = 428$ independent observations from the 51 years of data for all stations. The tree structure corresponding to the physical flow connections at these stations is presented in Fig. 9, while the tree structures $\hat{\mathcal{T}}_\lambda^\star$ and $\hat{\mathcal{T}}_\tau^\star$ (4.3) are given in Fig. 10 of Appendix A.3. We construct the extreme value distribution G_M by fitting the Hüsler–Reiss distribution to all pairs of connected variables and estimate their parameters by the moments estimator, the M-estimator and the weighted least squares estimator with $k = 65$. The estimated GPD-based marginal probability in (6.1) and the probability $\mathbb{P}(\xi_4 > u_4 \text{ or } \xi_7 > u_7 \text{ or } \xi_{13} > u_{13})$ of flooding at one or more stations 4, 7 and 13, which are connected differently on the maximum dependence trees $\hat{\mathcal{T}}_\lambda^\star$ and $\hat{\mathcal{T}}_\tau^\star$ (Fig. 10 of Appendix A.3), are given in Tables 2 and 3 respectively, where u_j for $j = 4, 7, 13$ are taken as the 0.95, 0.99, 0.995 and 0.999 empirical quantiles of the 51-year samples. Compared with the empirical probabilities, the tree-based model yields higher estimates of the flooding probabilities. As shown in Table 2, this may be partly due to the difference between the non-parametric and GPD-based estimates of the marginal probabilities $F_j(u_j)$ for $j = 4, 7, 13$.

Table 2: GPD-based estimates $1 - \hat{F}_j(u_j)$ of marginal tail probability in (6.1) with u_j taken as the 0.95, 0.99, 0.995 and 0.999 empirical quantiles of the 51-year samples for $j = 4, 7, 13$

u_j	$q_{j,0.95}$	$q_{j,0.99}$	$q_{j,0.995}$	$q_{j,0.999}$
$1 - \hat{F}_4$	6.52%	1.87%	0.99%	0.22%
$1 - \hat{F}_7$	6.66%	1.69%	1.03%	0.15%
$1 - \hat{F}_{13}$	7.02%	1.90%	1.18%	0.33%

Table 3: Empirical and model-based estimates of the flooding probability (1.1) over the 0.95, 0.99, 0.995 and 0.999 empirical quantiles of the 51-year samples at stations 4, 7 and 13 in the upper Danube basin, where bivariate Hüsler–Reiss stable tail dependence functions of connected pairs on maximum dependence tree $\hat{\mathcal{T}}_\lambda^\star$ and $\hat{\mathcal{T}}_\tau^\star$ are fitted with moments estimator (MM), M-estimator (M) and weighted least squares estimator (WLS) at $k = 65$

u_j		$q_{j,0.95}$	$q_{j,0.99}$	$q_{j,0.995}$	$q_{j,0.999}$
Empirical		8.44%	1.88%	1.02%	0.21%
$\hat{\mathcal{T}}_\lambda^\star$	MM	9.71%	2.65%	1.58%	0.38%
	M	9.72%	2.65%	1.58%	0.38%
	WLS	9.62%	2.62%	1.56%	0.37%
$\hat{\mathcal{T}}_\tau^\star$	MM	9.54%	2.61%	1.55%	0.37%
	M	9.55%	2.61%	1.55%	0.37%
	WLS	9.26%	2.53%	1.50%	0.37%

Acknowledgments

The research of Shuang Hu is financially supported by the State Scholarship Fund (CSC No.202106990036) from the China Scholarship Council.

References

Asadi, P., Davison, A. C., & Engelke, S. (2015). Extremes on river networks. *The Annals of Applied Statistics*, 9(4), 2023–2050.

Asenova, S., Mazo, G., & Segers, J. (2021). Inference on extremal dependence in the domain of attraction of a structured Hüsler–Reiss distribution motivated by a markov tree with latent variables. *Extremes*, 24, 461–500.

Asenova, S. & Segers, J. (2021). Extremes of Markov random fields on block graphs. *ArXiv:2112.04847*.

Ballani, F. & Schlather, M. (2011). A construction principle for multivariate extreme value distributions. *Biometrika*, 98(3), 633–645.

Beirlant, J., Escobar-Bach, M., Goegebeur, Y., & Guillou, A. (2016). Bias-corrected estimation of stable tail dependence function. *Journal of Multivariate Analysis*, 143, 453–466.

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., De Waal, D., & Ferro, C. (2004). *Statistics of Extremes: Theory and Applications*. John Wiley & Sons.

Boldi, M.-O. & Davison, A. C. (2007). A mixture model for multivariate extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(2), 217–229.

Bücher, A., Segers, J., & Volgushev, S. (2014). When uniform weak convergence fails: empirical processes for dependence functions and residuals via epi- and hypographs. *The Annals of Statistics*, 42(4), 1598–1634.

Chow, C.-S. & Liu, C. (1968). Approximating discrete probability distributions with dependence trees. *IEEE Transactions on Information Theory*, 14(3), 462–467.

Coles, S. G. (2001). *An Introduction to Statistical Modelling of Extreme Values*. London: Springer.

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

Cooley, D., Davis, R. A., & Naveau, P. (2010). The pairwise beta distribution: A flexible parametric multivariate model for extremes. *Journal of Multivariate Analysis*, 101, 2103–2117.

Davison, A. C. & Huser, R. (2015). Statistics of extremes. *Annual Review of Statistics and Its Application*, 2(1), 203–235.

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

de Haan, L. & Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. New York: Springer.

Deheuvels, P. (1991). On the limiting behavior of the pickands estimator for bivariate extreme-value distributions. *Statistics & Probability Letters*, 12(5), 429–439.

Drees, H. & Huang, X. (1998). Best attainable rates of convergence for estimators of the stable tail dependence function. *Journal of Multivariate Analysis*, 64(1), 25–46.

Einmahl, J. H., Kiriliouk, A., Krajina, A., & Segers, J. (2016). An M-estimator of spatial tail dependence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(Part 1), 275–298.

Einmahl, J. H., Kiriliouk, A., & Segers, J. (2018). A continuous updating weighted least squares estimator of tail dependence in high dimensions. *Extremes*, 21(2), 205–233.

Einmahl, J. H., Krajina, A., & Segers, J. (2008). A method of moments estimator of tail dependence. *Bernoulli*, 14(4), 1003–1026.

Einmahl, J. H., Krajina, A., & Segers, J. (2012). An M-estimator for tail dependence in arbitrary dimensions. *The Annals of Statistics*, 40(3), 1764–1793.

Einmahl, J. H., Piterbarg, V. I., & De Haan, L. (2001). Nonparametric estimation of the spectral measure of an extreme value distribution. *The Annals of Statistics*, 29(5), 1401–1423.

Einmahl, J. H. & Segers, J. (2009). Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution. *The Annals of Statistics*, 37(5B), 2953–2989.

Engelke, S. & Hitz, A. S. (2020). Graphical models for extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(4), 871–932.

Engelke, S., Hitz, A. S., & Gnecco, N. (2019). *graphicalExtremes: Statistical Methodology for Graphical Extreme Value Models*. R package version 0.1.0.

Engelke, S. & Ivanovs, J. (2021). Sparse structures for multivariate extremes. *Annual Review of Statistics and Its Application*, 8, 241–270.

Engelke, S. & Volgushev, S. (2020). Structure learning for extremal tree models. *ArXiv:2012.06179*.

Gissibl, N. & Klüppelberg, C. (2018). Max-linear models on directed acyclic graphs. *Bernoulli*, 24(4A), 2693–2720.

Gissibl, N., Klüppelberg, C., & Otto, M. (2018). Tail dependence of recursive max-linear models with regularly varying noise variables. *Econometrics and Statistics*, 6, 149–167.

Gumbel, E. J. (1960). Bivariate exponential distributions. *Journal of the American Statistical Association*, 55(292), 698–707.

Hall, P. (1935). On representatives of subsets. *Jounal of the London Mathematical Society*, s1-10(1), 26–30.

Hanson, T. E., de Carvalho, M., & Chen, Y. (2017). Bernstein polynomial angular densities of multivariate extreme value distributions. *Statistics & Probability Letters*, 128, 60–66.

Horváth, G., Kovács, E., Molontay, R., & Nováczki, S. (2020). Copula-based anomaly scoring and localization for large-scale, high-dimensional continuous data. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 11(3), 1–26.

Huang, X. (1992). *Statistics of Bivariate Extremes*. PhD thesis, Erasmus University.

Hüsler, J. & Reiss, R.-D. (1989). Maxima of normal random vectors: between independence and complete dependence. *Statistics & Probability Letters*, 7(4), 283–286.

Kiriliouk, A., Rootzén, H., Segers, J., & Wadsworth, J. L. (2019). Peaks over thresholds modeling with multivariate generalized pareto distributions. *Technometrics*, 61(1), 123–135.

Kiriliouk, A., Segers, J., & Tafakori, L. (2018). An estimator of the stable tail dependence function based on the empirical beta copula. *Extremes*, 21(4), 581–600.

Kotz, S. & Nadarajah, S. (2000). *Extreme Value Distributions: Theory and Applications*. World Scientific.

Narasimhan, B., Johnson, S. G., Hahn, T., Bouvier, A., & Kiêu, K. (2019). *cubature: Adaptive Multivariate Integration over Hypercubes*. R package version 2.0.4.

Papadimitriou, C. H. & Steiglitz, K. (1998). *Combinatorial Optimization: Algorithms and Complexity*. Courier Corporation.

Papastathopoulos, I. & Strokorb, K. (2016). Conditional independence among max-stable laws. *Statistics & Probability Letters*, 108, 9–15.

Peng, L. & Qi, Y. (2008). Bootstrap approximation of tail dependence function. *Journal of Multivariate Analysis*, 99(8), 1807–1824.

Prim, R. (1957). Shortest connection networks and some generalizations. *The Bell System Technical Journal*, 36(6), 1389–1401.

Resnick, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes*. Springer-Verlag, New York.

Schmidt, R. & Stadtmüller, U. (2006). Non-parametric estimation of tail dependence. *Scandinavian Journal of Statistics*, 33(2), 307–335.

Segers, J. (2020). One-versus multi-component regular variation and extremes of Markov trees. *Advances in Applied Probability*, 52(3), 855–878.

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

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press.

Address

Shuang Hu,
School of Mathematics and Statistics, Southwest University, 400715 Chongqing, China.
LIDAM/ISBA, Université catholique de Louvain, B1348 Louvain-la-Neuve, Belgium.
E-mail address: hs1995@email.swu.edu.cn

A Appendix

A.1 Proofs

Proof of Proposition 2.3. By assumption, there exists a regularly varying Markov tree $\boldsymbol{Y}$ satisfying Condition 2.1 and (2.7) such that $\boldsymbol{Y} \in D(G)$, where G is a simple multivariate extremes value distribution transformed from H. Moreover, G and H have the same stable tail dependence functions. Thus it is sufficient to derive the result in terms of G. Let $v_1, \ldots, v_d$ be an arbitrary permutation of $1, \ldots, d$. Since $\boldsymbol{Y}$ satisfies Condition 2.1, it follows from (2.7) and Theorem 2 in Segers (2020) that

$$
\begin{aligned}
& n\,\mathbb{P}\left(\bigcup_{v=1}^{d}\{Y_v > n y_v\}\right) \\
&= n \sum_{i=1}^{d-1} \mathbb{P}(Y_{v_i} > n y_{v_i}, Y_{v_j} \le n y_{v_j}, j = i+1, \ldots, d) + n\,\mathbb{P}(Y_{v_d} > n y_{v_d}) \\
&= \sum_{i=1}^{d-1} n\,\mathbb{P}(Y_{v_i} > n y_{v_i})\,\mathbb{P}(Y_{v_j} \le n y_{v_j}, j = i+1, \ldots, d \mid Y_{v_i} > n y_{v_i}) + n\,\mathbb{P}(Y_{v_d} > n y_{v_d}) \\
&\to \sum_{i=1}^{d} y_{v_i}^{-1}\,\mathbb{P}\left(W \Theta_{v_i,v_j} \le y_{v_j}/y_{v_i}, j = i+1, \ldots, d\right)
\end{aligned}
\tag{A.1}
$$

as $n \to \infty$, where the probability in the sum (A.1) is interpreted as 1 for $i = d$. Recall that W is unit-Pareto distributed and is independent of $\boldsymbol{\Theta}_{v_i}$. By the fact that $1/W$ is uniformly distributed on $[0,1]$, we have

$$
\begin{aligned}
& y_{v_i}^{-1}\,\mathbb{P}\left(W \Theta_{v_i,v_j} \le y_{v_j}/y_{v_i}, j = i+1, \ldots, d\right) \\
&= y_{v_i}^{-1} \int_0^1 \mathbb{P}\left\{\Theta_{v_i,v_j} \le (y_{v_j} u)/y_{v_i}, j = i+1, \ldots, d\right\} \mathrm{d}\,u \\
&= y_{v_i}^{-1} \int_0^1 \mathbb{P}\left\{\max_{j=i+1,\ldots,d}\left(y_{v_j}^{-1}\Theta_{v_i,v_j}\right) \le y_{v_i}^{-1} u\right\} \mathrm{d}\,u \\
&= \int_0^{1/y_{v_i}} \mathbb{P}\left\{\max_{j=i+1,\ldots,d}\left(y_{v_j}^{-1}\Theta_{v_i,v_j}\right) \le u\right\} \mathrm{d}\,u \\
&= \int_0^{\infty} \mathbb{P}\left\{\max_{j=i+1,\ldots,d}\left(y_{v_j}^{-1}\Theta_{v_i,v_j}\right) \le u \le y_{v_i}^{-1}\Theta_{v_i,v_i}\right\} \mathrm{d}\,u \\
&= \mathbb{E}\left\{\max_{j=i,\ldots,d}\left(y_{v_j}^{-1}\Theta_{v_i,v_j}\right) - \max_{j=i+1,\ldots,d}\left(y_{v_j}^{-1}\Theta_{v_i,v_j}\right)\right\},
\end{aligned}
\tag{A.2}
$$

where the last step follows from $\int_0^\infty \mathbb{P}(A \le u \le B)\,\mathrm{d}\,u = \mathbb{E}[\int_0^\infty \mathbb{1}(A \le u \le B)\,\mathrm{d}\,u] = \mathbb{E}[\max(A,B) - A]$ for non-negative random variables A and B. By (2.1) with $a_{n,v} = n$ and $b_{n,v} = 0$, (A.1)–(A.2) and (2.2), we have

$$
\begin{aligned}
\ell(\boldsymbol{y}) &= -\log G(1/y_1, \ldots, 1/y_d) = \lim_{n\to\infty} n\,\mathbb{P}\left(\bigcup_{v=1}^{d}\{Y_v > n y_v\}\right) \\
&= \sum_{i=1}^{d} \mathbb{E}\left\{\max_{j=i,\ldots,d}\left(y_{v_j}\Theta_{v_i,v_j}\right) - \max_{j=i+1,\ldots,d}\left(y_{v_j}\Theta_{v_i,v_j}\right)\right\},
\end{aligned}
\tag{A.3}
$$

where the maximum over the empty set is defined as zero. The identity $\Theta_{v_i,v_j} = \prod_{e\in[v_i \rightsquigarrow v_j]} M_e$ yields the desired expression of ℓ, which is totally determined by the distributions of the increments M_e with $e \in E$.

Assume $\mathbb{E}(M_{ab}) = 1$ for all $(a,b) \in E$. Then $\mathbb{E}(\Theta_{u,v}) = \prod_{(a,b)\in[u\rightsquigarrow v]} \mathbb{E}(M_{ab}) = 1$ for $u, v \in V$ by the independence of M_e for $e \in E$. Thus it follows from Corollary 3 in Segers (2020) that the root-change formula

$$\mathbb{E}\left\{g(\boldsymbol{\Theta}_u)\right\} = \frac{\mathbb{E}\left\{g\left(\boldsymbol{\Theta}_v/\Theta_{v,u}\right)\Theta_{v,u}\right\}}{\mathbb{E}(\Theta_{v,u})}$$

holds for any Borel measurable function $g : [0,\infty)^d \setminus \{\mathbf{0}\} \to [0,\infty]$, where $\mathbf{0}$ denotes the d-dimensional vector whose components are zero. For fixed $\boldsymbol{y} = (y_v, v \in V)$, taking $g_i(\boldsymbol{x}) = \max_{j=i,\ldots,d}\left(y_{v_j} x_{v_j}\right)$ for $i = 2,\ldots,d$ and using the root-change formula we have

$$\begin{aligned}
\mathbb{E}\left\{\max_{j=i,\ldots,d}\left(y_{v_j}\Theta_{v_i,v_j}\right)\right\} &= \mathbb{E}\left[\max_{j=i,\ldots,d}\left\{y_{v_j}\Theta_{v_i,v_j}\,\mathbb{1}(\Theta_{v_i,v_{i-1}} > 0)\right\}\right] \\
&= \mathbb{E}\left[\Theta_{v_i,v_{i-1}}\max_{j=i,\ldots,d}\left\{\left(y_{v_j}\Theta_{v_i,v_j}\right)/\Theta_{v_i,v_{i-1}}\right\}\right] \\
&= \mathbb{E}\left\{\max_{j=i,\ldots,d}\left(y_{v_j}\Theta_{v_{i-1},v_j}\right)\right\}\mathbb{E}\left(\Theta_{v_i,v_{i-1}}\right) \\
&= \mathbb{E}\left\{\max_{j=i,\ldots,d}\left(\ y_{v_j}\Theta_{v_{i-1},v_j}\right)\right\},
\end{aligned}$$

where the first step holds since $\mathbb{E}(\Theta_{v_{i-1},v_i}) = 1$ implies $\mathbb{P}(\Theta_{v_i,v_{i-1}} > 0) = 1$ by Corollary 3 in Segers (2020). Consequently, in view of (A.3) and the above equality we obtain

$$\begin{aligned}
\ell(\boldsymbol{y}) &= \sum_{i=1}^{d}\mathbb{E}\left\{\max_{j=i,\ldots,d}\left(y_{v_j}\Theta_{v_i,v_j}\right)\right\} - \sum_{i=2}^{d}\mathbb{E}\left\{\max_{j=i,\ldots,d}\left(y_{v_j}\Theta_{v_{i-1},v_j}\right)\right\} \\
&= \mathbb{E}\left\{\max_{j=1,\ldots,d}\left(y_{v_j}\Theta_{v_1,v_j}\right)\right\}.
\end{aligned}$$

The final identity holds because $v_1,\ldots,v_d$ was an arbitrary permutation of $1,\ldots,d$. □

Proof of Proposition 2.4. For $y_a, y_b \geq 0$, setting $d = 2$ in (2.9) and using the relation $\max(x,y) = x + y - \min(x,y)$ yields

$$\begin{aligned}
\ell_{ab}(y_a, y_b) &= \mathbb{E}\left\{\max\left(y_a, y_b\prod_{e\in[a\rightsquigarrow b]} M_e\right) - y_b\prod_{e\in[a\rightsquigarrow b]} M_e\right\} + y_b \\
&= y_a + y_b - \mathbb{E}\left\{\min\left(y_a, y_b\prod_{e\in[a\rightsquigarrow b]} M_e\right)\right\}.
\end{aligned}$$

The desired result for λ_{ab} follows since $\lambda_{ab} = 2 - \ell_{ab}(1,1)$. The proof is complete. □

Proof of Proposition 2.5. Let $\boldsymbol{Y} = (Y_v, v \in V)$ be the Markov tree on $\mathcal{T}$ associated to H in Definition 2.2. For each pair $(a,b) \in E$, the tail dependence coefficient of (Y_a, Y_b) equals the one of (X_a, X_b). Thus it is sufficient to derive the result in terms of $\boldsymbol{Y}$.

For $\delta > 0$, note that

$$n\,\mathbb{P}(Y_a > n\delta, Y_b > n) \leq n\,\mathbb{P}(Y_b > n) \to 1 \tag{A.4}$$

as $n \to \infty$ by (2.7). From the equality

$$\int_0^x \mathbb{P}(X > u)\,\mathrm{d}\,u = \mathbb{E}\{\min(X, x)\}$$

for a non-negative random variable X and for $x > 0$, we find by (2.7), (2.8) and the fact that $1/W$ is uniformly distributed on $[0,1]$ that

$$\begin{aligned} n\,\mathbb{P}(Y_a > n\delta, Y_b > n) &= n\,\mathbb{P}(Y_a > n\delta)\cdot\mathbb{P}(Y_b > n \mid Y_a > n\delta) \\ &\to \delta^{-1}\,\mathbb{P}(W M_{ab} > 1/\delta) = \delta^{-1}\,\mathbb{P}(\delta M_{ab} > 1/W) \\ &= \delta^{-1}\,\mathbb{E}\{\min(\delta M_{ab}, 1)\} = \mathbb{E}\{\min(M_{ab}, 1/\delta)\} \end{aligned}$$

as $n \to \infty$. Combining the above display with (A.4) and letting $\delta \to 0$ we obtain

$$\mathbb{E}(M_{ab}) \le 1, \qquad (a,b) \in E. \tag{A.5}$$

The concavity of the function $y \mapsto \min(1, xy)$ implies that

$$\min(1, xy) \le \min(1, x) - (1-y)\,x\,\mathbb{1}(x < 1), \qquad (x,y) \in [0,\infty)^2, \tag{A.6}$$

as can be confirmed by a case-by-case analysis. Let u be a node on the path from a to b and write $A = \prod_{e \in [a \rightsquigarrow u]} M_e$ and $B = \prod_{e \in [u \rightsquigarrow b]} M_e$. We have $\mathbb{E}[B] \le 1$ by (A.5) and the independence of the increments M_e for $e \in E$. Using Proposition 2.4, we have, since A and B are independent and non-negative,

$$\begin{aligned} \lambda_{ab} = \mathbb{E}\{\min(1, AB)\} &\le \mathbb{E}\{\min(1, A)\} - \{1 - \mathbb{E}(B)\}\,\mathbb{E}\{A\,\mathbb{1}(A < 1)\} \\ &\le \mathbb{E}\{\min(1, A)\} = \lambda_{au}. \end{aligned} \tag{A.7}$$

Interchanging the roles of A and B yields $\lambda_{ab} \le \lambda_{ub}$. The inequality (A.6) is strict whenever $x > 1 > xy$ or $x < 1 < xy$. If $\mathbb{P}(1 - \varepsilon < M_e < 1) > 0$ and $\mathbb{P}(1 < M_e < 1 + \varepsilon) > 0$ for every $e \in E$ and every $\varepsilon > 0$, then also $\mathbb{P}(A > 1 > AB) > 0$ and $\mathbb{P}(A < 1 < AB) > 0$. But then $\min(1, AB) < \min(1, A) - (1 - B)\,A\,\mathbb{1}(A < 1)$ with positive probability, yielding a strict inequality in (A.7).

As for the lower bound, the inequality $\min(1, s)\min(1, t) \le \min(1, st)$ for non-negative reals s and t and the independence of M_e, $e \in E$, yields, with A and B as above,

$$\lambda_{ab} = \mathbb{E}\{\min(1, AB)\} \ge \mathbb{E}\{\min(1, A)\min(1, B)\} = \mathbb{E}\{\min(1, A)\}\,\mathbb{E}\{\min(1, B)\} = \lambda_{au}\lambda_{ub},$$

as required. The proof is complete. □

Derivation of Eq. (3.7). For $x_v \in \mathbb{R}$ such that $v \in V$, we have, by the maximum-minimums identity

$$\max_{v \in V} x_v = \sum_{\varnothing \ne U \subseteq V} (-1)^{|U|+1} \min_{v \in U} x_v.$$

Recall that $V_i = \{i, \dots, d\}$ and $e = (e_p, e_s)$ for $e \in E$. Without loss of generality, let $v_i = i$ for the permutation $v_1, \dots, v_d$ of $1, \dots, d$ in Corollary 3.2 and (2.9). Then, in view of the distribution of M_e for $e \in E$ given in Eq. (3.6) and by applying the above display, we have from Corollary 3.2 that

$$\begin{aligned} \ell^M(\boldsymbol{y}) &= \sum_{i=1}^{d} \mathbb{E}\left\{ \max_{j=i,\dots,d} \left(y_j \prod_{e \in [i \rightsquigarrow j]} M_e \right) \right\} - \sum_{i=2}^{d} \mathbb{E}\left\{ \max_{j=i,\dots,d} \left(y_j \prod_{e \in [(i-1) \rightsquigarrow j]} M_e \right) \right\} \\ &= \sum_{i=1}^{d} \sum_{\varnothing \ne U \subseteq V_i} (-1)^{|U|+1}\,\mathbb{E}\left\{ \min_{j \in U} \left(y_j \prod_{e \in [i \rightsquigarrow j]} M_e \right) \right\} \\ &\quad - \sum_{i=2}^{d} \sum_{\varnothing \ne U \subseteq V_i} (-1)^{|U|+1}\,\mathbb{E}\left\{ \min_{j \in U} \left(y_j \prod_{e \in [(i-1) \rightsquigarrow j]} M_e \right) \right\} \end{aligned}$$

$$= \sum_{i=1}^{d} \psi_i^{-1} \sum_{\varnothing \neq U \subseteq V_i} (-1)^{|U|+1} \left(\prod_{e \in \bigcup_{j \in U} [i \rightsquigarrow j]} \psi_{e_p} \right) \left\{ \min_{j \in U} (y_j \psi_j) \right\}$$

$$- \sum_{i=2}^{d} \psi_{i-1}^{-1} \sum_{\varnothing \neq U \subseteq V_i} (-1)^{|U|+1} \left(\prod_{e \in \bigcup_{j \in U} [(i-1) \rightsquigarrow j]} \psi_{e_p} \right) \left\{ \min_{j \in U} (y_j \psi_j) \right\}. \qquad \square$$

Proof of Proposition 4.1. Let $\mathcal{T}' = (V, E')$ be any tree with the same node set V as the true tree $\mathcal{T} = (V, E)$. As in the proof of Proposition 5 in Engelke & Volgushev (2020), an application of Hall's marriage theorem (Hall, 1935) yields a bijection $f : E \to E'$ with the property that every edge $(u, v) \in E$ is mapped to an edge $f((u, v)) = (u', v') \in E'$ such that (u, v) belongs to the path $[u' \rightsquigarrow v']$ in $\mathcal{T}$. By Proposition 2.5, $\lambda_{uv} \geq \lambda_{f(u,v)}$ and thus

$$S_{\mathcal{T}_\lambda} = \sum_{(u,v) \in E} \lambda_{uv} \geq \sum_{(u,v) \in E} \lambda_{f(u,v)} = \sum_{(u',v') \in E'} \lambda_{u'v'} = S_{\mathcal{T}'_\lambda}.$$

It follows that $\mathcal{T}$ belongs to the collection $\boldsymbol{T}_\lambda^\star$ of maximum tail dependence trees.

Suppose in addition that $\lambda_{ab} < \min(\lambda_{au}, \lambda_{ub})$ for any triple of distinct nodes a, u, b such that u lies on the path between a and b. If $\mathcal{T}' = (V, E')$ is different from $\mathcal{T}$, then there is at least one edge $(u, v) \in E$ such that $f((u, v)) = (u', v')$ is different from (u, v). By the additional condition, we then have $\lambda_{uv} > \lambda_{u'v'}$. The inequality in the above display is thus strict, implying that $\mathcal{T}$ is the unique maximizer of the sum of edge weights. $\square$

Proof of Proposition 4.2. Let $\mathcal{T}^\star = (V, E^\star)$ be an arbitrary tree in $\boldsymbol{T}_\lambda^\star$. We show that for any tree $\mathcal{T}' = (V, E')$ *not* in $\boldsymbol{T}_\lambda^\star$, we have

$$\mathbb{P}\left(\hat{S}_{\mathcal{T}'_\lambda} < \hat{S}_{\mathcal{T}_\lambda^\star} \right) \to 1, \quad n \to \infty. \tag{A.8}$$

By definition of $\hat{\boldsymbol{T}}_\lambda^\star$, this implies

$$\mathbb{P}\left(\mathcal{T}' \in \hat{\boldsymbol{T}}_\lambda^\star \right) \to 0, \quad n \to \infty.$$

As this is true for any $\mathcal{T}'$ not in $\boldsymbol{T}_\lambda^\star$, we can then conclude that $\mathbb{P}\left(\hat{\boldsymbol{T}}_\lambda^\star \subseteq \boldsymbol{T}_\lambda^\star \right) \to 1$ as $n \to \infty$.

To show (A.8), note that, by definition of $\boldsymbol{T}_\lambda^\star$, we have

$$S_{\mathcal{T}'_\lambda} < S_{\mathcal{T}_\lambda^\star}.$$

The consistency of the empirical tail dependence coefficients $\hat{\lambda}_e$ for $e \in E$ (Drees & Huang, 1998; Schmidt & Stadtmüller, 2006) implies

$$\hat{S}_{\mathcal{T}'_\lambda} - \hat{S}_{\mathcal{T}_\lambda^\star} = \sum_{e \in E'} \hat{\lambda}_e - \sum_{e \in E^\star} \hat{\lambda}_e = \sum_{e \in E'} \lambda_e - \sum_{e \in E^\star} \lambda_e + o_p(1) = S_{\mathcal{T}'_\lambda} - S_{\mathcal{T}_\lambda^\star} + o_p(1), \quad n \to \infty,$$

yielding the convergence in (A.8), as required.

If $\boldsymbol{T}_\lambda^\star$ is a singleton, say $\{\mathcal{T}_\lambda^\star\}$, then the relation $\hat{\boldsymbol{T}}_\lambda^\star \subseteq \boldsymbol{T}_\lambda^\star$ implies that $\hat{\boldsymbol{T}}_\lambda^\star$ (which is non-empty since it is the set of maximizers over a finite domain) is a singleton too and that its only element, say $\hat{\mathcal{T}}_\lambda^\star$, is equal to $\mathcal{T}_\lambda^\star$. $\square$

Proof of Theorem 4.4. By Theorem 5.1 in Bücher et al. (2014), we have weak convergence of $\alpha_n = \sqrt{k}(\hat{\ell}_{n,k} - \ell)$ in the so-called hypi-topology on the set of bounded functions on $[0, 1]^d$ to a process which, thanks to the condition on the μ_r-almost everywhere continuity of $\dot{\ell}_j$, is μ_r-almost everywhere equal to α in (4.4). By Corollary 3.3 in the same article and again by the assumption on $\dot{\ell}_j$, we obtain the weak convergence of α_n to α in $L^2(\mu_r)$; indeed, with probability one, the process α is continuous μ_r-almost everywhere, thanks to the assumption

on $\dot{\ell}_j$ and the fact that $W(\mathbf{0}) = 0$. The linear functional $L^2(\mu_r) \to \mathbb{R} : f \mapsto \int f\psi_r \,\mathrm{d}\,\mu_r$ for $r \in \{1,\dots,q\}$ being continuous, an application of the continuous mapping theorem yields the result. The limit in (4.5) is a linear transformation of a Gaussian process and therefore Gaussian as well. The formula for the covariance matrix follows by Fubini's theorem.

An alternative proof is to proceed by a Skorohod construction as in the proof of Proposition 7.3 in Einmahl et al. (2012). The term involving the partial derivatives $\dot{\ell}_j$ is handled via the dominated convergence theorem: the pointwise convergence for μ_r-almost every $\boldsymbol{x} \in [0,1]^d$ holds true thanks to the condition on the partial derivatives $\dot{\ell}_j$. □

Proof of Corollary 4.7. For $e = (a,b) \in E$, the map

$$T_e : [0,1]^2 \to [0,1]^d : (x_a, x_b) \mapsto x_a \boldsymbol{e}_a + x_b \boldsymbol{e}_b$$

pushes the measure ν_e on $[0,1]^2$ forward to the measure $\mu_e = T_e \# \nu_e$ on $[0,1]^d$ given by

$$\mu_e(A) = \nu_e\left(\left\{(x_a, x_b) \in [0,1]^2 : x_a \boldsymbol{e}_a + x_b \boldsymbol{e}_b \in A\right\}\right)$$

for Borel sets $A \subseteq [0,1]^d$. The support of μ_e is contained in $\{\boldsymbol{x} \in [0,1]^d : x_j = 0 \text{ if } j \neq a \text{ and } j \neq b\}$. By the change-of-variables formula, integrals with respect to μ_e and ν_e are related through $\int_{[0,1]^d} f(\boldsymbol{x})\,\mu_e(\mathrm{d}\,\boldsymbol{x}) = \int_{[0,1]^2} f(T_e(x_a,x_b))\,\nu_e(\mathrm{d}(x_a,x_b))$.

Put $\alpha_n = \sqrt{k}(\hat{\ell}_{n,k} - \ell)$. By Slutsky's lemma, the $o_p(\mathbf{1})$ term in Assumption 4.6 does not matter for the convergence in distribution. Since $\hat{\ell}_{n,k,e}(x_a,x_b) = \hat{\ell}_{n,k}(T_e(x_a,x_b))$ and similarly for ℓ_e and ℓ, the main term on the right-hand side of (4.6) can be written as

$$\int_{[0,1]^d} \alpha_n(\boldsymbol{x})\,\boldsymbol{\psi}_e(x_a,x_b)\,\mu_e(\mathrm{d}\,\boldsymbol{x}).$$

This is a random vector of dimension p_e, each component of which has the same form as the entries of the random vector on the left-hand side of (4.5). Indeed, the measure μ_r in (4.5) is equal to the measure μ_e here, while the function ψ_r in (4.5) is the equal to the function $\boldsymbol{x} \mapsto \psi_{e,j}(x_a,x_b)$ here, where $\psi_{e,j}$ is one of the p_e component functions of $\boldsymbol{\psi}_e$. The conclusion now follows from Theorem 4.4; Assumption 4.6(ii) ensures that the smoothness condition in Theorem 4.4 is fulfilled. □

A.2 Additional material and results for simulation study in Section 5

In this subsection, we list some additional material for the simulation study: Prim's algorithm in Algorithm 1; the simulation and fitting procedure for Hüsler–Reiss distributions in Algorithm 2; the variogram matrices $\boldsymbol{\Gamma}_3$ and $\boldsymbol{\Gamma}_4$ in Table 4; the tree structure in Fig. 5; the simulation results for the proportion of misidentified trees in 300 replications for the 10-dimensional Hüsler–Reiss distribution in Fig. 6; and the boxplots of $\frac{\|\hat{\boldsymbol{\Gamma}}_M - \boldsymbol{\Gamma}\|_F}{\|\boldsymbol{\Gamma}\|_F}$, $\hat{D}_{\mathcal{T}_\omega}$ and the estimation error of rare event probability in Fig. 7–8.

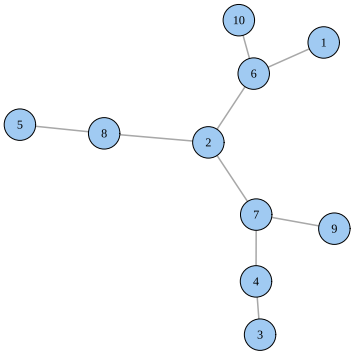

Fig. 5 Tree structure for 10-dimensional Hüsler–Reiss distribution with variogram matrix $\mathbf{\Gamma}_3$ in Table 4 used for generation of samples in Subsection 5.1

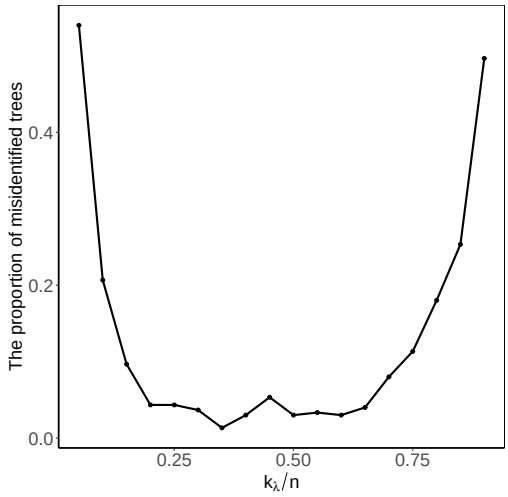

Fig. 6 Proportion of misidentified trees in 300 replications based on samples from 10-dimensional Hüsler–Reiss distribution with variogram matrix $\mathbf{\Gamma}_3$ in Table 4 and dependence structure linked to the tree in Fig. 5

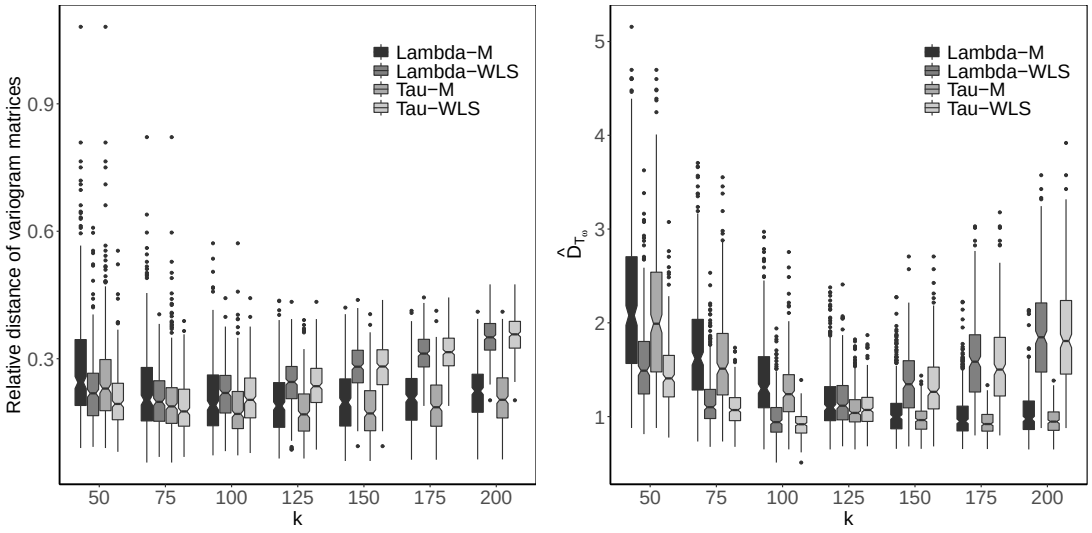

Fig. 7 Boxplots of $\frac{\|\hat{\mathbf{\Gamma}}_M - \mathbf{\Gamma}\|_F}{\|\mathbf{\Gamma}\|_F}$ (left) and $\hat{D}_{\mathcal{T}_\omega}$ in (4.8) (right) for 10-dimensional Hüsler–Reiss distribution with variogram matrix $\mathbf{\Gamma}_4$ based on $\hat{\mathcal{T}}_\lambda^\star$ (Lambda) and $\hat{\mathcal{T}}_\tau^\star$ (Tau), where bivariate stable tail dependence functions are fitted based on M-estimator (M) and weighted least squares estimator (WLS) at $k = 50, 75, \ldots, 200$ (300 samples of size $n = 1000$)

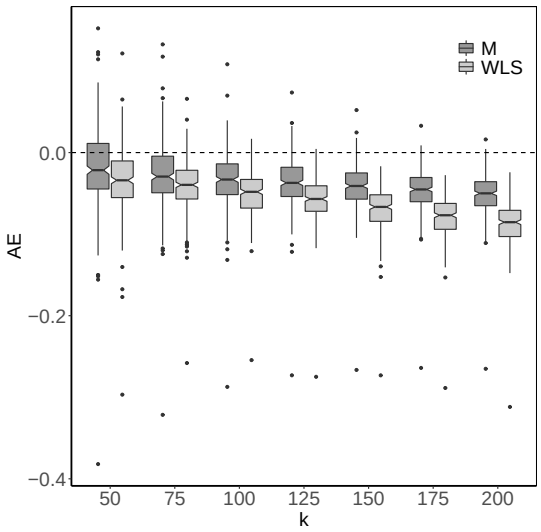

Fig. 8 Boxplot of approximation errors of rare event probability in (5.3) based on $\hat{\mathcal{T}}_\tau^\star$ for 10-dimensional Hüsler–Reiss distribution with variogram matrix $\mathbf{\Gamma}_4$, where bivariate stable tail dependence functions are fitted with M-estimator (M) and weighted least squares estimator (WLS) at $k = 50, 75, \ldots, 200$; levels u_1, u_2, u_3 are taken as 0.999 empirical marginal quantiles and $u_j = \infty$ for $j = 4, \ldots, d$ (300 samples of size $n = 1000$)

Algorithm 1: Prim's algorithm

1. Choose a starting point $v \in V$. Define the node set and edge set of the maximum dependence tree as $V' = \{v\}$ and $E' = \varnothing$.
2. While $V' \neq V$, choose an edge $e = (a, b)$ with maximal weight $\hat{\omega}_{ab}$ such that $a \in V'$ and $b \notin V'$.
3. Set $V' = V' \cup \{b\}$ and $E' = E' \cup \{(a, b)\}$.
4. End when $V' = V$.

Algorithm 2: Simulation and fitting procedure for Hüsler–Reiss distributions in Section 5

1. (a) If a sample linked to a real tree structure is wanted, a random tree structure with d nodes is generated primarily. Then we generate $d-1$ random parameters $\gamma_{i,j}$ from the interval $[0.1, 1]$ for each adjacent pairs on the tree and complete the variogram matrix according to the tree structure using the function `complete_Gamma` in package `graphicalExtremes`;

   (b) Otherwise, we obtain a conditionally negative definite variogram matrix $\mathbf{\Gamma}$ by transforming a $(d-1) \times (d-1)$ dimensional positive definiteness covariance matrix $\mathbf{\Sigma}$ with random eigenvalues in the interval $[0.6, 3]$. The matrix $\mathbf{\Sigma}$ is generated using the command `genPositiveDefMat` in package `clusterGeneration` and the transformation is achieved using the command `Sigma2Gamma` in package `graphicalExtremes`;
2. Generate a random sample from the Hüsler–Reiss distribution with variogram matrix $\mathbf{\Gamma}$ using the command `rmstable` or `rmstable_tree` (for tree structure) in package `graphicalExtremes` and add the independent noise generated from Fréchet(2) to the sample;
3. Calculate the empirical bivariate upper tail dependence coefficient matrix and Kendall's τ using the commands `emp_chi_mat` in package `graphicalExtremes` and `corkendall` in package `copula` respectively;
4. Select maximum spanning tree based on empirical estimates of λ and τ using Prim's algorithm in Algorithm 1;
5. Estimate parameters of stable tail dependence functions of adjacent pairs on selected tree by the moments estimator in Einmahl et al. (2008), M-estimator in Einmahl et al. (2012) and weighted least squares estimator in Einmahl et al. (2018), where the weighted functions for M-estimator are taken as $g_1(x, y) = 1$ and $g_2(x, y) = x$ and the weighted functions for the other two estimators are taken as one in the simulation study. The locations used for the weighted least squares estimator are $(1, 1)$, $(2, 1)$ and $(0.5, 1.5)$ throughout the simulation study;
6. Calculate $\hat{D}_{\mathcal{T}_\omega}$ in (4.8) to measure the goodness-of-fit of the constructed model;
7. Repeat Steps 2–6 for 300 times.

Table 4: Variogram matrices $\boldsymbol{\Gamma}_3$ and $\boldsymbol{\Gamma}_4$ for 10-dimensional Hüsler–Reiss distributions used for generation of samples in simulation study in Subsection 5.1, where the former has dependence structure linked to the tree in Fig. 5

$$\boldsymbol{\Gamma}_3 = \begin{pmatrix} 0 & 1.499 & 3.563 & 3.258 & 2.168 & 0.500 & 2.395 & 1.814 & 2.852 & 1.246 \\ & 0 & 2.064 & 1.759 & 0.669 & 0.999 & 0.896 & 0.315 & 1.353 & 1.745 \\ & & 0 & 0.305 & 2.733 & 3.063 & 1.168 & 2.379 & 1.624 & 3.809 \\ & & & 0 & 2.428 & 2.758 & 0.863 & 2.074 & 1.319 & 3.504 \\ & & & & 0 & 1.668 & 1.565 & 0.354 & 2.022 & 2.413 \\ & & & & & 0 & 1.895 & 1.313 & 2.352 & 0.746 \\ & & & & & & 0 & 1.211 & 0.456 & 2.641 \\ & & & & & & & 0 & 1.667 & 2.059 \\ & & & & & & & & 0 & 3.097 \\ & & & & & & & & & 0 \end{pmatrix}$$

$$\boldsymbol{\Gamma}_4 = \begin{pmatrix} 0 & 1.154 & 1.352 & 0.981 & 1.415 & 1.044 & 0.773 & 0.877 & 0.860 & 1.373 \\ & 0 & 2.797 & 2.060 & 2.789 & 2.092 & 1.684 & 1.901 & 1.762 & 2.754 \\ & & 0 & 2.366 & 2.935 & 2.473 & 2.117 & 2.245 & 2.220 & 2.908 \\ & & & 0 & 2.422 & 1.989 & 1.699 & 1.826 & 1.782 & 2.379 \\ & & & & 0 & 2.518 & 2.184 & 2.305 & 2.283 & 2.923 \\ & & & & & 0 & 1.720 & 1.866 & 1.801 & 2.476 \\ & & & & & & 0 & 1.581 & 1.513 & 2.135 \\ & & & & & & & 0 & 1.660 & 2.260 \\ & & & & & & & & 0 & 2.235 \\ & & & & & & & & & 0 \end{pmatrix}$$

A.3 Additional information and analysis for Danube discharge data in Section 6

In this subsection, we list additional material related to the data analysis of daily discharges of rivers in the upper Danube basin: the information for the original discharge data in Table 5; the number of declustered excesses over the 0.9 sample quantile of average daily discharges at 31 gauging stations in Table 6; the tree-like river network in the upper Danube basin in Fig. 9; the maximum dependence trees $\hat{\mathcal{T}}_\lambda^\star$ and $\hat{\mathcal{T}}_\tau^\star$ selected based on 51 years of discharge data in Fig. 10; and the parameter estimates of the generalized Pareto distribution fitted to excesses over the 0.9 marginal empirical quantile for each gauging station in Table 7.

Table 5: The start and end recording dates and sample size of original data for Summer (June, July and August) discharges at 31 gauging stations in the upper Danube basin

Station	Start	End	Size	Station	Start	End	Size
1	1901-06-01	2010-08-31	10120	17	1959-06-01	2013-08-31	5060
2	1901-06-01	2013-08-31	10396	18	1959-06-01	2013-08-31	5060
3	1926-06-01	2013-08-31	8096	19	1950-06-01	2013-08-31	5888
4	1924-06-01	2013-08-31	8280	20	1960-06-01	2013-08-31	4968
5	1926-06-01	2013-08-31	8096	21	1901-06-01	2013-08-31	10396
6	1901-06-01	2013-08-31	10396	22	1951-06-01	2013-08-31	5796
7	1924-06-01	2013-08-31	8280	23	1921-06-01	2013-08-31	8556
8	1924-06-01	2013-08-31	8280	24	1930-06-01	2013-08-31	7728
9	1924-06-01	2013-08-31	8280	25	1901-06-01	2013-08-31	10396
10	1954-06-01	2013-08-31	5520	26	1959-06-01	2013-08-31	5060
11	1901-06-01	2013-08-31	10396	27	1931-06-01	2013-08-31	7636
12	1901-06-01	2013-08-31	10396	28	1951-06-01	2013-08-31	5796
13	1921-06-01	2013-08-31	8556	29	1901-06-01	2013-08-31	10396
14	1926-06-01	2013-08-31	8096	30	1901-06-01	2013-08-31	10396
15	1959-06-01	2013-08-31	5060	31	1957-06-01	2013-08-31	5244
16	1959-06-01	2013-08-31	5060	–	–	–	–

Table 6: The number of declustered excesses over the 0.9 sample quantile of average daily discharges at 31 gauging stations in the upper Danube basin

Station	1	2	3	4	5	6	7	8	9	10	11
Excesses	226	192	166	179	185	248	198	204	216	149	347
Station	12	13	14	15	16	17	18	19	20	21	22
Excesses	348	229	159	102	125	126	116	140	125	282	136
Station	23	24	25	26	27	28	29	30	31	––	––
Excesses	153	143	262	126	208	177	326	322	155	––	––

Table 7: Parameter estimates of the generalized Pareto distribution fitted to excesses over the empirical quantile $q_{j,p}$ (rounded to the nearest integer) of daily discharges at gauging station $j = 1, \ldots, 31$ in the upper Danube basin

Station	$q_{j,p}$	$\hat{\sigma}_j$	$\hat{\vartheta}_j$	Station	$q_{j,p}$	$\hat{\sigma}_j$	$\hat{\vartheta}_j$
1	2680	749.596	0.029	17	106	70.874	0.079
2	1040	387.594	0.017	18	75	72.553	0.040
3	715	278.147	0.074	19	57	54.261	0.161
4	715	289.252	0.047	20	236	142.385	0.098
5	598	232.214	0.057	21	193	96.663	0.036
6	577	225.447	0.027	22	173	91.849	0.016
7	553	252.046	−0.025	23	58	40.197	0.202
8	309	136.322	0.046	24	44	33.902	0.199
9	270	136.070	0.054	25	51	41.917	0.243
10	212	120.944	−0.016	26	45	29.339	0.413
11	114	77.116	0.030	27	38	24.371	0.409
12	54	35.780	0.032	28	94	65.412	0.249
13	1630	579.764	0.123	29	90	66.082	0.139
14	315	120.868	0.124	30	578	275.008	0.116
15	296	123.116	0.089	31	544	256.610	0.205
16	162	103.409	0.092	−−	−−	−−	−−

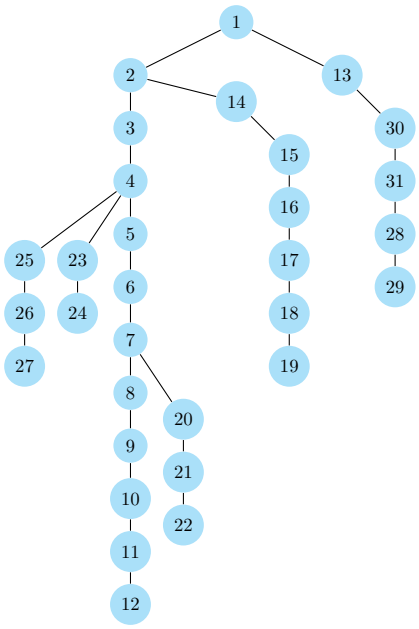

Fig. 9 Tree-like structure of the river network in the upper Danube basin. Each node represents a numbered gauging station along the rivers and water flows toward gauging station 1.

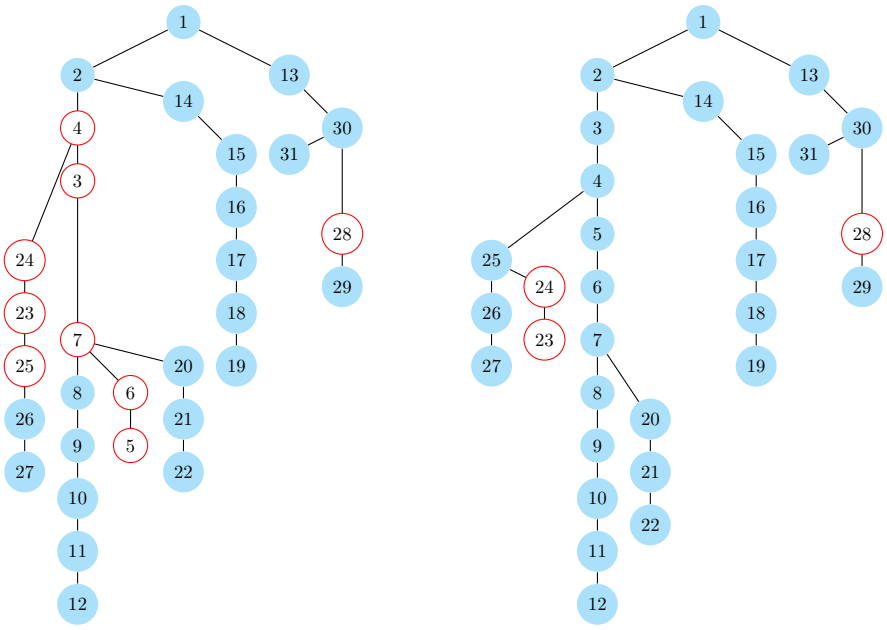

Fig. 10 Maximum dependence trees $\hat{\mathcal{T}}_\lambda^\star$ (left) and $\hat{\mathcal{T}}_\tau^\star$ (right) in (4.3) selected based on the declustered excesses of 51 years of average daily discharges at 31 gauging stations in the upper Danube basin. The empty nodes are those which have different neighbours from the tree structure of the river network in Fig. 9.