INFERENCE ON EXTREMAL DEPENDENCE IN A LATENT MARKOV TREE MODEL ATTRACTED TO A HÜSLER-REISS DISTRIBUTION

Asenova, S., Mazo, G. and J. Segers

DISCUSSION PAPER | 2020 / 05

UCLouvain

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

Inference on extremal dependence in a latent Markov tree model attracted to a Hüsler–Reiss distribution

Stefka Asenova* Gildas Mazo† Johan Segers‡

January 28, 2020

Abstract

A Markov tree is a probabilistic graphical model for a random vector by which conditional independence relations between variables are encoded via an undirected tree and each node corresponds to a variable. One possible max-stable attractor for such a model is a Hüsler–Reiss extreme value distribution whose variogram matrix inherits its structure from the tree, each edge contributing one free dependence parameter. Even if some of the variables are latent, as can occur on junctions or splits in a river network, the underlying model parameters are still identifiable if and only if every node corresponding to a missing variable has degree at least three. Three estimation procedures, based on the method of moments, maximum composite likelihood, and pairwise extremal coefficients, are proposed for usage on multivariate peaks over thresholds data. The model and the methods are illustrated on a dataset of high water levels at several locations on the Seine network. The structured Hüsler–Reiss distribution is found to fit the observed extremal dependence well, and the fitted model confirms the importance of flow-connectedness for the strength of dependence between high water levels, even for locations at large distance apart.

1 Introduction

A major topic in multivariate extreme value theory is the modelling of tail dependence between a finite number of variables. Informally, tail dependence represents the degree of association between the extreme values of these variables. Probabilistic graphical models are distributions which embody a set of conditional independence relations and have a graph-based representation, according to which the nodes of the graph are associated to the variables and the set of edges encode the conditional independence relations. The intersection of the two fields, extreme value theory and probabilistic graphical models, gives rise to the study of the tail behaviour of graphical models.

Consider a river network where the interest is in extreme water levels or water flow in relation to flood risks. Fig. 1 illustrates part of the Seine network. The graph fixed by the seven labelled nodes and the river channels between them is a base for building a conditional independence model for studying the joint extremal behaviour of water levels at these sites. Graphical models on multivariate extremes have been applied to hydrological data in the study of water flows of the Bavarian Danube in Engelke and Hitz (2018), in the application of water flow of Fraser river, British Colombia, in Lee and Joe (2018) and on precipitation data

*Corresponding author. UCLouvain, LIDAM/ISBA, Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgium. E-mail: stefka.asenova@uclouvain.be

†MaIAGE, INRA, Université Paris-Saclay 78350, Jouy-en-Josas, France. E-mail: gildas.mazo@inra.fr

‡UCLouvain, LIDAM/ISBA, Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgium. E-mail: johan.segers@uclouvain.be

in the Japanese archipelago in Yu et al. (2017), where the model is based on a spatial grid viewed as an ensemble of trees. Other applications are presented in Einmahl et al. (2018) which study the tail dependence of the European stock market as a max-linear model on a directed acyclic graph (DAG), in Klüppelberg and Sönmez (2018) introducing an infinite random max-linear model to analyze the distribution of extreme opinions in a social network, and a second application from Lee and Joe (2018) where a factor model with one latent variable is designed to study the extreme stock returns of international companies in the pharmaceutical sector.

Relatively recently the relation between extreme value distributions and conditional independence assumptions has been given theoretical relevance, the earliest with the paper of Gissibl and Klüppelberg (2018) introducing the max-linear model as a structural equation model on a DAG, the regularly varying Markov tree in Segers (2019) and the extremal graphical model in Engelke and Hitz (2018) based on multivariate Pareto distributions. Earlier, Papastathopoulos and Strokorb (2016) showed that for a max-stable random vector with positive and continuous density, conditional independence implies unconditional independence, thereby concluding that a broad class of max-stable distributions does not exhibit an interesting Markov structure.

In this paper we build on Segers (2019) by studying the domain of attraction of the Hüsler–Reiss extreme-value attractor of a special class of parametric graphical models on a tree. The model has some convenient properties thanks to its relation to the Gaussian distribution. The Markov structure induces a certain structure on the variogram matrix of the distribution, reducing the number of free parameters.

In the context of river network applications, variables associated to junctions/splits can be latent, as is the case for instance for nodes 2 and 5 on the Seine network in Fig. 1. The subvector of observable variables then no longer satisfies the Markov property with respect to the tree. A particularly interesting characteristic of the structured Hüsler–Reiss distribution is the existence of a simple parameter identifiability condition in case of latent variables in the tree: as long as all nodes corresponding to latent variables have degree at least three, the parameters are still identifiable. This result has the important implication that the extremal dependence between the variables can still be studied in the way it naturally exists thanks to the original network, and hence the conditional independence assumptions implied by it are preserved during statistical modeling.

In this paper $\xi = (\xi_v)_{v \in V}$ is a random vector with continuous margins. The elements of ξ live on the node set V of a network in the form of a undirected tree with edges E. Upon marginal standardization, we assume that ξ is in the domain of attraction of a Hüsler–Reiss distribution with limiting variogram matrix having a specific path-dependent structure according to the tree. The motivation behind the model stems from Segers (2019), where a regularly varying Markov tree, built from connecting each pair of neighboring variables with a bivariate Hüsler–Reiss copula, is in the same domain of attraction as ξ.

A Markov tree in the context of Segers (2019) is viewed as a collection of Markov chains that share common segments. A regularly varying Markov chain that starts at a high value is asymptotically distributed as a multiplicative random walk called tail chain (Smith, 1992; Perfekt, 1994; Yun, 1998; Segers, 2007; Janssen and Segers, 2014; Segers, 2019). In a similar way, the limiting distribution of a Markov tree conditionally on the event that a given variable exceeds a high threshold is called tail tree. Formally, for a vector X with Pareto margins, the tail tree is the random vector at the right-hand side of the weak limit relation

$$(X_v/X_u, v \in V \setminus u) \mid X_u > x \xrightarrow{d} (\Theta_{uv}, v \in V \setminus u), \qquad x \to \infty,$$

where each Θ_{uv} is a product of independent increments along the path from node u to node v. This multiplicative structure is confirmed by Engelke and Hitz (2018) for a multivariate

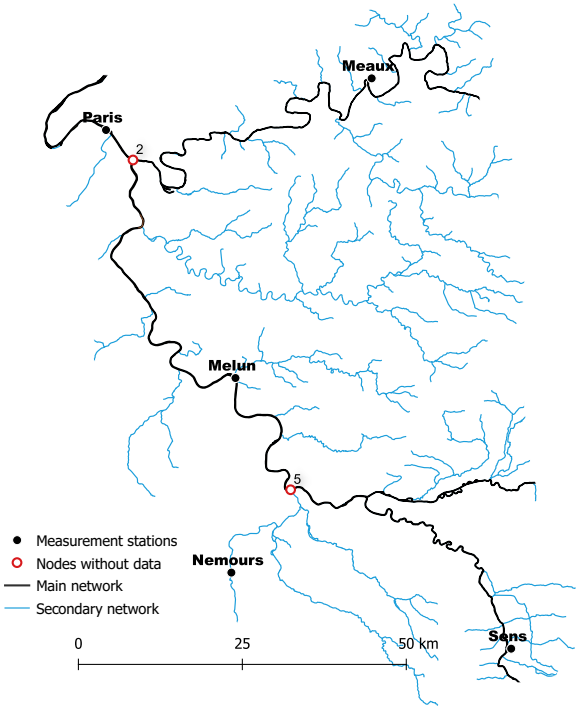


Figure 1: Seine network. The data is from the web-site of Copernicus Land Monitoring Service: https://land.copernicus.eu/imagery-in-situ.

Pareto distribution with positive and continuous density that factorizes with respect to a tree.

The Hüsler–Reiss distribution yields interesting theoretical results when applied to extremes on graphical models. Example 4 in Segers (2019) shows that having the Hüsler–Reiss bivariate copula for every pair of adjacent nodes leads to the independent increments of the tail tree being log-normal variables. Engelke and Hitz (2018) prove that the precision matrix of a Pareto vector with Hüsler–Reiss density which factorizes according to a given decomposable graph, has a sparse structure corresponding to missing links in the graph, a well known property for Gaussian graphical models. This important result is already mentioned in Example 1.14 of Hitz and Evans (2016). In that paper the predecessor of the multivariate Pareto graphical model from Engelke and Hitz (2018) is defined as a censored Pareto vector such that the range of each of the components does not depend on the other marginals.

In this paper the structure of the covariance matrices follows from the multiplicative nature of the tail tree variables: each entry of the matrix depends linearly on the shared parameters along the paths between the conditioning variable and the corresponding pair of variables. Exactly this structure of the covariance matrices leads to the identifiability condition in case of latent variables.

We propose three types of estimators of the tail dependence parameters of a distribution which, upon marginal standardization, is in the domain of attraction of the tree-structured Hüsler–Reiss distribution: a first one called method of moments estimator (MME) is based on the estimator proposed in Engelke et al. (2015), a second one is based on the composite likelihood function (Maximum Likelihood estimator or MLE) and the third one is essentially the pairwise extremal coefficient estimator (ECE) introduced in Einmahl et al. (2018). In general the first two methods produce similar results and have the lowest variance in simulation experiments. The ECE uses bivariate stable tail dependence functions which

involve univariate normal distribution functions.

An analysis of data on high water levels at several locations of the Seine network allows us to illustrate the model and the proposed estimation procedures. Along with point estimates computed using the three estimators, we construct confidence intervals for the ECE based on the asymptotic theory in Einmahl et al. (2018). From these, we conclude that avoiding the latent variable issue by suppressing nodes with latent variables and redrawing edges accordingly might lead to an incorrect picture of the extremal dependence in the original network. The goodness-of-fit of the model is confirmed by comparing non-parametric and model-based estimates of extremal coefficients and Pickands dependence functions.

Parsimoniously parametrized Hüsler–Reiss models are proposed in Lee and Joe (2018) as well. In Appendix A.1, we illustrate the structural difference of their Markov tree model with ours. In Yu et al. (2017), spatial extremes are modelled by an ensemble-of-tree model, mixing over various Markov tree models with extreme-value marginal distributions and parametric families of pairwise copulas.

The outline of the paper is as follows: Section 2 describes the model and its probabilistic properties, while in Section 3 the focus is on the identifiability criterion in case some variables are latent. Section 4 introduces the three estimators, used for statistical inference and Section 5 is dedicated to the study of high water levels on the Seine network, France. Concluding remarks and perspectives for further research are discussed in Section 6. The Appendix provides proofs that are not in the text, some simulation results which aim at comparing the different estimators, and details about some procedures and the data preprocessing.

2 The model – definition and properties

2.1 Preliminaries

We first introduce some notions and notation from graph theory. A graph is a pair $\mathcal{G} = (V, E)$ where $V = \{1, \dots, d\}$ is the set of nodes or vertices and $E \subseteq \{(a, b) \in V \times V : a \neq b\}$ is the set of edges. The number of vertices in a subset $U \subseteq V$ will be denoted by $|U|$, while d is reserved for $|V|$ only. A graph is undirected if $(a, b) \in E$ is equivalent to $(b, a) \in E$. A path $(u \rightsquigarrow v)$ between two distinct nodes u, v is a collection $\{(u_0, u_1), (u_1, u_2), \dots, (u_{n-1}, u_n)\}$ of distinct, directed edges such that $u_0 = u$ and $u_n = v$. An undirected tree is an undirected graph $\mathcal{T} = (V, E)$ such that for every distinct nodes a and b there is a unique path $(a \rightsquigarrow b)$. Rooting a tree at some arbitrary node u means to consider only those edges $E_u \subsetneq E$ that appear on paths starting at u. The rooted version of $\mathcal{T}$ at node u will be denoted by $\mathcal{T}_u = (V, E_u)$.

Let $\mathcal{T} = (V, E)$ be an undirected tree and let $X = (X_v)_{v \in V}$ be a random vector indexed by the vertices of the tree. For $A \subseteq V$, the notation $X_A = (X_a)_{a \in A}$ is the collection of variables on the node subset A. When we exclude a single node, say v, from a set of nodes U we write $U \setminus v$ instead of $U \setminus \{v\}$. For disjoint subsets A, B, C of V, the expression $A \perp\!\!\!\perp_{\mathcal{T}} B \mid C$ means that C separates A from B in $\mathcal{T}$, also called graphical separation, i.e., all paths from A to B pass through at least one vertex in C. If X lives on the probability space $(\Omega, \mathcal{B}, \mathbb{P})$, conditional independence of X_A and X_B given X_C will be denoted by $X_A \perp\!\!\!\perp_{\mathbb{P}} X_B \mid X_C$. If $P = \mathbb{P}X^{-1}$ is the distribution of X, then the tree $\mathcal{T}$ is an independence map (I-map) of P if for any disjoint subsets A, B, C of V it holds that

$$A \perp\!\!\!\perp_{\mathcal{T}} B \mid C \implies X_A \perp\!\!\!\perp_{\mathbb{P}} X_B \mid X_C \tag{1}$$

(Koller and Friedman, 2009). This assumption is equivalent to the assumption that X obeys the global Markov property with respect to $\mathcal{T}$ (Lauritzen, 1996).

A key element of our model is the multivariate Hüsler–Reiss distribution. The latter is a popular parametric family of extreme value distributions (Genton et al., 2011; Huser and Davison, 2013; Asadi et al., 2015; Engelke et al., 2015; Einmahl et al., 2018; Lee and Joe, 2018). It arises as the limiting distribution of the vector of scaled component-wise maxima of independent random samples of a multivariate normal distribution with correlation matrix ρ_n depending on the sample size n (Hüsler and Reiss, 1989). In particular, assume that

$$\lim_{n\to\infty} (1-\rho_{ij}(n)) \ln n = \lambda_{ij}^2$$

for every pair of variables i, j in $V = \{1, \ldots, d\}$. Let $\Lambda = (\lambda_{ij}^2)_{i,j\in V}$ denote this limiting matrix. For every subset $W \subseteq V$ and any element $u \in W$ let $\Gamma_{W,u}(\Lambda)$ be the square matrix of size $|W| - 1$ with elements

$$(\Gamma_{W,u}(\Lambda))_{ij} = 2(\lambda_{iu}^2 + \lambda_{ju}^2 - \lambda_{ij}^2), \qquad i, j \in W \setminus u. \tag{2}$$

Nikoloulopoulos et al. (2009) and later Genton et al. (2011) and Huser and Davison (2013) show that the cumulative distribution function as deduced by Hüsler and Reiss (1989) can be more compactly written as

$$H_\Lambda(x_1, \ldots, x_d) = \exp\left\{ -\sum_{u\in V} \frac{1}{x_u} \Phi_{d-1}\Big(\ln \frac{x_v}{x_u} + 2\lambda_{uv}^2, v \in V \setminus u; \Gamma_{V,u}(\Lambda) \Big) \right\} \tag{3}$$

for $(x_1, \ldots, x_d) \in (0, \infty)^d$, where $\Phi_p(\,\cdot\,; \Sigma)$ denotes the p-variate zero mean Gaussian cumulative distribution function with covariance matrix Σ. In (3) the marginal distributions are unit Fréchet, i.e., $x \mapsto e^{-1/x}$ for $x > 0$.

2.2 Introducing the model

Let $\xi = (\xi_v)_{v\in V}$ be a d-variate random vector with continuous margins, whose elements live on the nodes of an undirected tree $\mathcal{T} = (V, E)$, with $V = \{1, \ldots, d\}$. Think of water levels measured at the nodes of a river network. Because of measurement error or other kinds of observational noise, graphical separation need not imply conditional independence as in (1).

Let $X = (X_v)_{v\in V}$ be a vector whose elements are standard Pareto variables obtained after the transformation $X_v = 1/(1 - F_v(\xi_v))$, hence $\mathbb{P}(X_v \le x) = 1 - 1/x$ for $x \in [1, \infty)$ and $v \in V$. Note that ξ and X share the same copula. We assume further that X is in the max-domain of attraction of the Hüsler–Reiss distribution in (3) with $\Lambda = (\lambda_{ij}^2)_{i,j\in V}$ having the following structure: there exist positive scalars θ_e for $e \in E$ such that $\theta_{ab} = \theta_{ba}$ and

$$(\Lambda(\theta))_{ij} = \lambda_{ij}^2 = \frac{1}{4} \sum_{e\in(i\rightsquigarrow j)} \theta_e^2\,, \qquad i, j \in V,\ i \neq j. \tag{4}$$

The diagonal elements are zero: $(\Lambda(\theta))_{ii} = 0$ for $i \in V$. With this parametrization the extremal dependence in ξ and in X depends on a vector of $d - 1$ free parameters indexed by the edges of the tree.

We will often use the stable tail dependence function (stdf). The stdf l of a random vector with copula C is given by the limit

$$\lim_{t\to\infty} t\,[1 - C(1 - tx_1, \ldots, 1 - tx_d)] = l(x), \qquad x \in [0, \infty)^d. \tag{5}$$

For details about the stdf we refer to de Haan and Ferreira (2007, Chapter 6) and Beirlant et al. (2004, Chapter 8). In our particular case, for a subset $J \subseteq V$, the stdf l_J of the

subvector X_J is given by

$$\begin{aligned} l_J(x_J) &= -\ln H_{\Lambda(\theta)}(1/x_v, v \in J) \\ &= \sum_{u \in J} x_u \, \Phi_{|J \setminus u|}\Big(\ln \frac{x_u}{x_v} + 2\lambda_{uv}^2(\theta), v \in J \setminus u; \Gamma_{J,u}(\Lambda(\theta))\Big), \end{aligned} \tag{6}$$

where $x_J = (x_v, v \in J) \in [0, \infty)^{|J|}$; if $x_u = 0$, the corresponding term in the sum vanishes.

We summarize the assumptions as follows: ξ is a random vector with continuous margins which is associated to the nodes of an undirected tree $\mathcal{T} = (V, E)$ and its copula C satisfies the limit in (5) with $l = l_V$ the Hüsler–Reiss stdf in (6) parametrized by θ through the tree structure. The vector X is obtained from ξ by rescaling the margins to unit Pareto, so that the max-stable attractor of X is (3) with Λ as in (4).

2.3 Motivation of the structured Hüsler–Reiss model

The aim of this subsection is to provide a motivation for the structured Hüsler–Reiss model in (4). To this end, we construct a graphical model in its domain of attraction. In Segers (2019, Theorem 1) it is shown that if $Y = (Y_v)_{v \in V}$ satisfies the global Markov property (1) with respect to the undirected tree $\mathcal{T} = (V, E)$ and if Y obeys Condition 1 in the paper, then it holds that for every $u \in V$ we have convergence in distribution

$$(Y_v / Y_u \mid Y_u > x)_{v \in V \setminus u} \xrightarrow{d} (\Theta_{uv})_{v \in V \setminus u}, \qquad x \to \infty,$$

where Θ_{uv} is a product of independent increments M_e indexed by the edges $e \in E_u$:

$$\Theta_{uv} = \prod_{e \in (u \rightsquigarrow v)} M_e \, .$$

For every $u \in V$ the vector $(\Theta_{uv})_{v \in V \setminus u}$ is called a tail tree. The conditions that Y should satisfy include stability of the Markov kernels at high levels for every pair of adjacent nodes and a condition which ensures that if zero is a possible state of the tail tree it is also an absorbing state. The result is applicable to any distribution that satisfies these two conditions, including for instance a max-linear model on a tree (Gissibl and Klüppelberg, 2018). The multiplicative structure is confirmed in Engelke and Hitz (2018, Proposition 2).

Suppose further that the margins of Y are continuous and belong to the domain of attraction of a standard Fréchet distribution, which may follow from the transformation to Pareto scale. We assume also that the Markov kernel for a pair of adjacent nodes $e = (a, b)$ is determined by a Hüsler–Reiss bivariate copula (Segers, 2019, Examples 3 and 4) with dependence parameter $\theta_e \in (0, \infty)$: for (u_1, u_2) in the unit square,

$$C_e(u_1, u_2) = \exp\left\{ \Phi\left(\frac{\theta_e}{2} + \frac{1}{\theta_e} \ln\Big(\frac{\ln u_1}{\ln u_2} \Big) \right) \ln u_1 + \Phi\left(\frac{\theta_e}{2} + \frac{1}{\theta_e} \ln\Big(\frac{\ln u_2}{\ln u_1} \Big) \right) \ln u_2 \right\}. \tag{7}$$

When $\theta_e \uparrow \infty$ we have $C_e(u_1, u_2) \to u_1 u_2$ corresponding to the independence copula and when $\theta_e \downarrow 0$ we obtain the comonotone copula, i.e., $C_e(u_1, u_2) \to \min(u_1, u_2)$, corresponding to perfect dependence. In the case where every pair of variables on adjacent nodes is related through the copula in (7), the auto-rescaled Markov kernels of Y converge to a log-normal distribution: for $e = (a, b) \in E$ and $x > 0$, we have

$$\lim_{t \to \infty} \mathbb{P}\left(\frac{Y_b}{Y_a} \le x \,\Big|\, Y_a = t \right) = \mathbb{P}(M_e \le x) = \Phi\left(\frac{\ln x - (-\theta_e^2/2)}{\theta_e} \right).$$

Since $\ln M_e \sim \mathcal{N}(-\theta_e^2/2, \theta_e^2)$, the vector $(\ln \Theta_{uv}, v \in V \setminus u)$ is a linear transformation of a Gaussian vector and it is therefore itself Gaussian. Its mean vector $\mu_{V,u}(\theta)$ and its covariance matrix $\Sigma_{V,u}(\theta)$ are given by

$$\{\mu_{V,u}(\theta)\}_v = -\frac{1}{2} \sum_{e \in (u \rightsquigarrow v)} \theta_e^2 \,, \qquad v \in V \setminus u, \tag{8}$$

$$\{\Sigma_{V,u}(\theta)\}_{ij} = \sum_{e \in (u \rightsquigarrow i) \cap (u \rightsquigarrow j)} \theta_e^2 \,, \qquad i, j \in V \setminus u \,. \tag{9}$$

In summary, when Y is a Markov tree on $\mathcal{T} = (V, E)$ with continuous margins, built from pairwise Hüsler–Reiss copulas with dependence parameters θ_e, one for each pair of adjacent variables (Y_a, Y_b) with $e = (a, b) \in E$, the following result holds for every $u \in V$ and as $x \to \infty$:

$$(\ln Y_v - \ln Y_u \mid Y_u = x)_{v \in V \setminus u} \xrightarrow{d} (R_{uv})_{v \in V \setminus u} \sim \mathcal{N}_{|V \setminus u|}\big(\mu_{V,u}(\theta), \Sigma_{V,u}(\theta)\big) \,, \tag{10}$$

where $\mathcal{N}_p$ is the p-variate normal distribution.

The full proof that the extreme value attractor of Y is the Hüsler–Reiss distribution in (3) with structured parameter matrix in (4) is given in the appendix. The covariance matrix $\Sigma_{V,u}(\theta)$ in (9) is the same as the matrix $\Gamma_{W,u}(\Lambda)$ in (2) with $W = V$ and with $\Lambda = \Lambda(\theta)$ given in (4):

$$\begin{aligned} \big\{\Sigma_{V,u}(\theta)\big\}_{ij} &= \sum_{e \in (u \rightsquigarrow i) \cap (u \rightsquigarrow j)} \theta_e^2 = \frac{1}{2} \left(\sum_{e \in (u \rightsquigarrow i)} \theta_e^2 + \sum_{e \in (u \rightsquigarrow j)} \theta_e^2 - \sum_{e \in (i \rightsquigarrow j)} \theta_e^2 \right) \\ &= 2(\lambda_{iu}^2 + \lambda_{ju}^2 - \lambda_{ij}^2) = \{\Gamma_{V,u}(\Lambda(\theta))\}_{ij} \,, \qquad i, j \in V \setminus u. \end{aligned} \tag{11}$$

It is needed to divide by 2 in the second equality, because the parameters on shared edges are added twice. In addition, the Hüsler–Reiss parameters λ_{uv}^2 are proportional to the means:

$$2\lambda_{uv}^2 = \frac{1}{2} \sum_{e \in (u \rightsquigarrow v)} \theta_e^2 = -\{\mu_{V,u}(\theta)\}_v \,, \qquad v \in V \setminus u. \tag{12}$$

Equation (10) continues to hold if we replace the conditioning event $\{Y_u = x\}$ by $\{Y_u > x\}$ (Segers, 2019, Corollary 1). In this form, weak convergence of logarithmic excesses is equivalent to multivariate regular variation and hence to the max-domain of attraction condition thanks to the equivalence of statements (a) and (e) in Theorem 2 in Segers (2019). But since we assumed in Section 2.2 that X belongs to the same max-domain of attraction as Y, we conclude that

$$(\ln X_v - \ln X_u)_{v \in V \setminus u} \mid X_u > t \xrightarrow{d} \mathcal{N}_{|V \setminus u|}\big(\mu_{V,u}(\theta), \Sigma_{V,u}(\theta)\big), \qquad t \to \infty. \tag{13}$$

The latter convergence is a key result that will be used throughout the paper. Up to the parameters of the normal distribution the convergence in (13) appears in Engelke et al. (2015, Theorem 2).

3 Latent variables and parameter identifiability

A typical application of our model arises in relation to quantities measured on river networks that have a tree-like structure.

A general rule is to associate a node to an existing measurement station or to a split/junction even if there is no measurement station. Stations are supposed to generate data for the quantity of interest, so for any node associated to a station there is a corresponding variable. In practice, junctions/splits may lack measurements, in which case there are nodes in the tree with latent variables. This is the case for instances for nodes labelled 2 and 5 in the Seine network in Figure 1.

Let $\mathcal{T} = (V, E)$ be an undirected tree and consider the Hüsler–Reiss distribution (3) with parameter matrix $\Lambda = \Lambda(\theta)$ in (4). When there are nodes with latent variables, the question is whether it is still possible to identify the $d - 1$ free edge parameters θ_e from the distribution of the subvector of observable variables only. Let $U \subseteq V$ denote the set of indices of the observable variables. On the one hand, Eq. (13) implies

$$(\ln X_v - \ln X_u)_{v \in U \setminus u} \mid X_u > t \xrightarrow{d} \mathcal{N}_{|U \setminus u|}\big(\mu_{U,u}(\theta), \Sigma_{U,u}(\theta)\big), \qquad t \to \infty, \tag{14}$$

with $\mu_{U,u}(\theta)$ and $\Sigma_{U,u}(\theta)$ as in (8) and (9) but with V replaced by U. On the other hand, $\mu_{U,u}(\theta)$ and $\Sigma_{U,u}(\theta)$ together determine the stdf l_U of the subvector X_U in (6) through the identities (11) and (12). The question is thus whether the parameter vector θ is still identifiable from the $|U \setminus u|$-variate normal distributions on the right-hand side of (14), where u ranges over U.

Example. Let $X = (X_a, X_b, X_c)$ be a trivariate vector living on $\mathcal{T} = (V = \{a, b, c\}, E = \{(a, b), (b, a), (b, c), (c, b)\})$ and let the variable X_b be latent. Since a parameter is associated to each (undirected) edge of the graph, we have $\theta = (\theta_{ab}, \theta_{bc})$. By (14) we have

$$\ln X_c - \ln X_a \mid X_a > t \xrightarrow{d} \mathcal{N}\big(-(\theta_{ab}^2 + \theta_{bc}^2)/2, (\theta_{ab}^2 + \theta_{bc}^2)\big), \qquad t \to \infty.$$

It is clear that from the limiting normal distribution, we cannot identify θ_{ab} and θ_{bc}.

One approach to the identifiability problem would be to omit nodes with latent variables and redraw the edges. In Figure 2 for instance, the latent variable on node 2 on the left is suppressed leading to the reparametrized tree on the right. However, when there are many latent variables, suppressing nodes (as in Figure 2) runs the risk of changing completely the existing tree and hence the dependence between the variables.

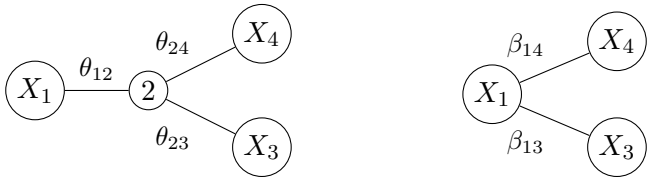


Figure 2: Left: tree on four nodes where node 2 has a latent variable. Right: node 2 has been suppressed and the edges have been redrawn.

In this section it is shown that as long as all nodes with missing variables have degree at least three, the parameters associated to the Hüsler–Reiss distribution of the full vector are still identifiable and hence there is no need to change the tree. To this end, note that by (8), (11) and (12), the vectors $\mu_{U,u}(\theta)$ and the matrices $\Sigma_{U,u}(\theta)$ are determined completely by the path sums

$$p_{ab} = \sum_{e \in (a \rightsquigarrow b)} \theta_e^2, \qquad a, b \in U, \tag{15}$$

and that, vice versa, the values of these path sums are determined by the vectors $\mu_{U,u}(\theta)$ and the matrices $\Sigma_{U,u}(\theta)$. If we know the distribution of $X_U = (X_u)_{u \in U}$, we can compute

the values of these sums, and if we know these sums, we can compute the stdf l_U of X_U. The question is thus whether or not the edge parameters θ_e are identifiable from the values of the path sums p_{ab} for $a, b \in U$. According to the following proposition, there is a surprisingly simple criterion to decide whether this is the case or not.

Proposition 3.1. *Let $\mathcal{T} = (V, E)$ be an undirected tree and let $X = (X_v)_{v \in V}$ have unit Pareto margins and be in the max-domain of attraction of the structured Hüsler–Reiss distribution H_Λ in (3) with parameter matrix $\Lambda = \Lambda(\theta)$ in (4). Let $U \subseteq V$ be the set of nodes corresponding to the observable variables. The parameter vector θ is identifiable from $X_U = (X_u)_{u \in U}$ if and only if every node $u \in V \setminus U$ has degree at least three.*

Proof. Necessity. Let $\bar{U} = V \setminus U \neq \varnothing$ be the set of nodes with latent variables. We need to show that every $v \in \bar{U}$ has degree $d(v)$ at least 3. We will do this by contradiction. As the tree is connected there cannot be a node of zero degree.

First, assume there is $v \in \bar{U}$ such that $d(v) = 1$. The node v must be a leaf node, and in this case there is no path $(a \rightsquigarrow b)$ with $a, b \in U$ that passes by v, and thus θ^2_{uv}, with u the unique neighbour of v, does never appear in the sum (15). Hence θ_{uv} is not identifiable.

Second, assume there exists $v \in \bar{U}$ with $d(v) = 2$. Then v has exactly two neighbours, i and j, say. Every path sum p_{ab} for $a, b \in U$ will contain either the sum of the squared parameters, $\theta^2_{iv} + \theta^2_{jv}$, or neither of these. Hence, the individual edge parameters θ_{iv} and θ_{jv} are not identifiable. (This generalizes the example given before the statement of the theorem.)

Sufficiency. Assume that all nodes with latent variables are of degree three or more.

Let $e = (u, v) \in E$. We will show that there exists a linear combination of the path sums that is equal to θ^2_{uv}.

If $u, v \in U$, then the one-edge path sum $p_{uv} = \theta^2_{uv}$ already meets the condition.

Suppose that $u \in \bar{U}$. By assumption, u has at least two other neighbours besides v, say w and x. If $v \in U$, then put $\bar{v} = v$. Otherwise, start walking at v away from u until you encounter the first visible node, say $\bar{v} \in U$. There must always be such a node, since V is finite and since all leafs are observable by assumption. Similarly, let $\bar{w} \in U$ and $\bar{x} \in U$ be the first visible nodes encountered when walking away from u and starting in w and x, respectively. We can thus observe the sums

$$\begin{aligned} p_{\bar{v}\bar{w}} &= p_{\bar{v}u} + p_{u\bar{w}} \,, \\ p_{\bar{v}\bar{x}} &= p_{\bar{v}u} + p_{u\bar{x}} \,, \\ p_{\bar{w}\bar{x}} &= p_{\bar{w}u} + p_{u\bar{x}} \,. \end{aligned}$$

Since $p_{yz} = p_{zy}$, the previous identities constitute three linear equations in three unknowns that can be solved explicitly, producing the values of $p_{u\bar{v}}$, $p_{u\bar{w}}$, $p_{u\bar{x}}$. In particular, summing the first two equations, subtracting the third, and dividing by two, we find

$$p_{u\bar{v}} = \frac{1}{2} p_{\bar{v}\bar{w}} + \frac{1}{2} p_{\bar{v}\bar{x}} - \frac{1}{2} p_{\bar{w}\bar{x}} \,.$$

If $v \in U$, then $v = \bar{v}$, and $(u \rightsquigarrow v) = \{e\}$, so that the above equation shows how to combine path sums in a linear way to extract $p_{uv} = \theta^2_e$.

If $v \notin U$, then we can repeat the same procedure with u replaced by v. The result is a formula expressing $p_{v\bar{v}}$ as a linear combination of three visible path sums. Now since

$$\theta^2_e = p_{u\bar{v}} - p_{v\bar{v}} \,,$$

we have found a way to extract θ^2_e by a linear combination of at most six visible path sums. □

4 Estimation

We propose three methods for estimating the parameters in the matrix $\Lambda(\theta)$ of the Hüsler–Reiss distribution. The first one, called moment estimator, builds upon the estimator introduced in Engelke et al. (2015). The second estimator is a variation of a maximum likelihood estimator (MLE). The third estimator is based on bivariate extremal coefficients and on the method in Einmahl et al. (2018).

Let $\xi_i = (\xi_{v,i})_{v\in V}$ be an independent random sample from the distribution of ξ satisfying the assumptions in Section 2.2. Further, let $U \subseteq V$ be the set of indices of observable variables and assume that every $u \in V \setminus U$ has degree at least three, so that, by Proposition 3.1, the Hüsler–Reiss edge parameters $\theta_e \in (0,\infty)$ for $e \in E$ are identifiable. The estimators should then be functions of the subvectors $\xi_{U,i} = (\xi_{v,i})_{v\in U}$.

An important remark for this whole section is related to the fact that X as introduced in Section 2.2 should have unit Pareto margins, obtained after the transformation $X_v = 1/(1 - F_v(\xi_v))$ where F_v is the marginal distribution of ξ_v for $v \in V = \{1,\dots,d\}$. It is unrealistic that the functions F_v are known, so in practice their empirical versions are used, $\hat{F}_{v,n}(x) = \big[\sum_{i\le n} \mathbb{1}(\xi_{v,i} \le x)\big]/(n+1)$. The estimates of the edge parameters will then be based upon the sample $\hat{X}_1,\dots,\hat{X}_n$ with coordinates

$$\hat{X}_{v,i} = \frac{1}{1-\hat{F}_{v,n}(\xi_{v,i})}\,, \qquad v \in U, \quad i = 1,\dots,n,$$

considered as a random sample from the distribution of $X_U = (X_u)_{u\in U}$.

A variable indexed by the double subscript W, i will denote the i-th observation of variables on nodes belonging to the set $W \subseteq U$. For instance $\hat{X}_{W,i} = (\hat{X}_{v,i}\,, v \in W)$. Such vectors are taken to be column vectors of length $|W|$. Whenever $W = U$ we just write $\hat{X}_i$.

4.1 Method of moments estimator

Engelke et al. (2015) introduce an estimator of the matrix Λ of the Hüsler–Reiss distribution, based on sample counterparts of the matrices $\Gamma(\,\cdot\,,\Lambda)$ in (2). We will apply their method to the vector of observable variables and then add a least-squares step to extract to the edge parameters θ_e.

As a starting point we take the result in (14) and as suggested by Engelke et al. (2015) for given $k \in \{1,\dots n\}$ we obtain the log-differences

$$\Delta_{uv,i} = \ln \hat{X}_{v,i} - \ln \hat{X}_{u,i}\,, \tag{16}$$

for $u, v \in U$ and for $i \in I_u = \{i = 1,\dots,n : \ln \hat{X}_{u,i} > -\ln(k/n)\}$. The proposed estimators of $\mu_{U,u}$ and $\Sigma_{U,u}$ are the sample mean vector

$$\hat{\mu}_{U,u} = \frac{1}{|I_u|}\sum_{i\in I_u}(\Delta_{uv,i}, v \in U \setminus u)$$

and the sample covariance matrix

$$\hat{\Sigma}_{U,u} = \frac{1}{|I_u|}\sum_{i\in I_u}(\Delta_{uv,i} - \hat{\mu}_{U,u}, v \in U \setminus u)(\Delta_{uv,i} - \hat{\mu}_{U,u}, v \in U \setminus u)^T\,.$$

To estimate the vector of edge parameters $\theta = (\theta_e, e \in E)$, we propose the least squares estimator

$$\hat{\theta}^{\mathrm{MM}}_{n,k} = \operatorname*{arg\,min}_{\theta\in(0,\infty)^{|E|}} \sum_{u\in U} \|\hat{\Sigma}_{U,u} - \Sigma_{U,u}(\theta)\|_F^2\,. \tag{17}$$

where $\|\cdot\|_F$ is the Frobenius norm. In this way, we take advantage of the empirical covariance matrices $\hat{\Sigma}_{U,u}$ for each $u \in U$ and thus of each exceedance set I_u.

In (17), for each $u \in U$, we consider the covariance matrix of the log-differences $\Delta_{uv,i}$ for all $v \in U \setminus u$. However, if v is far away from u in the tree, then the extremal dependence between ξ_u and ξ_v may be weak and the difference $\Delta_{uv,i}$ may carry little information. Therefore, we propose a modified estimator where, for each $u \in U$, we limit the scope to a subset $W_u \subseteq U$ of observable variables that are close to u, producing the estimator

$$\hat{\theta}^{\mathrm{MM}}_{n,k} = \operatorname*{arg\,min}_{\theta \in (0,\infty)^{|E|}} \sum_{u \in U} \|\hat{\Sigma}_{W_u,u} - \Sigma_{W_u,u}(\theta)\|_F^2 \,. \tag{18}$$

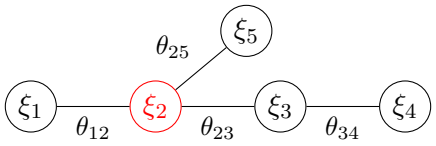
Besides being simpler to compute, the modified estimator (18) performed better than the one in (17) in Monte Carlo experiments. One possible explanation is that by excluding pairs with weak extremal dependence, the bias of the estimator diminishes.

When choosing the sets W_u, care needs to be taken that the parameter vector θ is still identifiable from the collection of covariance matrices $\Sigma_{W_u,u}(\theta)$ for $u \in U$. For $v \in W_u \setminus u$, the variance is the path sum

$$\{\Sigma_{W_u,u}(\theta)\}_{vv} = \sum_{e \in (u \rightsquigarrow v)} \theta_e^2 = p_{uv},$$

see (9) and (15). The set of path sums p_{ab} for $a, b \in U$ in Proposition 3.1 is now reduced to the set of the path sums p_{ab} for $a \in U$ and $b \in W_u$. Whether or not these are still sufficient to identify θ needs to be checked on a case-by-case basis.

Example. Consider the following structure on five nodes where all variables are observable except for the one on node 2:



Clearly, the parameters $\theta_{12}, \theta_{23}, \theta_{34}, \theta_{25}$ are identifiable because the criterion of Proposition 3.1 is satisfied: the node whose variable is latent has degree three.

Suppose we consider the following collection of subsets W_u for $u \in \{1,3,4,5\}$:

$$W_1 = \{1,5\}, \quad W_3 = \{3,4\}, \quad W_4 = \{3,4\}, \quad W_5 = \{1,5\}.$$

The covariance matrices that correspond to these subsets are

$$\Sigma_{W_1,1} = \theta_{12}^2 + \theta_{25}^2 = p_{15}, \qquad \Sigma_{W_4,4} = \theta_{34}^2 = p_{34},$$
$$\Sigma_{W_3,3} = \theta_{34}^2 = p_{34}, \qquad \Sigma_{W_5,5} = \theta_{12}^2 + \theta_{25}^2 = p_{15}.$$

We are not able to identify the parameter $\theta = (\theta_{12}, \theta_{23}, \theta_{34}, \theta_{25})$ because the set of path sums $\{p_{15}, p_{34}\}$ is too small: we have only two equations and four unknowns.

If, instead, we choose the following subsets

$$W_1 = \{1,5,3\}, \quad W_3 = \{1,3,4,5\}, \quad W_4 = \{3,4\}, \quad W_5 = \{1,5\},$$

and consider that the variable exceeding a high threshold has the same index as the index of the subset W_u, the covariance matrices of the vectors (R_{15}, R_{13}), (R_{31}, R_{34}, R_{35}), (R_{43}), (R_{51}) are given by

$$\Sigma_{W_1,1} = \begin{bmatrix} \theta_{12}^2 + \theta_{25}^2 & \theta_{12}^2 \\ \theta_{12}^2 & \theta_{12}^2 + \theta_{23}^2 \end{bmatrix} = \begin{bmatrix} p_{15} & p_{12} \\ p_{12} & p_{13} \end{bmatrix}, \qquad \Sigma_{W_4,4} = \theta_{34}^2 = p_{34},$$

$$\Sigma_{W_3,3} = \begin{bmatrix} \theta_{12}^2+\theta_{23}^2 & 0 & \theta_{23}^2 \\ 0 & \theta_{34}^2 & 0 \\ \theta_{23}^2 & 0 & \theta_{23}^2+\theta_{25}^2 \end{bmatrix} = \begin{bmatrix} p_{13} & 0 & p_{23} \\ 0 & p_{34} & 0 \\ p_{23} & 0 & p_{35} \end{bmatrix}, \quad \Sigma_{W_5,5} = \theta_{12}^2+\theta_{25}^2 = p_{15}.$$

Clearly, the four edge parameters are identifiable from the four covariance matrices, because the set of the path sums $\{p_{15}, p_{13}, p_{12}, p_{34}, p_{23}, p_{35}\}$ is rich enough. From the data one obtains $\hat{\Sigma}_{W_u,u}$ for $u = 1, 3, 4, 5$ and the estimator of $\theta = (\theta_{12}, \theta_{23}, \theta_{34}, \theta_{25})$ according to (18) becomes

$$\operatorname*{arg\,min}_{\theta\in(0,\infty)^4} \sum_{u\in\{1,3,4,5\}} \|\hat{\Sigma}_{W_u,u} - \Sigma_{W_u,u}(\theta)\|_F^2 .$$

4.2 Maximum likelihood estimator

The maximum likelihood estimator (MLE) is again based on the result in (14). This time however we maximize the likelihood function with respect to the parameter θ directly. The MLE uses composite likelihoods based on subtrees.

The likelihood function of p-variate normal data $y_i = (y_{1,i}, \ldots, y_{p,i})$ for $i = 1, \ldots, k$ with mean μ and covariance matrix Σ is given by

$$\prod_{i=1}^{k} \phi_p(y_i - \mu; \Sigma) = (2\pi)^{-kp/2} (\det \Sigma^{-1})^{k/2} \exp\left(-\frac{1}{2} \sum_{i=1}^{k} (y_i - \mu)^T \Sigma^{-1} (y_i - \mu) \right) .$$

As for the method of moment estimator (Section 4.1), we consider for each $u \in U$ a set $W_u \subset U$ of nodes that are close to u in the tree, taking care to include sufficiently many variables so that the edge parameters are still identifiable. Recall the log-differences $\Delta_{uv,i}$ in (16) and the exceedance set I_u right below (16). The maximum likelihood estimator $\hat{\theta}_{n,k}^{\mathrm{MLE}}$ is the maximizer of the composite likelihood

$$L\big(\theta;\, \{\Delta_{uv,i} : v \in W_u \setminus u,\, i \in I_u, u \in U\}\big) = \prod_{u\in U} \prod_{i\in I_u} \phi_{|W_u\setminus u|}\big((\Delta_{uv,i})_{v\in W_u} - \mu_{W_u,u}(\theta); \Sigma_{W_u,u}(\theta)\big) ,$$

where we aggregate the likelihoods of the different normal distributions for all $u \in U$ as if the samples of log-differences are independent, which of course they are not. Results from Monte Carlo simulation experiments (Appendix A.3) show that the performance of the MLE is comparable to the one of the moment estimator and the extremal coefficient estimator.

4.3 Pairwise extremal coefficients estimator

The pairwise extremal coefficients estimator (ECE), based on Einmahl et al. (2018), is based on the bivariate stable tail dependence function (stdf) in (6). It minimizes the weighted distance between the non-parametric estimate of it and the parametric stdf.

The non-parametric estimator of the stdf is given by (Drees and Huang, 1998)

$$\hat{l}_{n,k}(x) = \frac{1}{k} \sum_{i=1}^{n} \mathbb{1}\left(\bigcup_{v\in U} \{ n\hat{F}_{v,n}(\xi_{v,i}) > n + 1/2 - kx_v \} \right) \tag{19}$$

with $x \in [0,\infty)^{|U|}$. Let $q \geq |E|$ be integer and let $x^{(m)} \in [0,\infty)^d$ for $m = 1, \ldots, q$. Consider the vector valued function $L(\theta) = (l(x^{(m)}; \theta))_{m=1}^q$ where $l(\,\cdot\,;\theta)$ is as in (6) with $J = U$ and let $\hat{L}_{n,k} = (\hat{l}_{n,k}(x^{(m)}))_{m=1}^q$ as in (19). Let $\mathcal{Q} \subseteq \{J \subseteq U : |J| = 2\}$ and put $q = |\mathcal{Q}|$. Number

the elements in $\mathcal{Q}$ in some way and put $x^{(m)} = (\mathbb{1}_{\{i \in J\}}, i \in U)$, for J the m-th element of $\mathcal{Q}$. The pairwise extremal coefficients estimator of θ is

$$\hat{\theta}^{\mathrm{ECE}}_{n,k} = \underset{\theta \in (0,\infty)^{|E|}}{\arg\min} \, \|\hat{L}_{n,k} - L(\theta)\|_2^2 \,. \tag{20}$$

The elements of $L(\theta)$ are the pairwise extremal coefficients

$$\ell(x^{(m)}; \theta) = 2\Phi(\sqrt{p_{uv}}/2), \qquad J = \{u, v\}, \tag{21}$$

with p_{uv} the path sum (15).

If $\mathcal{Q}$ is the collection of all possible pairs of elements of U, then the pairwise extremal coefficients give us access to all path sums p_{ab} for $a, b \in U$, and Proposition 3.1 guarantees the identifiability of θ from the pairwise extremal coefficients. If, however, $\mathcal{Q}$ is a smaller set of pairs, then the identifiability of θ needs to be checked on the case at hand.

5 High water levels on the Seine network

Data on water levels were collected from five locations on the Seine river: Paris, Meaux, Melun, Nemours and Sens. The map on Fig. 1 shows part of the actual Seine network. The schematic representation of the graphical model used in the estimation is shown in Fig. 3. The model is relatively small since there are only seven nodes, but because of the variables on nodes 2 and 5 being latent, the application allows us to show that we can still identify all six parameters $\theta_1, \ldots, \theta_6$ of extremal dependence. For more information on the data set, some summary statistics and details on data preprocessing, we refer to Appendix A.4.

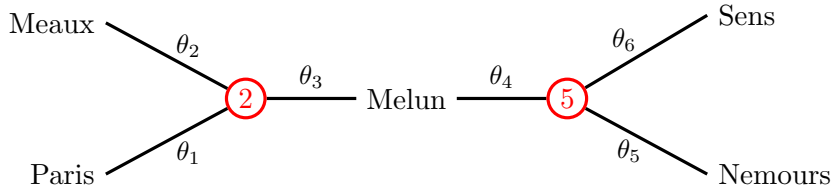


Figure 3: The graphical model on the selected locations in the Seine network.

5.1 First results and identifiability of the parameters

We used all three estimators to obtain estimates of the six parameters of extremal dependence. For the pairwise extremal coefficient estimator (ECE) it is possible to calculate standard errors thanks to the asymptotic distribution derived in Einmahl et al. (2018, Theorem 2.2). Computational details for the standard errors follow in Appendix A.5.

The estimates and their 95% confidence intervals are displayed in Fig. 4 for two of the parameters, namely θ_1 and θ_4. The plots for $\theta_2, \theta_5, \theta_6$ are similar to the one for θ_1: the 95% confidence intervals never include zero, suggesting that the extremal dependence between the corresponding variables is not perfect and hence that the edges cannot be collapsed. In Section 3, we alluded to the possibility of circumventing the issue of latent variables by suppressing nodes and redrawing edges. The fact that the confidence intervals do not include zero indicate that doing so would have produced a misleading picture of extremal dependence.

The plot of θ_3, similarly to the plot of θ_4, does contain a segment over k where the lower confidence bound reaches zero: for θ_4 this is approximately $k \in [260, 360]$, while for θ_3 it is $k \in [90, 180]$. Although the confidence intervals for θ_4 and θ_3 indicate some instability of the

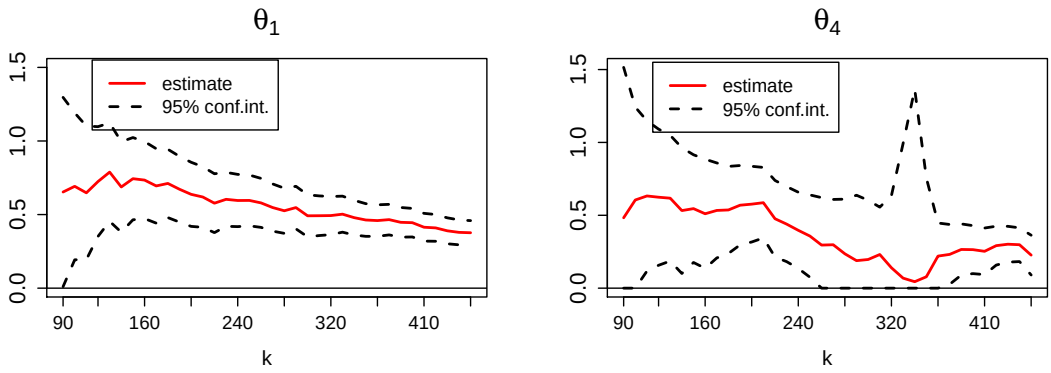


Figure 4: Point estimates and confidence intervals for the pairwise extremal coefficient estimator (ECE).

estimated parameters, we believe that collapsing the edges is not advisable. Moreover the river distance, which is one of the important factors in tail dependence (Asadi et al., 2015), is rather long between {2, Melun} and {Melun, 5}, so that there is no physical motivation for collapsing the corresponding edges.

For a point estimate per parameter we need to average out over a range of k. The chosen range per estimator and per parameter need not be the same. As a rule we select a range around the beginning where the estimates start stabilizing around a certain level, omitting the most volatile part for relatively small k. Most of the time we thus consider $k \in [100, 200]$. In this way we end up with the point estimates displayed for comparison in Fig. 5. Given

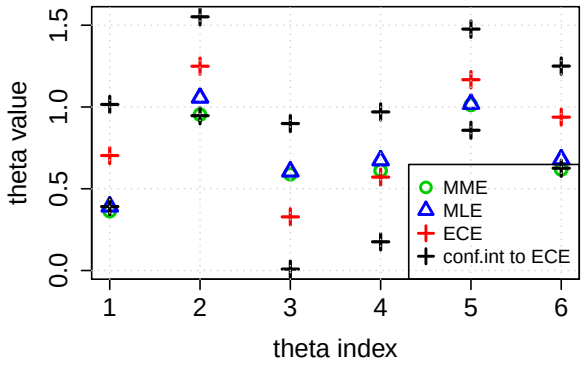


Figure 5: Point estimates of the model parameters $\theta_1, \ldots, \theta_6$ for the Seine data and confidence intervals only for the ECE.

the similarities between the MME and MLE, the estimates are pooled in an average of the two for each parameter and these are the ones used further on.

5.2 Considerations on the goodness-of-fit of the model

Here we look at some informal criteria of how well the model from Section 2.2 describes the real data. We compare non-parametric and model-based estimates of quantities describing extremal dependence such as pairwise and triple-wise extremal coefficients and the Pickands dependence function.

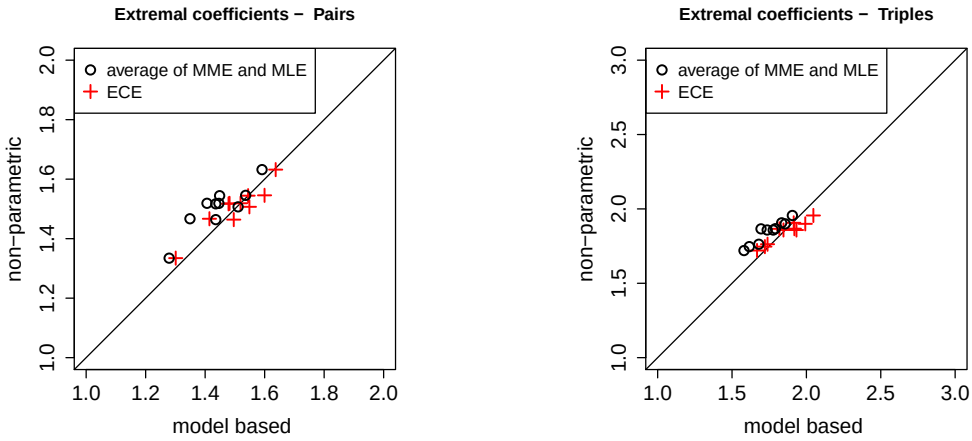


Figure 6: Non-parametric vs model-based extremal coefficients for pairs (left) and triples (right).

The extremal coefficient in an arbitrary subset $J \subseteq U$ is the stable tail dependence function evaluated at coordinates $\mathbb{1}_J = (\mathbb{1}_{j\in J}, j \in U)$, i.e.,

$$l(\mathbb{1}_J) = \lim_{t\to\infty} t\,\mathbb{P}(\textstyle\bigcup_{j\in J} X_j > t)\,,$$

where $X_j = 1/(1 - F_j(\xi_j))$ as in Section 2.2. Here we focus on extremal coefficients between pairs and triples. For the model from Section 2.2 the bivariate extremal coefficient has the simple form (21). The extremal coefficient $l(\mathbb{1}_J)$ is always between 1 and $|J|$, corresponding to perfect extremal dependence and extremal independence, respectively.

The empirical extremal coefficient for the variables in the set J is given by

$$\hat{l}_{n,k}(\mathbb{1}_J) = \frac{1}{k}\sum_{i=1}^{n} \mathbb{1}\left(\bigcup_{j\in J}\{n\hat{F}_{j,n}(\xi_{j,i}) > n - k + 1/2\}\right)\,.$$

Fig. 6 compares the model-based coefficients, $\hat{l}(\mathbb{1}_J, \theta) = l(\mathbb{1}_J, \hat{\theta})$, with the empirical extremal coefficients, $\hat{l}_{n,k}(\mathbb{1}_J)$, for pairs and triples. At least visually the fit is quite good for both estimators, the average of MLE and MME on the one hand and the ECE on the other hand.

It should be noted that when there are locations with latent variables, it is impossible to compute the empirical extremal coefficients involving any of that locations. However the fact that we can identify all the parameters allows us to come up with model-based estimates of the extremal coefficients even in this case.

As another visual check on the goodness-of-fit of the assumed model we consider the bivariate Pickands dependence function, usually denoted by $A(w)$ for $w \in [0, 1]$. For the Hüsler–Reiss extreme-value distribution, it is equal to

$$\begin{aligned} A_{u,v}(w;\theta) &= l(1-w, w; \theta) \\ &= (1-w)\,\Phi\left(\frac{\ln(\frac{1-w}{w}) + \frac{1}{2}p_{uv}}{\sqrt{p_{uv}}}\right) + w\,\Phi\left(\frac{\ln(\frac{w}{1-w}) + \frac{1}{2}p_{uv}}{\sqrt{p_{uv}}}\right)\,, \end{aligned}$$

with p_{uv} as in (15). Hence the model-based estimator of $A_{u,v}(w;\theta)$ is $A_{u,v}(w;\hat{\theta}_{n,k})$ where $\hat{\theta}_{n,k}$ can be the average MME/MLE or the ECE.

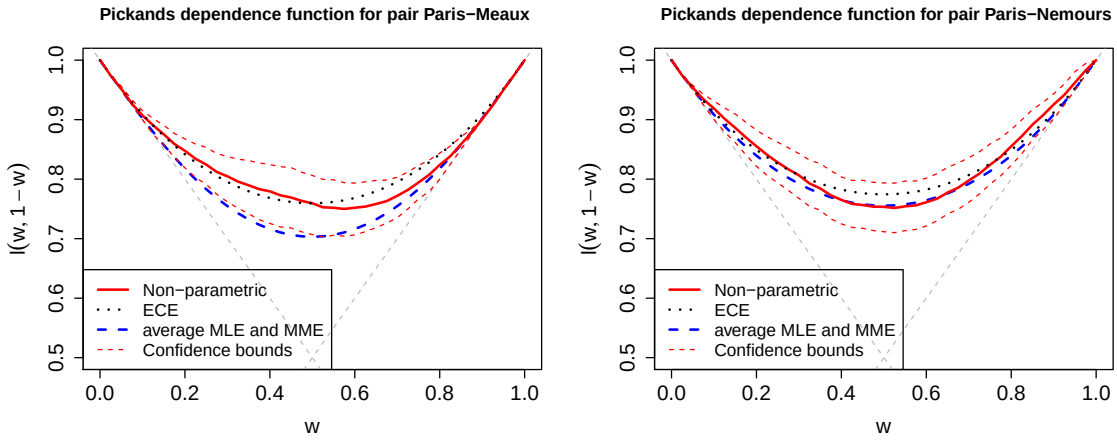


Figure 7: The empirical and model-based Pickands dependence function computed using the pooled ML and MM estimates and the extremal coefficients estimates.

The non-parametric counterpart of the Pickands dependence function is

$$\hat{A}_{u,v}(w) = \frac{1}{k}\sum_{i=1}^{n} \mathbb{1}\{n\hat{F}_{u,n}(\xi_{u,i}) > n - k(1-w) + 1/2 \text{ or } n\hat{F}_{v,n}(\xi_{v,i}) > n - kw + 1/2\}.$$

The model-based Pickands dependence function is compared to the empirical counterpart in Fig. 7. The plot is complemented with non-parametric 95% confidence intervals for $A(w)$ computed by the bootstrap method introduced in Kiriliouk et al. (2018, Section 5). The general idea of the method is to approximate the distribution of $\sqrt{k}(\hat{l}_{n,k} - l)$ by the distribution of $\sqrt{k}(\hat{l}^*_{n,k} - \hat{l}^{\beta}_{n,k})$ where $\hat{l}^*_{n,k}$ is the empirical stable tail dependence function based on the ranks of a sample of size n from the Beta empirical copula and $\hat{l}^{\beta}_{n,k}$ is the stdf based on the empirical Beta copula using the ranks of the original sample $(\xi_{v,i}, i = 1, \ldots, n; v \in U)$. A detailed description of the bootstrap derived confidence intervals is provided in Appendix A.6.

5.3 Flow-connectedness and tail dependence

Fig. 9 illustrates the tail dependence in the Seine network through a heat map of the pairwise extremal coefficients. In a study by Asadi et al. (2015) of data from the Danube, it was found that a key factor for the extremal dependence between two locations is whether they are flow connected or not.

Two locations are flow connected if one of them is downstream of the other one. Flow connectedness often dominates the importance of river distance or Euclidean distance. Two distant nodes might be much more tail dependent if they are flow connected than two nodes that are at close distance but are not flow connected. This effect is confirmed in our data too and is illustrated in Fig. 8. The cities of Sens and Nemours are not flow connected but the Euclidean and river distance between them is smaller than the one between the flow connected cities of Sens and Paris. Still, the tail dependence seems to be stronger for the flow connected pair of locations.

From the heat maps in Fig. 9 it can be seen that pairs with larger tail dependence (smaller extremal coefficient) are indeed flow connected. According to both estimators the tail dependence between Paris and locations Melun, 2 and 5 is the strongest.

We look once again at the pairwise upper tail dependence, this time presenting the Seine network in Fig. 10 with edges weighted by the tail dependence coefficients, defined for a pair

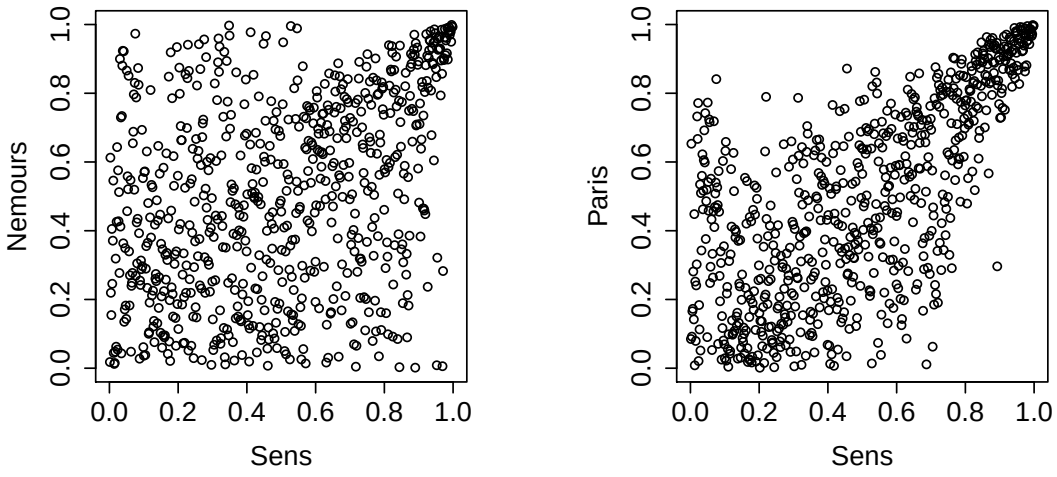


Figure 8: Scatterplots of rank transformed data for two pairs of locations. Sens and Nemours (left) are not flow connected while Sens and Paris (right) are flow connected. It can be seen from the Seine map in Fig. 1 that the river and Euclidean distance from Sens to Nemours is much smaller than the one from Sens to Paris. However the tail dependence seems to be stronger for the second pair of locations.

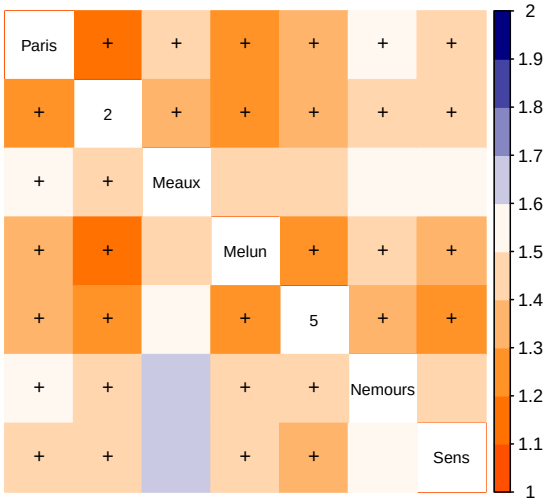


Figure 9: Heat map of the extremal coefficients. The upper diagonal is computed using the pooled MM and ML estimates and the lower diagonal uses the EC estimates. The crosses denote flow connected nodes.

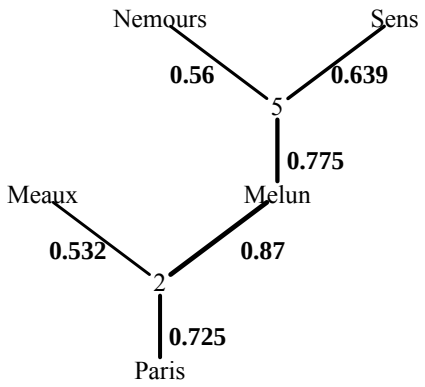


Figure 10: The tree of the tail dependence coefficients in (22), using the EC estimates.

(u, v) by

$$2\{1 - \ell(\mathbb{1}_{\{u,v\}}; \theta)\} \tag{22}$$

with $\ell(\mathbb{1}_{\{u,v\}}; \theta)$ the pairwise extremal coefficient in (21). The largest dependence is observed on the path from location 5 to Paris, which corresponds to the Seine river itself.

6 Conclusion

We have presented a statistical model suitable for studying extremal dependence between random variables which live on a tree-like network. Under minimal assumptions on the marginal and the joint distributions we have shown that such a model is motivated by the existence of a particular parametric graphical model introduced in Segers (2019). Namely, a limiting Hüsler–Reiss distribution with structured variogram matrix is obtained from the extreme value limit of a vector which satisfies the global Markov properties with respect to the given tree and whose distribution is composed of bivariate Hüsler–Reiss copulas.

The central point and contribution of the paper is related to the identifiability criterion in case of latent variables. This situation occurs in applications on river networks, when measurements on junctions or split locations are missing. The tail dependence parameters are uniquely identified if and only if all nodes with latent variables are of degree at least three. As this characterization is due to the special structure of the variogram matrix of the Hüsler–Reiss distribution, it may not be applicable to other extreme value distributions.

We fitted the model to water level data on the Seine network using three different estimators, based on the method of moments, on maximum likelihood, and on pairwise extremal coefficients. Comparisons of non-parametric and model-based tail dependence quantities confirmed the goodness-of-fit of the structured Hüsler–Reiss distribution. The results of the inference suggested that circumventing the issue of latent variables by omitting nodes and redrawing edges would have led to an overly simplified model. From the fitted

model, we could compute estimates of extremal coefficients involving latent variables, which would have been impossible by a non-parametric procedure.

A Appendix

A.1 Comparison with the tree model in Lee and Joe (2018)

We give an example of a structured Hüsler–Reiss tree graphical model which is parametrized as suggested by Lee and Joe (2018, Section 4.1). We compare the resulting parameter matrix Λ with the one proposed in (4).

Consider a Gaussian graphical model on a linear tree of four nodes:

$$X_i = Z_1 + \cdots + Z_i\,, \qquad i = 1, \ldots, 4,$$

where $Z_1, \ldots, Z_4$ are independent Gaussian variables with variances $\tau_1^2 > 0$ and

$$\tau_i^2 = \tau_2^2(1 + \tau_2^2)^{i-2}, \qquad i \in \{2, 3, 4\}.$$

This definition of the variances guarantees that the correlation coefficient ρ_{ij} between X_i and X_j only depends on the distance, in line with the example in Lee and Joe (2018, Example in Table 2, Section 4.3):

$$\rho_{ij} = (1 + \tau_2^2)^{-|i-j|/2}, \qquad i, j \in \{1, \ldots, 4\}.$$

The Hüsler–Reiss parameter matrix proposed in Lee and Joe (2018, Section 4.1, p. 159) is then

$$\Lambda = (\lambda_{ij}^2)_{i,j=1}^4, \qquad \text{with } \lambda_{ij}^2 = \nu\,(1 - \rho_{ij})\,,$$

with $\nu > 0$ an additional parameter.

Alternatively, consider the matrix Λ that would follow from the parametrization in (4). The parameters $\theta_{12}, \theta_{23}, \theta_{34}$ are then given by the standard deviations of the increments $Z_2 = X_2 - X_1$, $Z_3 = X_3 - X_2$ and $Z_4 = X_4 - X_3$, i.e.,

$$\begin{aligned}
\theta_{12}^2 &= \tau_2^2, \\
\theta_{23}^2 &= \tau_3^2 = \tau_2^2(1 + \tau_2^2), \\
\theta_{34}^2 &= \tau_4^2 = \tau_2^2(1 + \tau_2^2)^2.
\end{aligned}$$

The Hüsler–Reiss parameter matrix in (4) then becomes $\Lambda(\theta) = (\lambda_{ij}^2)_{i,j=1}^4$ with

$$\begin{aligned}
\lambda_{12}^2 &= \tfrac{1}{4}\theta_{12}^2, & \lambda_{13}^2 &= \tfrac{1}{4}(\theta_{12}^2 + \theta_{23}^2), & \lambda_{14}^2 &= \tfrac{1}{4}(\theta_{12}^2 + \theta_{23}^2 + \theta_{34}^2) \\
\lambda_{23}^2 &= \tfrac{1}{4}\theta_{23}^2, & \lambda_{24}^2 &= \tfrac{1}{4}(\theta_{23}^2 + \theta_{34}^2), & & \\
\lambda_{34}^2 &= \tfrac{1}{4}\theta_{34}^2. & & & &
\end{aligned}$$

The structural differences between the Lee–Joe parametrization and ours are apparent. We wish to emphasize that the structure we propose corresponds to the one of the extreme-value attractor of a probabilistic graphical model with unit Pareto margins and Markov tree structure specified by Hüsler–Reiss copulas (Section 2.3).

A.2 The stable tail dependence function of Y from Section 2.3

Let $Y = (Y_1, \ldots, Y_d)$ be a random vector as described in Section 2.3: it satisfies the global Markov property with respect to the undirected tree $\mathcal{T} = (V, E)$ with $V = \{1, \ldots, d\}$, it has unit Pareto margins and its joint distribution is determined by $d-1$ bivariate Hüsler–Reiss copulas (7) with dependence parameters $\theta_e \in (0, \infty)$ for $e \in E$. We will show that the stable tail dependence function (stdf) of Y exists and is equal to l in (6) with $J = V$.

By the inclusion–exclusion principle, the stable tail dependence function of Y is

$$\begin{aligned} l(x_1, \ldots, x_d) &= \lim_{t\to\infty} t\,\mathbb{P}(Y_1 > t/x_1 \text{ or } \ldots \text{ or } Y_d > t/x_d) \\ &= \sum_{i=1}^{d} (-1)^{i-1} \sum_{\substack{W\subseteq V \\ |W|=i}} \lim_{t\to\infty} t\,\mathbb{P}(Y_v > t/x_v, v \in W) \end{aligned} \tag{23}$$

for $x \in (0, \infty)^d$. For any $W \subseteq V$ and any $u \in W$, it holds by (10) that

$$\begin{aligned} \lim_{t\to\infty} t\,\mathbb{P}(Y_v > t/x_v, v \in W) &= \lim_{t\to\infty} t \frac{1}{t/x_u} \mathbb{P}\left(\frac{Y_u}{t/x_u} \frac{Y_v}{Y_u} > \frac{x_u}{x_v},\, v \in W \setminus u \,\middle|\, Y_u > \frac{t}{x_u} \right) \\ &= x_u\,\mathbb{P}(\zeta \Theta_{uv} > x_u/x_v, v \in W \setminus u)\,, \end{aligned}$$

where $\Theta_{uv} = \exp(R_{uv})$ and where ζ is a unit Pareto variable, independent of $(\Theta_{uv})_{v\in V\setminus u}$. Using the fact that $1/\zeta$ is a uniform variable on $[0,1]$ and that $\Theta_{uu} = 1$ we have

$$\begin{aligned} & x_u\,\mathbb{P}(\zeta\Theta_{uv} > x_u/x_v, v \in W \setminus u) \\ &= x_u\,\mathbb{P}\left(1/\zeta < \min\{(x_v/x_u)\Theta_{uv}, v \in W \setminus u\}\right) \\ &= x_u\,\mathbb{E}[\min\{1, (x_v/x_u)\Theta_{uv}, v \in W \setminus u\}] = \mathbb{E}[\min\{x_v\Theta_{uv}, v \in W\}] \\ &= \int_0^{x_u} \mathbb{P}\big(x_v\Theta_{uv} > y, v \in W \setminus u\big)\,\mathrm{d}y \\ &= \int_{-\ln x_u}^{\infty} \mathbb{P}\big(R_{uv} > (-\ln x_v) - z, v \in W \setminus u\big)\,\exp(-z)\,\mathrm{d}z \end{aligned}$$

upon a change of variable $y = \exp(-z)$. Since $(R_{uv})_{v\in V\setminus u}$ is multivariate normal with mean vector $\mu_{V,u}(\theta)$ and covariance matrix $\Sigma_{V,u}(\theta)$, we obtain from (23) that the stdf of Y is equal to $-\ln H_{\Lambda(\theta)}(1/x_1, \ldots, 1/x_d)$, with H_Λ the cumulative distribution function in Eqs. (3.5)–(3.6) in Hüsler and Reiss (1989), but with unit Fréchet rather than Gumbel margins. By Remark 2.5 in Nikoloulopoulos et al. (2009), this stdf is equal to the one given in (6), as required.

A.3 Finite-sample performance of the estimators

We assess the performance of the three estimators introduced in Section 4 by numerical experiments involving Monte Carlo simulations.

The data are generated according to the graphical model $(\xi_v)_{v\in V}$ presented in Fig. 11. More specifically, let $f_u(x_u)$ for any $u \in V$ be the marginal density function of the variable ξ_u and let $x_j \mapsto f_{j|v}(x_j \mid x_v)$ be the conditional density function of ξ_j given $\xi_v = x_v$. Then for arbitrary, fixed $u \in V$, an observation of $(\xi_v)_{v\in V}$ is generated according to the following factorization of the joint density function:

$$f\big((x_v)_{v\in V}\big) = f_u(x_u) \prod_{(j,v)\in E_u} f_{j|v}(x_j \mid x_v),$$

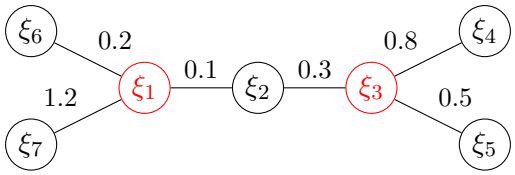


Figure 11: Tree used for the graphical model underlying the data-generating process in the simulation study in Appendix A.3. The value of the parameters are $\theta_{12} = 0.1$, $\theta_{23} = 0.3$, $\theta_{34} = 0.8$, $\theta_{35} = 0.5$, $\theta_{16} = 0.2$ and $\theta_{17} = 1.2$. Variables ξ_1 and ξ_3 are latent.

with E_u the set of edges directed away from u. The densities $f_{j|v}$ for $(j, v) \in E_u$ are determined by the bivariate Hüsler–Reiss copula density with parameter $(\theta_{jv}, (j, v) \in E_u)$. The marginal densities, including f_u, are unit-Fréchet densities, i.e., $f_j(x_j) = \exp(-1/x_j)/x_j^2$. An observation x_j from the conditional distribution of ξ_j given $\xi_v = x_v$ is generated via the inverse function of $x_j \mapsto F_{j|v}(x_j|x_v)$, the conditional distribution function of ξ_j given $\xi_v = x_v$. To do so, the equation $F_{j|v}(x_j|x_v) - p = 0$ is solved numerically as a function in x_j for fixed $p \in (0, 1)$. The choice of the Hüsler–Reiss bivariate copula gives the following expression for $F_{j|v}(x_j \mid x_v)$:

$$\Phi\left(\frac{\theta_{jv}}{2} + \frac{1}{\theta_{jv}} \ln \frac{x_j}{x_v}\right) \cdot \exp\left[-\frac{1}{x_v}\left\{\Phi\left(\frac{\theta_{jv}}{2} + \frac{1}{\theta_{jv}} \ln \frac{x_j}{x_v}\right) - 1\right\} - \frac{1}{x_j}\Phi\left(\frac{\theta_{jv}}{2} + \frac{1}{\theta_{jv}} \ln \frac{x_v}{x_j}\right)\right].$$

After generating all the variables $(\xi_v)_{v \in V}$ as described above, independent standard normal noise $\varepsilon \sim \mathcal{N}_d(0, I_d)$ is added, moving the data-generating process away from the limiting extreme-value attractor. The data on nodes 1 and 3 are deleted and not used in the estimation so as to mimic a model with two latent variables, ξ_1 and ξ_3; according to Proposition 3.1, the six dependence parameters are still identifiable. In this way, we generate 200 samples of size $n = 1000$. The estimators are computed with threshold tuning parameter $k \in \{25, 50, 100, 150, 200, 300\}$.

The bias, standard deviation and root mean squared errors of the three estimators are shown in Fig. 12 and Fig. 13 for the six parameters. The MME and MLE are computed with the sets W_u being $W_2 = \{2, 4, 5, 6, 7\}$, $W_4 = W_5 = \{2, 4, 5\}$, and $W_6 = W_7 = \{2, 6, 7\}$. As is to be expected, the absolute value of the bias is increasing with k, while the standard deviation is decreasing and the mean squared error has a U-shape and eventually increases with k. The MME and MLE have very similar properties. For larger values of the true parameter, e.g. $\theta_{34} = 0.8$ and $\theta_{17} = 1.2$, all three estimators perform in a comparable way. The ECE tends to have larger absolute bias and standard deviation for smaller values of the true parameters.

A.4 Seine case study: data preprocessing

The data represent water level in centimeters at the five locations mentioned above and it was obtained from *Banque Hydro*, http://www.hydro.eaufrance.fr, a web-site of the Ministry of Ecology, Energy and Sustainable Development of France providing data on hydrological indicators across the country. The dataset encompasses the period from January 1987 to April 2019 with gaps for some of the stations.

Two major floods in Paris make part of our dataset: the one in June 2016 when the water level was measured at 6.01m and the one in end January 2018 with slightly less than 6m measured water level in Paris too. A flood of the similar magnitude to the one in 2016 and 2018 occurred in 1982 and for comparison the biggest reported[1] flood in Paris is the

[1] According to the report of the Organisation for Economic Co-operation and Development (OECD) *Preventing the flooding of the Seine in the Paris – Ile de France region* - p.4.

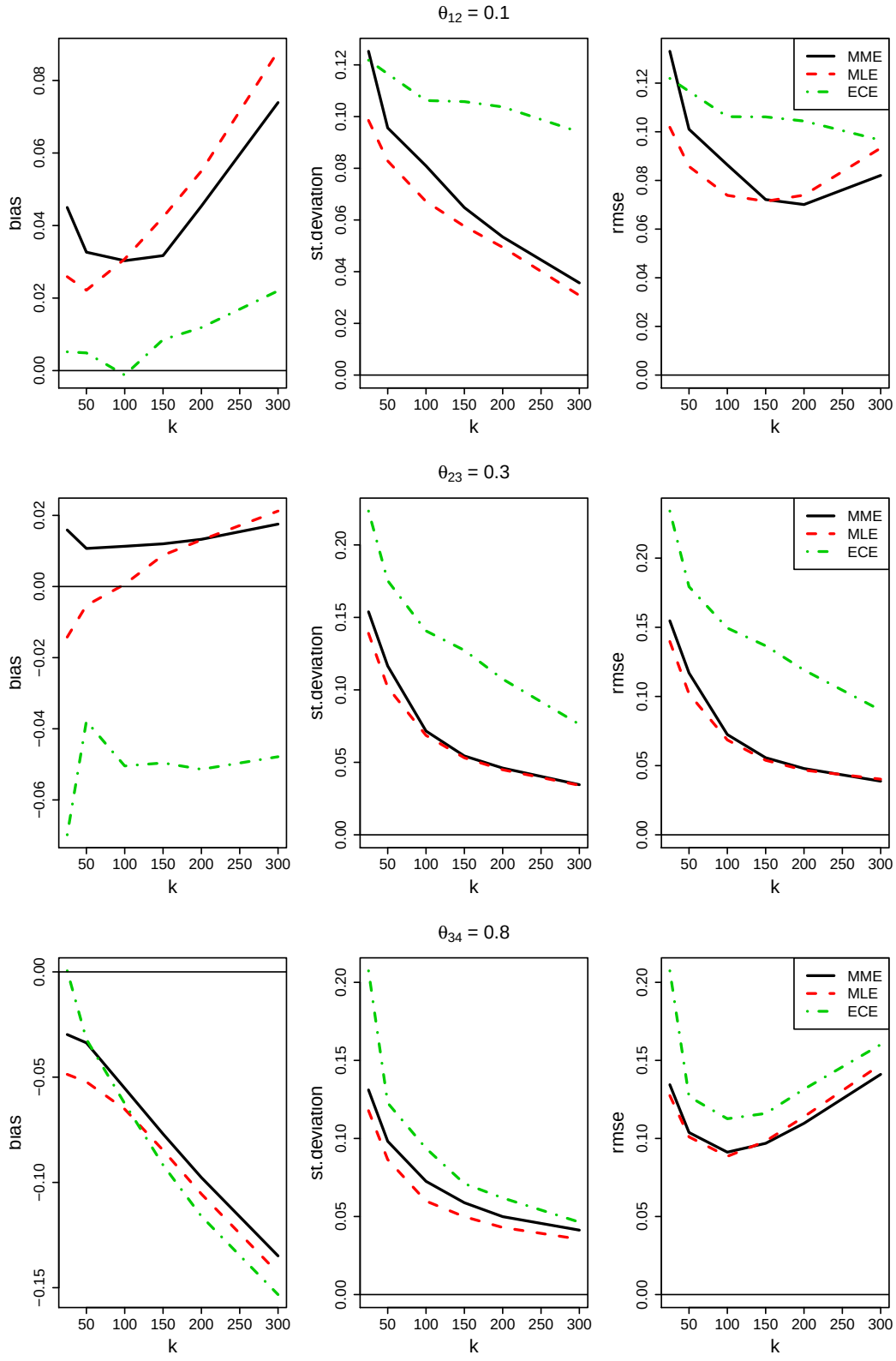


Figure 12: Bias (left), standard deviation (middle) and root mean squared error (right) of the method of moment estimator (MME), maximum likelihood estimator (MLE) and pairwise extremal coefficient estimator (ECE) of the parameters θ_{12} (top), θ_{23} (middle), and θ_{34} (bottom) as a function of the threshold parameter k. Model and settings as described in Appendix A.3.

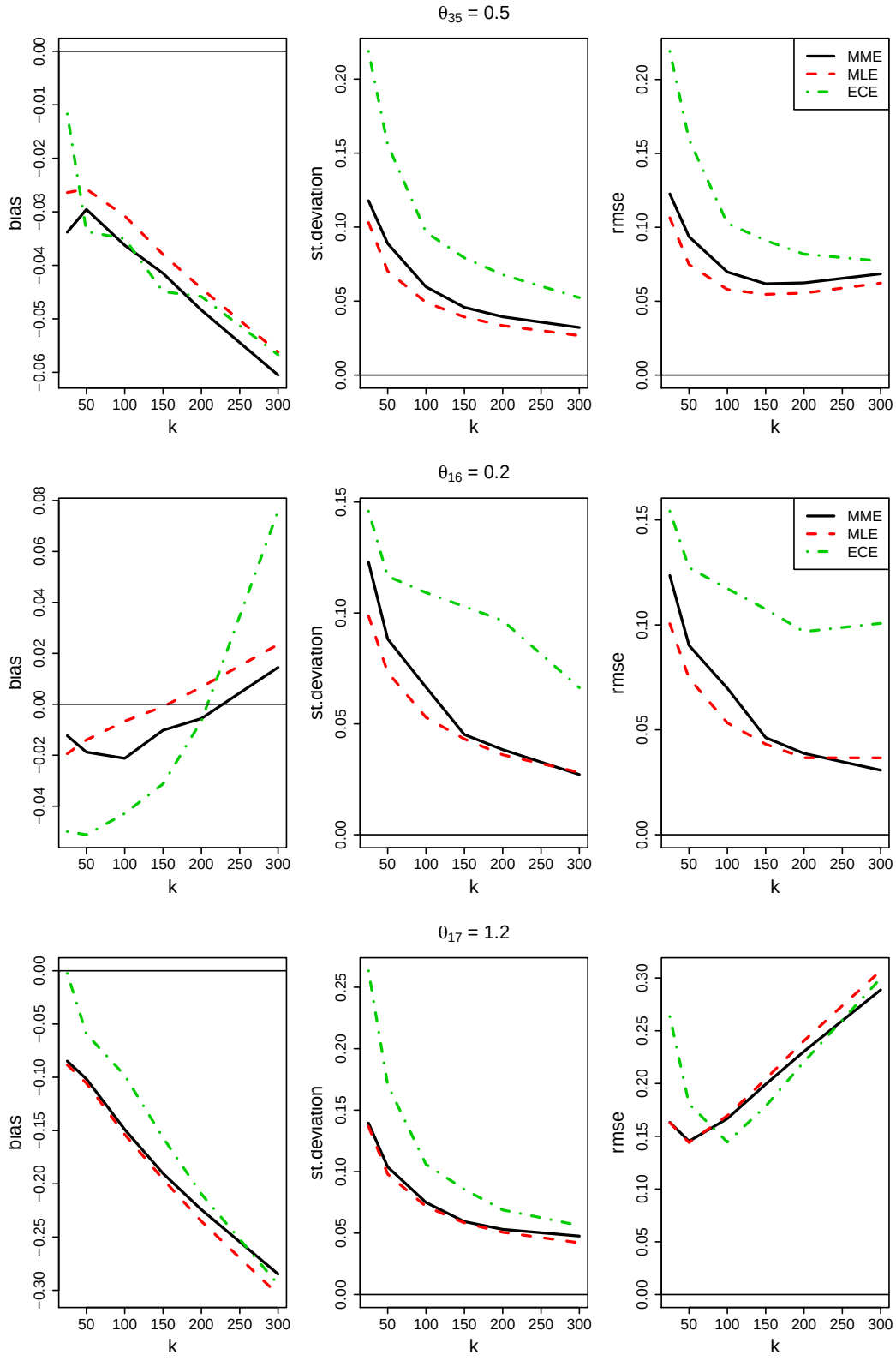


Figure 13: Bias (left), standard deviation (middle) and root mean squared error (right) of the method of moment estimator (MME), maximum likelihood estimator (MLE) and pairwise extremal coefficient estimator (ECE) of the parameters θ_{35} (top), θ_{16} (middle), and θ_{17} (bottom) as a function of the threshold parameter k. Model and settings as described in Appendix A.3.

one in 1910 when the level in Paris reached 8.6m.

Table 1 shows the average and the maximum water level per station observed in the complete dataset. The maxima of Paris, Meaux, Melun and Nemours occurred either during the floods in June 2016 or the floods in January 2018, which can be seen from Table 2 which displays the annual maxima at the five locations and the date of occurrence.

Station	Paris	Meaux	Melun	Nemours	Sens
Period	1 Jan 1990 - 9 Apr 2019	1 Nov 1999 - 9 Apr 2019	1 Oct 2005 - 9 Apr 2019	16 Jan 1987 - 9 Apr 2019	1 Jan 1990 - 9 Apr 2019
(#obs)	(10,621)	(6,287)	(4,443)	(10,154)	(9,159)
Mean (cm)	139.11	275.85	296.61	210.07	133.46
Max (cm)	601.95	468.70	545.48	439.03	333.80

Table 1: Average and maximum water level per station in the whole dataset.

From Table 2 it can be observed that for many of the years the dates of maxima occurrence identify a period of several consecutive days during which the extreme event took place. For instance the maxima in 2007 occurred all in the period 4–8 March, which suggests that they make part of one extreme event. Similar examples are the periods 25–31 Dec 2010, 4–12 Feb 2013, 2–4 June 2016, etc. For most of the years this period spans between 3 and 7 days. We will take this into account when forming independent events from the dataset. In particular we choose a window of 7 consecutive calendar days within which we believe the extreme event have propagated through the seven locations. We have experimented with different length of that window, namely 3 and 5 days event period, but we have found that the estimation and analysis results are robust to that choice.

Fig. 14 illustrates the water levels attained at the different locations during selected years from Table 2. The maxima of Sens, Nemours and Meaux seem to be relatively homogeneous compared to the maxima in Paris.

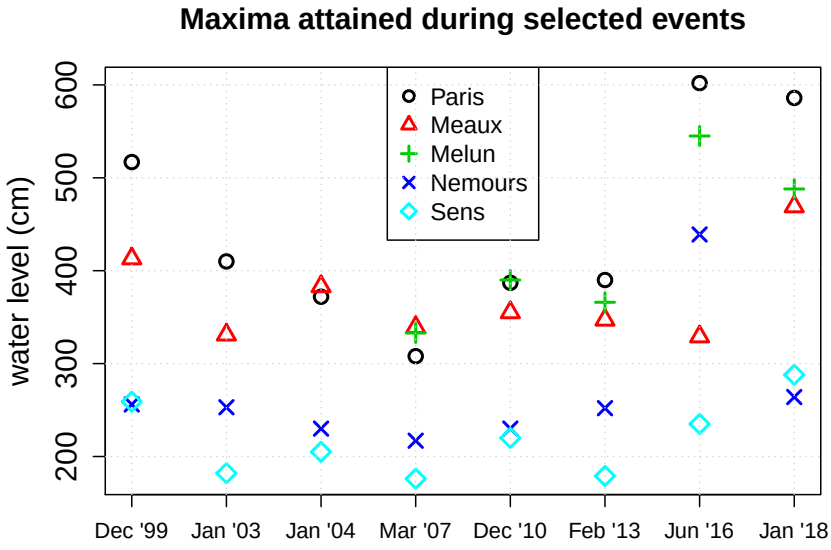


Figure 14: Plot of maxima attained at each location during selected events from Table 2.

For all of the stations water level is recorded several times a day and we take the daily average to form a dataset of daily observations. Accounting for the gaps in the mentioned

Year	Paris		Meaux		Melun		Nemours		Sens	
	date	cm	date	cm	date	cm	date	cm	date	cm
1987	n/a	n/a	n/a	n/a	n/a	n/a	15/11	221	n/a	n/a
1988	n/a	n/a	n/a	n/a	n/a	n/a	13/02	247	n/a	n/a
1989	n/a	n/a	n/a	n/a	n/a	n/a	04/03	213	n/a	n/a
1990	17/02	254	n/a	n/a	n/a	n/a	03/07	217	18/02	183
1991	10/01	339	n/a	n/a	n/a	n/a	23/04	212	04/01	175
1992	06/12	293	n/a	n/a	n/a	n/a	15/01	218	06/12	170
1993	28/12	377	n/a	n/a	n/a	n/a	26/09	217	26/12	184
1994	11/01	478	n/a	n/a	n/a	n/a	19/10	253	09/01	260
1995	30/01	500	n/a	n/a	n/a	n/a	21/03	277	28/01	259
1996	04/12	324	n/a	n/a	n/a	n/a	03/12	219	04/12	194
1997	28/02	313	n/a	n/a	n/a	n/a	03/07	214	n/a	n/a
1998	02/05	358	n/a	n/a	n/a	n/a	21/12	216	n/a	n/a
1999	31/12	517	30/12	413	n/a	n/a	30/12	252	31/12	259
2000	01/01	515	02/01	407	n/a	n/a	07/06	233	01/01	239
2001	25/03	517	30/03	427	n/a	n/a	16/03	260	17/03	334
2002	03/03	410	03/03	403	n/a	n/a	01/01	272	01/01	200
2003	08/01	410	09/01	331	n/a	n/a	05/01	253	06/01	182
2004	21/01	372	21/01	383	n/a	n/a	16/01	230	20/01	205
2005	17/02	192	22/01	296	07/12	306	24/01	217	16/02	152
2006	14/03	340	08/10	333	13/03	357	11/03	219	12/03	223
2007	05/03	308	08/03	339	05/03	333	04/03	217	05/03	176
2008	29/03	301	01/01	250	23/03	342	15/04	219	23/03	167
2009	26/01	169	03/09	288	25/12	311	25/01	218	25/01	152
2010	28/12	387	31/12	355	27/12	390	25/12	230	26/12	220
2011	01/01	337	07/01	347	18/12	356	09/10	287	18/12	167
2012	09/01	330	23/12	308	09/01	353	05/01	220	08/01	186
2013	09/02	390	12/02	347	05/02	366	04/02	252	07/05	221
2014	03/03	273	13/12	295	16/02	321	02/03	226	15/02	157
2015	07/05	347	21/11	295	07/05	389	05/05	255	06/05	211
2016	03/06	602	03/06	329	03/06	545	02/06	439	04/06	235
2017	07/03	243	28/12	304	12/01	307	08/03	221	08/03	151
2018	29/01	586	02/02	469	28/01	488	24/01	264	26/01	288
2019	03/02	222	31/03	292	22/01	314	02/02	216	26/02	149

Table 2: Annual maxima for all stations. We highlighted some of the years where there is a clear indication that the dates of the occurrence of the maxima at the different locations form a period of several consecutive days. The maxima attained during this period across stations can thus be considered as one extreme event. The water level in centimeters is rounded to the nearest integer.

period (see Table 1 and Table 2) we end up with a dataset of 3408 daily observations in the period 1 October 2005 - 8 April 2019. The dataset represents five time series each of length 3408. We consider two sources of non-stationarity: seasonality and serial correlation.

The serial correlation can be due to closeness in time or presence of long term time trend in the observations. We first apply a declustering procedure, similar to the one in Asadi et al. (2015) in order to form a collection of supposedly independent events. As a first step each of the series is transformed in ranks and the sum of the ranks is computed for every day in the dataset. The day with the maximal rank is chosen, say d^*. A period of $2r+1$ consecutive days, centered around d^* is considered and only the observations falling in that period are selected to form the event. Within this period the station-wise maximum is identified and the collection of the station-wise maxima forms one event. Because there is some evidence that the time an extreme event takes to propagate through the seven nodes in our model is about 3–7 days, we choose $r = 3$, hence we consider that one event lasts 7 days. In this way we obtain 717 observations of supposedly independent events. As it was mentioned the results are robust to the choice of $r = \{1, 2, 3\}$.

We test for seasonality and trends each of the series (each having 717 observations). The season factor is significant across all series and the time trend is marginally significant for some of the locations. We used a simple time series model to remove these non-stationarities. The model is based on season indicators and a linear time trend

$$X_t = \beta_0 + \beta_1 \mathbb{1}_{\text{spring}_t} + \beta_2 \mathbb{1}_{\text{summer}_t} + \beta_3 \mathbb{1}_{\text{winter}_t} + \alpha t + \epsilon_t, \tag{24}$$

where ϵ_t for $t = 1, 2, \ldots$ is a stationary mean zero process. After fitting the model in (24) to each of the five series through ordinary least squares we obtain the residuals and use those in the estimation of the extremal dependence.

A.5 ECE-based confidence interval for the dependence parameters

Let $\hat{\theta}_{n,k} = \hat{\theta}^{\text{ECE}}_{n,k}$ denote the pairwise extremal coefficient estimator in (20) and let θ_0 denote the true vector of parameters. By Einmahl et al. (2018, Theorem 2) with Ω equal to the identity matrix, the ECE is asymptotically normal,

$$\sqrt{k}(\hat{\theta}_{n,k} - \theta_0) \xrightarrow{d} \mathcal{N}_{|E|}(0, M(\theta_0)), \qquad n \to \infty,$$

where $k = k_n \to \infty$ such that $k/n \to 0$ fast enough (Einmahl et al., 2012, Theorem 4.6) and the asymptotic covariance matrix is

$$M(\theta_0) = (\dot{L}^T \dot{L})^{-1} \dot{L}^T \Sigma_L \dot{L} (\dot{L}^T \dot{L})^{-1}.$$

The matrices $\dot{L}$ and Σ_L depend on θ_0 and are described below. For every k and every $e \in E$, an asymptotic 95% confidence interval for the edge parameter $\theta_{0,e}$ is given by

$$\theta_{0,e} \in [\hat{\theta}_{k,n;e} \pm 1.96\sqrt{M(\hat{\theta}_{k,n})_{ee}/k}].$$

The map $\theta \mapsto L(\theta)$ is introduced in Section 4.3 and equation (21) and $\dot{L}(\theta)$ is a $q \times |E|$ matrix of partial derivatives, where $q = |\mathcal{Q}|$ and $\mathcal{Q} \subseteq \{J \subseteq U : |J| = 2\}$ is the set of pairs on which the ECE is based. For a given edge index $e = (u, v)$, the partial derivative of $l(x^{(m)}; \theta)$ with respect to θ_e when $x^{(m)} = (\mathbb{1}_{i \in J}, i \in U)$ is given by

$$\frac{\partial l(x^{(m)}; \theta)}{\partial \theta_e} = \frac{\phi\left(\sqrt{p_{uv}}/2\right)}{\sqrt{p_{uv}}} \theta_e \mathbb{1}_{\{e \in (u \rightsquigarrow v)\}},$$

where p_{uv} is the path sum as in (15) and ϕ denotes the standard normal density function. The partial derivatives of $l(x^{(m)};\theta)$ with respect to θ_e for every $e \in E$ form a row vector in the matrix $\dot{L}(\theta)$.

The matrix $\Sigma_L(\theta_0)$ is a $q \times q$ covariance matrix from the asymptotic distribution of the empirical stable tail dependence function

$$\big\{\sqrt{k}\big(\hat{l}_{n,k}(x^{(m)}) - l(x^{(m)};\theta_0)\big)\big\}_{m=1,\dots,q} \xrightarrow{d} \mathcal{N}_q(0, \Sigma_L(\theta_0)), \qquad n \to \infty.$$

The ingredients of every element of the matrix $\Sigma_L(\theta_0)$ consist of the stdf evaluated at different coordinates and of the partial derivatives of the stdf $l(x;\theta)$ with respect to the elements of x. For details on the distribution and the covariance matrix we refer to Einmahl et al. (2018, Section 2.5). Here we provide the analytical expression of the partial derivatives. When $x^{(m)} = (\mathbb{1}_{i\in\{u,v\}}, i \in U)$ the two partial derivatives with respect to the x_u and x_v are identical and equal to

$$\dot{l}_u(x^{(m)};\theta) = \dot{l}_v(x^{(m)};\theta) = \frac{\partial l(x^{(m)};\theta)}{\partial x_v}\Big|_{(x_u,x_v)=(1,1)} = \Phi(\sqrt{p_{uv}}/2)\,.$$

A.6 Bootstrap confidence interval for the Pickands dependence function

For assessing the goodness-of-fit of the proposed model (Section 5.2), we would like to obtain non-parametric 95% confidence intervals for $A(w) = l(1-w, w)$ for $w \in [0,1]$. As shown in Kiriliouk et al. (2018, Section 5) this can be achieved by resampling from the empirical beta copula. For every fixed $w \in [0,1]$ we seek with $a(w), b(w)$,

$$\mathbb{P}\big(a(w) \le \hat{l}_{n,k}(1-w,w) - l(1-w,w) \le b(w)\big) = 0.95\,,$$

where $\hat{l}_{n,k}$ is the non-parametric estimator of the stdf. For a, b that satisfy the expression above, a point-wise confidence interval is given by

$$l \in [\hat{l}_{n,k} - b, \hat{l}_{n,k} - a]\,. \tag{25}$$

Let $(Z^*_{v,i})_{v\in U}$, for $i = 1,\dots,n$, be a random sample from the empirical beta copula drawn according to steps A1–A4 of Kiriliouk et al. (2018, Section 5). Let the function $\hat{l}^{\beta}_{n,k}(1-w,w)$ be the empirical beta stdf using the ranks $R_{v,i} = n\hat{F}_{v,n}(\xi_{v,i})$ of the original variables and let the function $\hat{l}^{*}_{n,k}(1-w,w)$ be the non-parametric estimate of the stdf using the ranks $R^*_{v,i} = n\hat{F}_{Z^*_v,n}(Z^*_{v,i})$ of the bootstrap sample.

We use the distribution of $\hat{l}^{*}_{n,k} - \hat{l}^{\beta}_{n,k}$ conditionally on the data as an estimate of the distribution of $\hat{l}_{n,k} - l$. Hence, we estimate $a(w)$ and $b(w)$ by $a^*(w)$ and $b^*(w)$ defined implicitly by

$$\begin{aligned} 0.95 &= \mathbb{P}^*\left(a^*(w) \le \hat{l}^{*}_{n,k}(1-w,w) - \hat{l}^{\beta}_{n,k}(1-w,w) \le b^*(w)\right) \\ &= \mathbb{P}^*\left(a + \hat{l}^{\beta}_{n,k}(1-w,w) \le \hat{l}^{*}_{n,k}(1-w,w) \le b + \hat{l}^{\beta}_{n,k}(1-w,w)\right). \end{aligned}$$

We further estimate the bootstrap distribution of $\hat{l}^{*}_{n,k}$ by a Monte Carlo approximation obtained by $N = 1000$ samples of size n from the empirical beta copula. As a consequence, the lower and upper bounds for $\hat{l}^{*}_{n,k}$ above are equated to the empirical 0.025- and 0.975-quantiles, respectively, yielding

$$\hat{l}^{*}_{0.025} = a^*(w) + \hat{l}^{\beta}_{n,k} \qquad\qquad \hat{l}^{*}_{0.975} = b^*(w) + \hat{l}^{\beta}_{n,k}\,.$$

Replacing a and b in (25) by a^* and b^*, we obtain the following bootstrapped confidence interval for l:

$$l \in \left[\hat{l}_{n,k} - \hat{l}^{*}_{0.975} + \hat{l}^{\beta}_{n,k}\,, \hat{l}_{n,k} - \hat{l}^{*}_{0.025} + \hat{l}^{\beta}_{n,k}\right].$$

References

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

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

de Haan, L. and Ferreira, A. (2007). *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering. Springer New York.

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

Einmahl, J., Krajina, A., and Segers, J. (2012). An m-estimator for tail dependence in arbitrary dimensions. *The annals of statistics*, 40(3):1764–1793.

Engelke, S. and Hitz, A. S. (2018). Graphical models for extremes. *ArXiv e-prints*. arXiv:1812.01734.

Engelke, S., Malinowski, A., Kabluchko, Z., and Schlather, M. (2015). Estimation of Hüsler–Reiss distributions and Brown–Resnick processes. *Journal of the Royal Statistical Society. Series B*, 77:239–265.

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

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

Hitz, A. and Evans, R. (2016). One-component regular variation and graphical modeling of extremes. *Journal of Applied Probability*, 53(3):733–746.

Huser, R. and Davison, A. C. (2013). Composite likelihood estimation for the Brown–Resnick process. *Biometrika*, 100(2):511–518.

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

Janssen, A. and Segers, J. (2014). Markov tail chains. *Journal of Applied Probability*, 51(4):1133–1153.

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

Klüppelberg, C. and Sönmez, E. (2018). Max-linear models on infinite graphs generated by Bernoulli bond percolation. *ArXiv e-prints*. arXiv:1804.06102.

Koller, D. and Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, Cambridge, Massachusetts.

Lauritzen, S. (1996). *Graphical Models*. Oxford University Press, Oxford.

Lee, D. and Joe, H. (2018). Multivariate extreme value copulas with factor and tree dependence structures. *Extremes*, 21:147–176.

Nikoloulopoulos, A. K., Joe, H., and Li, H. (2009). Extreme value properties of multivariate t copulas. *Extremes*, 12(2):129–148.

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

Perfekt, R. (1994). Extremal behaviour of stationary Markov chains with applications. *The Annals of Applied Probability*, 4(2):529–548.

Segers, J. (2007). Multivariate regular variation of heavy-tailed Markov chains. *ArXiv e-prints*. arXiv:math/0701411.

Segers, J. (2019). One- versus multi- component regular variation and extremes of Markov trees. *ArXiv e-prints*. arXiv:1902.02226.

Smith, R. L. (1992). The extremal index for a Markov chain. *Journal of Applied Probability*, 29(1):37–45.

Yu, H., Uy, W., and Dauwels, J. (2017). Modeling spatial extremes via ensemble-of-trees of pairwise copulas. *IEEE Transactions on Signal Processing*, 65(3):571–585.

Yun, S. (1998). The extremal index of a higher-order stationary Markov chain. *The Annals of Applied Probability*, 8(2):408–437.