

The tail dependograph

Cécile Mercadier, Olivier Roustant

▶ **To cite this version:**

Cécile Mercadier, Olivier Roustant. The tail dependograph. Extremes, 2019, 10.1007/s10687-019-00345-3 . hal-01649596v3

HAL Id: hal-01649596
https://hal.science/hal-01649596v3

Submitted on 7 Mar 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Noname manuscript No.
(will be inserted by the editor)

The Tail Dependograph

Cécile Mercadier · Olivier Roustant

Received: date / Accepted: date

Abstract All characterizations of non-degenerate multivariate tail dependence structures are both functional and infinite-dimensional. Taking advantage of the Hoeffding–Sobol decomposition, we derive new indices to measure and summarize the strength of dependence in a multivariate extreme value analysis. The tail superset importance coefficients provide a pairwise ordering of the asymptotic dependence structure. We then define the tail dependograph, which visually ranks the extremal dependence between the components of the random vector of interest. For the purpose of inference, a rank-based statistic is derived and its asymptotic behavior is stated. These new concepts are illustrated with both theoretical models and real data, showing that our methodology performs well in practice.

Keywords Global sensitivity analysis · Hoeffding–Sobol decomposition · Multivariate extreme value analysis · Pairwise index · Tail dependency graph

Mathematics Subject Classification (2000) Primary 62J10 · 62G32 · 62H20 · Secondary 62J15 · 62G20 · 62-09

1 Introduction

This paper is at the crossroads of two domains: sensitivity analysis and multivariate extreme values. This combination makes it possible to find variables or groups of asymptotically dependent

Cécile Mercadier
Univ Lyon, Université Claude Bernard Lyon 1,
CNRS UMR 5208, Institut Camille Jordan,
43 blvd. du 11 novembre 1918,
F-69622 Villeurbanne, France
E-mail: mercadier@math.univ-lyon1.fr

Mines Saint-Étienne,
UMR CNRS 6158, LIMOS,
F-42023 Saint-Etienne, France
E-mail: roustant@emse.fr

variables, and to illustrate the asymptotic dependence structure with the use of non-oriented graphs of dependence.

It is already possible to characterize the dependence structure using a finite number of parameters. Since the original idea of Tiago de Oliveira (1962/63) in the bivariate case and the name proposed by Smith (1990), many authors have studied or just used the extremal coefficients. In dimension d, we have $2^d - d - 1$ coefficients that, for a given subset of $\{1, \ldots, d\}$, might be interpreted as the number of asymptotically independent components among this subset. The aim of this paper is to build a second family whose pairwise version is of size $d(d-1)/2$. It is not possible to say which one is preferable to the other since they are partial extractions of an infinite-dimensional function. Nevertheless, the illustrations in this paper make it possible to show situations where the second family of proposed indices provides greater nuances.

It would be untrue to say that this is the first time that the terms sensitivity analysis and extreme values can be found together in a research paper. We can mention, for example, the section Extreme values (Section 7) of Liu and Owen (2006). Building on the decomposition of the variance of a well-chosen function and the notion of mean dimension, they show that the laws of extremes cannot be Gaussian. Their work focuses on extreme values in a univariate framework and does therefore not deal with dependence in extremes.
More generally, a sensitivity analysis is the search for the input into a system that will have the greatest impact on the systems output, often extremely complex. This search can be applied to subsets of components whose variations will most significantly impact variations in output. In other words, given a pre-established size of subset, which group of input parameters will have the most impact on the outcome? This question, which is generally addressed using global variances, is in fact linked to the extreme value theory if the indicator observed becomes a high quantile or a probability of failure. We refer to the works of Owen et al. (2014), Kucherenko and Song (2016), Sueur et al. (2017), Browne et al. (2017) and Maume-Deschamps and Niang (2018).

In the study presented here, there is no output, only inputs: a random vector of interest, denoted by $\mathbf{X} = (X_1, \ldots, X_d)$. For example, in a financial context, Peng and Ng (2012) investigated the contagion between stock indices across markets. From a global viewpoint, none of the stock markets is favored. The aim is to classify the degrees of asymptotic dependence between these economic institutions. Another example is storms, which are characterized by the behavior of high surge, low pressure, strong wind, storm size, angle to coast, etc. One goal could be to compare the strength of tail dependence between surge, wind, water level and pressure. Once again, no variable is favored by designating it as the output of a complex system whose inputs would be the other variables mentioned. See Cai et al. (2013), Kereszturi et al. (2016) or Mhalla et al. (2017) for such environmental considerations.

Since there is no observed output, we construct it artificially but with mathematical significance. We apply global sensitivity analysis to the multivariate function that models the dependence structure for multivariate extreme values, which, if unknown, has an empirical tractable version. Indeed, under multivariate max-stable modeling, various ways to describe the extremal dependence structure of $\mathbf{X}$ exist: extreme value copula, dependence functions, intensity measure, spectral measure, etc, which obviously have strong connections. We refer to Fougères (2004) and Segers (2012) for a comprehensive description of these notions. The focus here is on the stable tail dependence function. See Ressel (2013) for its complete characterization. It has some nice properties: homogeneity of order 1, convexity, identity on the axis, and

tractable upper and lower bounds for independence and complete positive dependence respectively. The restriction to the stable tail dependence function is imposed by the fact that tail independence induces additivity properties to it. In fact, sensitivity analysis detects additivity in the responses structure with respect to several inputs.

The main goal of this paper is to introduce new indices, called *tail superset importance coefficients*, for this type of multivariate extreme value analysis. Let ℓ stand for the stable tail dependence function of $\mathbf{X}$. In our setting, we only have natural entries represented by the d variables X_i jointly studied but there is no response. The idea is to define a pseudo-response such as $\ell(\mathbf{U})$ at some generic random vector $\mathbf{U}$, with the intention of ranking dependence in the asymptotic joint structure. Roughly speaking, our indices are derived from the decomposition of the variance of $\ell(\mathbf{U})$. Their definition starts from the Hoeffding–Sobol decomposition (of the function ℓ) that is a very well-known tool in computer experiments. See Hoeffding (1948) for the pioneer construction and Sobol′ (1993) for its practical efficient use. The new indices make it possible to decompose the tail dependence structure but, above all, rank the pairs of components (or subsets with fixed size) in terms of their global contribution to tail dependence.

Furthermore, the superset importance coefficients of Hooker (2004) and Liu and Owen (2006) are not the only tool of sensitivity analysis that we try to transpose to an extreme value setting. We also want to highlight the features of the FANOVA graph introduced by Muehlenstaedt et al. (2012). Applied to the stable tail dependence function, we refer to it as the *tail dependograph* since it graphically represents, on an undirected but weighted graph, the strength of asymptotic dependence within a vector. On this graph, the edge thickness between two vertices is proportional to the force of tail dependence in the corresponding bivariate model but also cumulates the information of the asymptotic dependence structure for any multivariate model that contains this pair.

If instead of applying the Hoeffding–Sobol decomposition on the true (and most of the time unknown) stable tail dependence function, its estimation $\hat{\ell}$ is used, we have a decomposition of the estimated variance of $\ell(\mathbf{U})$. As a consequence, we obtain an estimation of the tail superset importance coefficients and, obviously, an estimate of the tail dependograph. The empirical estimator of the stable tail dependence function, first proposed by Huang (1992), leads to explicit rank-based estimators of the tail superset importance coefficients. Under a second-order condition, similar to the setting assumed in Fougères et al. (2015), and taking advantage of the asymptotic distributional expansion of the empirical estimator of ℓ, we state the asymptotic behavior of the empirical tail superset importance coefficients. Normalized versions of the new indices are also introduced and consistent rank-based estimators are provided.

The remainder of the article is organized as follows: In Section 2, we review some tools from global sensitivity analysis. Then, in Section 3, we derive new coefficients and dependency graphs for extreme value analysis. In Section 4, the inference is done through a non-parametric approach. The asymptotic behavior of these empirical indices is also explored at the end of the section. We provide a broad numerical illustration of the new concepts in Section 5. The paper ends with some concluding remarks and the proofs are presented in a dedicated section.

2 Background on global sensitivity analysis

Global Sensitivity Analysis is a branch of control theory and computer sciences. This type of analysis measures how sensitive the outcome quantity $f(\mathbf{u})$ is to the variation of individual

input variables u_i. A variance decomposition determines which of the multiple input parameters are responsible for most of the variation in the outcome. This section presents an overview of functional decomposition, computational issues and consequences in terms of variance.

2.1 The Hoeffding–Sobol Decomposition

Let $d \geq 2$ be an integer and let $f : \mathbb{R}^d \to \mathbb{R}$. The Hoeffding–Sobol decomposition aims at representing $f(\mathbf{u})$ as a sum of terms of increasing complexity. The name of this decomposition arises, first, from its creation by Hoeffding (1948) in the context of U-statistics and, second, from its use for sensitivity analysis by Sobol′ (1993).

Theorem 1 (Hoeffding–Sobol Decomposition) *Let $\mathbf{U} = (U_1, \ldots, U_d)$ be a random vector with independent margins and distribution denoted by $\mu = \mu_1 \otimes \ldots \mu_d$. For any $\mathbf{u} \in \mathbb{R}^d$, and any subset $I \subseteq \{1, \ldots, d\}$, $\mathbf{u}_I$ concatenates the components of $\mathbf{u}$ whose indices are included in I. Assume that $f \in L^2(\mu)$. Then there exists a unique decomposition of the form:*

$$f(u_1, \ldots, u_d) = f_\emptyset + \sum_{i=1}^{d} f_i(u_i) + \sum_{i<j} f_{i,j}(u_i, u_j) + \ldots + f_{1,\ldots,d}(u_1, \ldots, u_d) \tag{1}$$

where the $f_I : \mathbb{R}^{|I|} \to \mathbb{R}$ satisfy

$$\mathbb{E}[f_I(\mathbf{U}_I)|\mathbf{U}_{I'}] = 0 \qquad \text{when } I' = \emptyset \text{ or } I' \subsetneq I\ . \tag{2}$$

Furthermore, the f_I's are given explicitly by:

$$f_I(\mathbf{u}_I) = \mathbb{E}[f(\mathbf{U})|\mathbf{U}_I = \mathbf{u}_I] - \sum_{I' \subsetneq I} f_{I'}(\mathbf{u}_{I'})\ .$$

In Theorem 1, we followed the presentation of Efron and Stein (1981) and we refer to their paper for a proof.
The f_I's are known as ANOVA terms. They can be interpreted as follows. The constant $f_\emptyset = \mathbb{E}[f(\mathbf{U})]$ is the global (unconditional) mean, $f_i(u_i) = \mathbb{E}[f(\mathbf{U})|U_i = u_i] - f_\emptyset$ represents the main effect of component $\{i\}$, $f_{ij}(u_i, u_j) = \mathbb{E}[f(\mathbf{U})|U_i = u_i, U_j = u_j] - f_i(u_i) - f_j(u_j) - f_\emptyset$ captures the second-order interaction from the pair of components $\{i, j\}$, and so on. In the previous formulae, the mathematical expectation refers to the integral with respect to $d\mu$ in $f_\emptyset$, to $\otimes_{t\neq i} d\mu_t$ in the definition of f_i, and to $\otimes_{t\neq i,j} d\mu_t$ in that of f_{ij}.

Even if $f(\mathbf{u})$ involves all variables $u_1, \ldots, u_d$ in its definition, it is not generally equal to the highest order interaction terms $f_{1,\ldots,d}(\mathbf{u})$ of the Hoeffding–Sobol decomposition. Indeed, it does not satisfy the non-simplification conditions $\mathbb{E}(f(U_1, \ldots, U_d)|U_I) = 0$ for all strict subsets I of $\{1, \ldots, d\}$. Consider, for instance, $f(u_1, \ldots, u_d) = \prod_{i=1}^d u_i$ and $U_1, \ldots, U_d$ being i.i.d. uniform on $[0, 1]$. It is then easy to check that the Hoeffding–Sobol decomposition of f is obtained by expanding the product $\prod_{i=1}^d ((u_i - 0.5) + 0.5)$. All the terms $f_I(u_I) = \prod_{i\in I}(u_i - 0.5)0.5^{d-|I|}$ satisfy the non-simplification conditions, and we can conclude by unicity. In particular, $f_{1,\ldots,d}(u_1, \ldots, u_d) = \prod_{i=1,\ldots,d}(u_i - 0.5)$.

The reader is also referred to van der Vaart (1998; Section 11.4). More recently, an extension to dependent entries was obtained by Chastaing et al. (2012) under specific conditions. Before going further, it is important to avoid any possible confusion. Here, we only refer to the pioneer setting where the distribution of $\mathbf{U}$ is a product measure. On the contrary, all the dependences that we discuss in this paper will be stored in the function ℓ on which we will focus as of Section 3.1.

2.2 A variance decomposition

The Hoeffding–Sobol decomposition also has the advantage of a variance decomposition

$$\operatorname{var}(f(\mathbf{U})) = \sum_{I \subseteq \{1,\dots,d\}} \operatorname{var}(f_I(\mathbf{U}_I)) \, .$$

Indeed, the property (2), which means that the orthogonal projection of the ANOVA terms f_I onto strict subspaces is zero, implies that they are non correlated (see e.g. Efron and Stein (1981)). The variance decomposition is generally referred to with the following notation

$$D(f) = \sum_{I \subseteq \{1,\dots,d\}} D_I(f) \tag{3}$$

where $D(f)$ denotes the global variance $\operatorname{var}(f(\mathbf{U}))$ and $D_I(f)$ the variance corresponding to the subset I, i.e., $\operatorname{var}(f_I(\mathbf{U}_I))$ or the variance of the term f_I. For instance, $D_{\{i\}}(f)$ is the variance of $f(\mathbf{U})$ due to the ith component only. Similarly, $D_{\{i,j\}}(f)$ is the variance due to the combined effect of components $\{i, j\}$.

One combination of such variances is of prime interest. It is defined by

$$D^{\text{super}}_{\{i,j\}}(f) = \sum_{I \supseteq \{i,j\}} D_I(f)$$

and referred to as the superset importance of the pair $\{i, j\}$. See Hooker (2004) and Liu and Owen (2006) for further discussion. This coefficient is positive and smaller than or equal to $D(f)$. This index makes it possible to discover additive structures in multivariate functions. Indeed, under continuity assumptions on μ and f, $D^{\text{super}}_{\{i,j\}}(f) = 0$ implies that f does not simultaneously depend on i and j in its ANOVA decomposition and, thus, f is additive with respect to u_i and u_j. Taking the decomposition of the variance $D(f)$ into account, scaled versions were introduced by Sobol′ (1993). The superset sensitivity indices that belong to the interval $[0, 1]$ are given by $S^{\text{super}}_{\{i,j\}}(f) = D^{\text{super}}_{\{i,j\}}(f)/D(f)$.

2.3 On computational aspects

When the function of interest f is known, the Hoeffding–Sobol decomposition (1) is obviously not necessarily derived. Let us introduce some notation. Let $(\mathbf{e}_1, \dots, \mathbf{e}_d)$ stand for the canonical basis of $\mathbb{R}^d$. Both vectors $\mathbf{u} + (v - u_i)\mathbf{e}_i$ and $\mathbf{u} + (v - u_i)\mathbf{e}_i + (w - u_j)\mathbf{e}_j$ are obtained from $\mathbf{u}$, with u_i replaced by v and, additionally, in the second case, with u_j replaced by w. We denote

the corresponding evaluations of f at these points by $T_i[f](\mathbf{u}, v) := f(\mathbf{u} + (v - u_i)\mathbf{e}_i)$, and $T_{i,j}[f](\mathbf{u}, v, w) := f(\mathbf{u} + (v - u_i)\mathbf{e}_i + (w - u_j)\mathbf{e}_j)$. Then, from Liu and Owen (2006), we have

$$D^{\text{super}}_{\{i,j\}}(f) = \frac{1}{4}\int_{[0,1]^{d+2}} \{f(\mathbf{u}) - T_i[f](\mathbf{u}, v) - T_j[f](\mathbf{u}, w) + T_{i,j}[f](\mathbf{u}, v, w)\}^2 \, d\mu(\mathbf{u}) d\mu_i(v) d\mu_j(w) \, . \quad (4)$$

Equality (4) is powerful. On the one hand, it provides information that is not just limited to $\{i, j\}$ but that is linked to all subsets I from $\{1, \ldots, d\}$ that contain this pair. On the other hand, it does not require the knowledge of all subfunctions f_I induced in the initial definition of $D^{\text{super}}_{\{i,j\}}(f)$. In other words, this integral representation makes it possible to compute

$$\sum_{I \supseteq \{i,j\}} \text{var}(f_I(\mathbf{U}_I))$$

without identifying the terms of the decomposition $\{f_I, I \supseteq \{i, j\}\}$.

Again, even if the superset importance coefficient associated with any pair has an integral formula, it is not easy to compute explicitly. Hopefully, Formula (4) admits Monte Carlo approximations as estimates. It can be written as

$$\widehat{D}^{\text{super}}_{\{i,j\}}(f) = \frac{1}{4N}\sum_{s=1}^{N} \left(f(\mathbf{u}^{[s]}) - T_i[f](\mathbf{u}^{[s]}, v^{[s]}) - T_j[f](\mathbf{u}^{[s]}, w^{[s]}) + T_{i,j}[f](\mathbf{u}^{[s]}, v^{[s]}, w^{[s]})\right)^2 \quad (5)$$

where all $\mathbf{u}^{[s]}, v^{[s]}, w^{[s]}$ are i.i.d. random vectors or variables drawn from $\otimes_{t=1,\ldots,d} d\mu_t$, $d\mu_i$ and $d\mu_j$ respectively. Its statistical properties have been investigated in Fruth et al. (2014): it is unbiased and asymptotically normal when the true value is not zero. More interestingly, it is asymptotically efficient in a class of models with exchangeable variables, indicating that it has the smallest within-class variance. Furthermore, $\widehat{D}^{\text{super}}_{\{i,j\}}(f)$ vanishes when the theoretical $D^{\text{super}}_{\{i,j\}}(f)$ is zero: the Monte Carlo error of estimation under this specific case is always zero, which is remarkable.

Finally, the scaled version $S^{\text{super}}_{\{i,j\}}(f)$ can also be approximated with Monte Carlo replications by dividing $\widehat{D}^{\text{super}}_{\{i,j\}}(f)$ by a common estimator of the overall variance $D(f)$.

3 New measures for extreme value analysis

The main idea of this paper is to apply the above to a functional characterization of the tail dependence structure. We begin this section with some background on the stable tail dependence function, which is additive under tail independence. We then derive both new coefficients and a new graph for multivariate extreme value analysis. A small numerical illustration is given at the end of this section.

3.1 The stable tail dependence function

Let $\mathbf{X} = (X_1, \ldots, X_d)$ be a multivariate vector with continuous marginal cumulative distribution functions denoted by $F_1, \ldots, F_d$. When it exists, the stable tail dependence function (s.t.d.f.) is defined at the point $\mathbf{u} = (u_1, \ldots, u_d)$ by

$$\ell(\mathbf{u}) = \lim_{z \to \infty} z \left\{1 - F\left(F_1^{-1}(1 - u_1/z), \ldots, F_d^{-1}(1 - u_d/z)\right)\right\} \, . \quad (6)$$

This function furnishes a full characterization of the extremal dependence between the components of $\mathbf{X}$. It is not the only measure to model the dependence between extreme components. As such, it is regularly used in the literature in a modeling bid, for instance, with parametric construction or for inference concerns. Assuming both Equation (6) and that each marginal F_t (for $t = 1, \ldots, d$) belongs to the domain of attraction of the Generalized Extreme Value distribution with parameters (μ_t, σ_t, ξ_t) is equivalent to assuming that F is in the domain of attraction of the distribution

$$H(\mathbf{u}) = \exp\left(-\ell\left(-\log H_1(u_1), \ldots, -\log H_d(u_d)\right)\right) \tag{7}$$

with

$$H_t(u) = \exp\left(-\left(1 + \xi_t \frac{u - \mu_t}{\sigma_t}\right)^{-1/\xi_t}\right) .$$

For more details, we refer to de Haan and Ferreira (2006; Theorem 6.2.1, Part 2 combined with convergence (6.2.1)) for example.
Necessary and sufficient conditions on ℓ to be a s.t.d.f. can be found in Ressel (2013). Only some tractable properties will be needed here. The s.t.d.f. is continuous, convex and, as already said, homogeneous of order 1. Furthermore, taking Equation (6), for example, we obtain

$$u_1 \vee \ldots \vee u_d \leq \ell(\mathbf{u}) \leq u_1 + \ldots + u_d \tag{8}$$

corresponding to *asymptotic full dependence* (lower bound, *denoted AD*) and *asymptotic independence* (upper bound, *denoted AI*) between all components.

3.2 The tail superset importance coefficients

In the present setting, the random vector $\mathbf{U}$ from Section 2 is an abstract construction. We will further introduce the following assumption (H_μ) *The measure μ is continuous with support* $[0,1]^d$, where $\mu = \mu_1 \otimes \ldots \otimes \mu_d$ refers to the distribution of $\mathbf{U}$. First, the requirement on the support makes it possible to substitute the function f by ℓ in Equation (1) or (3). By homogeneity of ℓ, the s.t.d.f. can be restricted to $[0,1]^d$ without loss of information. Now, by continuity, ℓ is square integrable on $[0,1]^d$. Note that the continuity is also useful in this paper to deduce that a term of variance zero is null.

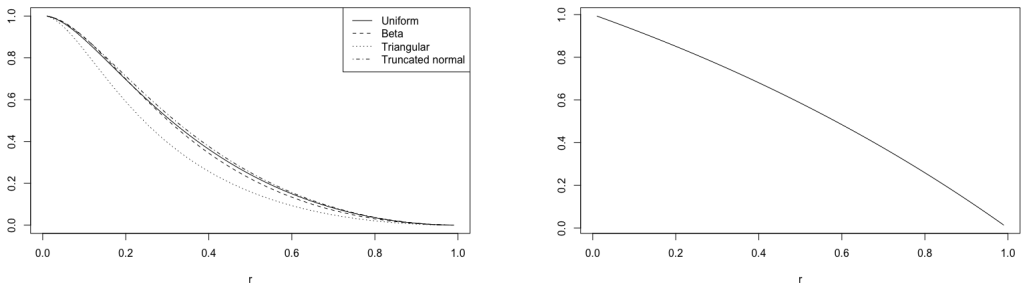
Definition 1 (Tail superset importance coefficient) A tail superset importance coefficient is a superset importance index $D^{\text{super}}_{\{i,j\}}(\ell)$ where ℓ is a stable tail dependence function and where μ satisfies (H_μ). Its *scaled* version is $S^{\text{super}}_{\{i,j\}}(\ell)$.

Even though it is associated with a random vector $\mathbf{X}$, function ℓ is a deterministic function. Because ℓ is generally unknown, we provide here a new modeling tool. Under AI between all components (right hand side of (8)), for instance, all tail superset importance coefficients vanish. However, these indices are also helpful in less exaggerated situations. By looking at their sorted values, it is possible to trace the pair with the strongest asymptotic dependence up to the one with the weakest (in the superset sense, not in the bivariate one). These coefficients can be directly compared and applied for dimension reduction purposes.

To illustrate Definition 1, let us put it in the simple context of dimension 2 and the standard model

$$\ell_r(u_1, u_2) = (u_1^{1/r} + u_2^{1/r})^r \tag{9}$$

with $r \in (0, 1]$, known as the symmetric logistic model of Gumbel (1960). The limit as r tends to zero leads to AD, whereas AI corresponds to $r = 1$. The values of $D_{12}^{\text{super}}(\ell_r)/D_{12}^{\text{super}}(\ell_0)$ are plotted under different distributions μ on Figure 1(a). We should compare the new indices to the well-known extremal coefficients, given as $\ell_r(1, 1) = 2^r$ under (9). We represent $2 - \ell_r(1, 1)$ on Figure 1(b). Both scales are chosen to produce comparable information.



(a) (b)

Fig. 1: (a) Illustration of Definition 1 in the two-dimensional framework (9) where $D_{12}^{\text{super}}(\ell_r)/D_{12}^{\text{super}}(\ell_0)$ is plotted with respect to the coefficient r. Four types of product measure μ are used based either on Uniform, Beta, Triangular or Truncated normal. (b) Values of $2 - \ell_r(1, 1)$ derived from the classical extremal coefficient under the same model (9).

Our aim here is not to rank the coefficient $D_{12}^{\text{super}}(\ell)$ versus $2 - \ell(1, 1)$. Their values are not the same, but through many two-dimensional practical examples, they always order strengths of asymptotic dependence in a concordant way. On the other hand, the information is more dissonant in a higher dimension, as illustrated in Sections 3.5 and 5.1. Insofar as the two families are pairwise orderings of the asymptotic dependence, an infinite-dimensional function, the two families of indices cannot replace each other but are complementary.

3.3 The tail dependograph

We now introduce a new weighted graph whose weights are directly given by the tail superset importance coefficients. An *undirected and weighted graph* is a list $G = (V, \mathbf{g})$ of the set of vertices $V = \{1, \ldots, d\}$ and the collection $\mathbf{g} = \{g_{1;2}, g_{1;3}, \ldots, g_{1;d}, \ldots, g_{d-1;d}\}$ of the $d(d-1)/2$ edge weights. Graphically, we represent the vertex number inside a bubble and plot a segment between $\{i\}$ and $\{j\}$ whose width varies proportionally to $g_{i;j}$. For a *dependency graph* of a random vector $\mathbf{X} = (X_1, \ldots, X_d)$, V represents the components of $\mathbf{X}$ and $\mathbf{g}$ contains some positive dependence measures between pairs.

When the weight of an edge is close to zero, it indicates some independence. Moreover, ordered values of edge weights reflect ordered strength of dependence between the associated

pairs. In this paper, we adopt the graph representation given in the R `sensitivity` package (Pujol et al., 2017): the weights are not printed along edges but a visual indication is given through the edge thickness.

We now introduce a new dependency graph to explore tail dependence.

Definition 2 (Tail dependency graph) An undirected and weighted graph G will be known as a *tail dependency graph* if the following holds: the vectors $\mathbf{X}_A = \{X_a, a \in A\}$ and $\mathbf{X}_B = \{X_a, a \in B\}$ are asymptotically independent when there is no edge from any vertex in A to any vertex in B, for any two subsets of vertices A and B from V: $g_{i;j} = 0$, $\forall i \in A, \forall j \in B$.

The new indices, namely the tail superset importance coefficients of Section 3.2, are suitable measures to play the role of edge weights of tail dependency graphs.

Definition 3 (Tail dependograph) The *tail dependograph* is a tail dependency graph with edge weights given by $D^{\text{super}}_{\{i,j\}}(\ell)$, namely the tail superset importance coefficient from Definition 1. Similarly, the weights for the *scaled tail dependograph* are given by the tail sensitivity indices $S^{\text{super}}_{\{i,j\}}(\ell)$.

Note that the tail dependograph corresponds to a FANOVA graph, introduced in Muehlenstaedt et al. (2012), when applied to the s.t.d.f. ℓ.

Proposition 1 *Tail independence involved in Definition 2 and 3 is concordant with the intuitive meaning. Let A, B and (possibly empty) C be a partition of $\{1, \ldots, d\}$. The random vectors $\mathbf{X}_A$ and $\mathbf{X}_B$ are asymptotically independent if*

$$\ell(\mathbf{u}_A, \mathbf{u}_B, \mathbf{0}_C) = \ell(\mathbf{u}_A, \mathbf{0}_B, \mathbf{0}_C) + \ell(\mathbf{0}_A, \mathbf{u}_B, \mathbf{0}_C) \quad \forall \mathbf{u}_A, \mathbf{u}_B \ .$$

We postpone all the proofs of the paper until the Proof Section. Note that the link between additivity of ℓ and the product form (two terms here) of the asymptotic distribution H can be derived from (7). This can be generalized to the asymptotic independence of any K groups of components ; the s.t.d.f. ℓ becomes a sum of K terms and the limiting extreme distribution H is then a product of K terms of separate variables.

Section 3.5 provides two examples of tail dependographs when $d = 3$ and $d = 4$. More illustrations for $d = 4, 10$ and 21 are given in Sections 5.1 and 5.3.

3.4 Two competitors for tail graphical summaries

To provide the reader with a better understanding of the interest of the tail dependograph, it is useful to compare it with two other graphs that rely on well-known summary measures for tail dependence, namely the extremal coefficients. For a s.t.d.f. ℓ, the extremal coefficient associated with subset I from $\{1, \ldots, d\}$ is given as $\theta_I(\ell) := \ell\left(\sum_{i \in I} \mathbf{e}_i\right)$. On the basis of Equation (8), it should be kept in mind that

$$1 \leq \theta_I(\ell) \leq |I| \ , \tag{10}$$

where the lower (resp. upper) bound corresponds to AD (resp. AI).

The λ-*graph* is a tail dependency graph with the edge weights given by

$$\lambda_{\{i,j\}} := 2 - \theta_{\{i,j\}}(\ell) = 2 - \ell(\mathbf{e}_i + \mathbf{e}_j) \ .$$

Using Equation (10) all the weights lie in the interval $[0,1]$, ranging from AI to AD. In particular they are positive and satisfy the tail dependency graph property. By extension to supersets of a pair, we define *the* λ-*supergraph* as the tail dependency graph G with weights

$$\lambda^{\text{super}}_{\{i,j\}} := \sum_{I \supseteq \{i,j\}} \{|I| - \theta_I(\ell)\} = (d+2)2^{d-3} - \sum_{I \supseteq \{i,j\}} \ell\left(\sum_{i \in I} \mathbf{e}_i\right) \ .$$

Similarly to the λ-graph, (10) ensures that the weights are valid. They range from 0 (AI) to $d2^{d-3}$ (AD), and could be easily normalized to range within $[0,1]$. The value $(d+2)2^{d-3}$ as well as the upper bound $d2^{d-3}$ have been derived from standard combinatorial results, with details given in the proof section.

3.5 Illustrating the tail dependograph on the asymmetric logistic model

Among the parametric subfamilies of s.t.d.f., one of the well-known forms is the asymmetric logistic model. Introduced in full generality by Tawn (1990), we have

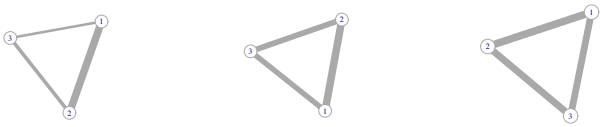
$$\ell(\mathbf{u}) = \sum_{b \in B} \left(\sum_{i \in b} (\beta_{i,b} u_i)^{1/\alpha_b} \right)^{\alpha_b}$$

where B is the set of all non-empty subsets of $\{1,\dots,d\}$, the dependence parameters α_b lie in $(0,1]$ and the asymmetric collection of weights $\beta_{i,b}$ are coefficients from $[0,1]$ satisfying $\forall i \in \{1,\dots,d\}$, $\sum_{b \in B : b \ni i} \beta_{i,b} = 1$. For standard Fréchet univariate margins, the simulation of vectors with such multivariate dependence is included in the R `evd` package of Stephenson (2002) and detailed in Stephenson (2003).

Let us consider two examples extracted from the help file of the function `rmvevd` of the `evd` package. For the dimension $d = 3$, let s.t.d.f. (11) be the asymmetric logistic model with the following family of coefficients

$$\begin{aligned} (\beta_{1,\{1\}}, \beta_{2,\{2\}}, \beta_{3,\{3\}}) = (.4,.1,.6) \quad (\beta_{1,\{12\}}, \beta_{2,\{12\}}) = (.3,.2) \text{ with } \alpha_{\{12\}} = .6, \\ (\beta_{1,\{13\}}, \beta_{3,\{13\}}) = (.1,.1) \text{ with } \alpha_{\{13\}} = .5, (\beta_{2,\{23\}}, \beta_{3,\{23\}}) = (.4,.1) \text{ with } \alpha_{\{23\}} = .8, \\ \text{and } (\beta_{1,\{123\}}, \beta_{2,\{123\}}, \beta_{3,\{123\}}) = (.2,.3,.2) \text{ with } \alpha_{\{123\}} = .3. \end{aligned} \tag{11}$$

We provide a graph representation below of the s.t.d.f. through the tail superset indices in Figure 2(a), the λ-graph in (b) and the λ-supergraph in (c).
Corresponding numerical values are available in Table 1.
The tail dependograph points out a strict ordering between the three pairs thanks to tail superset indices, i.e., not visible by looking at the other two graphs. Focusing on the λ-graph, it is possible to isolate the pair $\{1,2\}$ from the other two. Looking at the λ-supergraph, we could conclude that the three variables are exchangeable with respect to this criterion. Of



(*a*) (*b*) (*c*)

Fig. 2: The tail dependograph (*a*), the λ-graph (*b*), and the λ-supergraph (*c*) for s.t.d.f. (11).

Pair $\{i,j\}$	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$
$D^{\text{super}}_{\{i,j\}}(\ell)$	0.00040	0.00014	0.00019
$\lambda_{\{i,j\}}(\ell)$	0.2946	0.2123	0.2229
$\lambda^{\text{super}}_{\{i,j\}}(\ell)$	0.8736	0.7913	0.8019

Table 1: Values of theoretical coefficients for ℓ being the s.t.d.f. (11).

course, the λ-graph remains interesting since it relies on coefficients θ_I that are interpretable through the equality

$$\mathbb{P}\left(X_t \leq F_t^{-1}(p), \text{ for all } t \in I\right) = p^{\theta_I(\ell)}\ , \tag{12}$$

written above when F, the distribution of $\mathbf{X}$, is a max-stable distribution. With the tail superset coefficients, we lose this kind of interpretation. Instead, a global variance $D(\ell)$ is a reference to which each $D^{\text{super}}_{\{i,j\}}(\ell)$ can be compared so that they also can be compared to one another directly.

4 Nonparametric statistical inference

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be n independent copies of the d-variate vector $\mathbf{X}$. For inference purposes, we change the notation: we use (t) as an exponent to refer to the *t-th* component of a random vector. For $t = 1, \ldots, d$ and $s = 1, \ldots, n$, we denote the *t-th* component of $\mathbf{X}_s$ by $X_s^{(t)}$. Let $k \leq n$ be an integer. In extreme value theory, k represents the number of largest values considered in inference procedures. When k is too large, the estimation relies on some observations that are not extreme, introducing some bias. On the contrary, when k is too small, there is no bias but the variance of the estimator is large. Nice asymptotic properties are based on such a balance. In this section we derive non-parametric estimators of the main indices in this paper: the coefficients $D^{\text{super}}_{\{i,j\}}(\ell)$ in the first part, and the whole Hoeffding–Sobol decomposition in the second one. The particular case of uniform margins for $\mathbf{U}$ is then dealt with. Finally, asymptotic results are addressed at the end of the section.

4.1 Empirical versions of the tail superset importance coefficients

Taking Section 2.3 into account, once an estimator $\hat{\ell}$ for ℓ is given, the computation of a tail superset importance coefficient is then possible by substituting f with $\hat{\ell}$ in Formula (5). Consequently, the estimating issue could already appear to be settled. More interestingly, exact expressions can be derived for new indices $D_{\{i,j\}}(\hat{\ell})$ when the empirical counterpart of the true function ℓ is used.

Recall first the natural estimation of the stable tail dependence function, given by Huang (1992)

$$\hat{\ell}_{k,n}(\mathbf{u}) = \frac{1}{k}\sum_{s=1}^{n} \mathbf{1}\Big\{X_s^{(1)} \geq X^{(1)}_{n-[ku_1]+1,n} \text{ or } \ldots \text{ or } X_s^{(d)} \geq X^{(d)}_{n-[ku_d]+1,n}\Big\},$$

where z has been replaced by n/k in (6), F by its empirical version, and $F_t^{-1}(1-z^{-1}u_t)$ for $t=1,\ldots,d$ by $X^{(t)}_{n-[nz^{-1}u_t]+1,n}$. This estimator can be written as

$$\hat{\ell}_{k,n}(\mathbf{u}) = \frac{1}{k}\sum_{s=1}^{n} \mathbf{1}\Big\{u_1 \geq \tilde{R}_s^{(1)} \text{ or } \ldots \text{ or } u_d \geq \tilde{R}_s^{(d)}\Big\}, \tag{13}$$

in terms of $\tilde{R}_s^{(t)} := (n - R_s^{(t)} + 1)/k$ where $R_s^{(t)}$ denotes the rank of $X_s^{(t)}$ among $X_1^{(t)},\ldots,X_n^{(t)}$.

We now give the expression of the empirical version of the tail superset importance coefficients in terms of the rank observations. With a slight abuse of notation, we simply write $\mu_t(u) = \mu_t([0,u])\ \big(= \int_0^u d\mu_t(x)\big)$ for any $u \in (0,1)$.

Proposition 2 *The empirical tail superset importance coefficient can be expressed as*

$$D^{super}_{\{i,j\}}(\hat{\ell}_{k,n}) = \frac{1}{k^2}\sum_{s=1}^{n}\sum_{s'=1}^{n}\left\{\prod_{t\in\{i,j\}}\left(\mu_t\left(\tilde{R}_s^{(t)}\wedge\tilde{R}_{s'}^{(t)}\right) - \mu_t\left(\tilde{R}_s^{(t)}\right)\mu_t\left(\tilde{R}_{s'}^{(t)}\right)\right)\prod_{t\notin\{i,j\}}\mu_t\left(\tilde{R}_s^{(t)}\wedge\tilde{R}_{s'}^{(t)}\right)\right\}.$$

The empirical scaled tail superset importance coefficient is estimated by $\hat{S}^{super}_{k,n;\{i,j\}} := \dfrac{D^{super}_{\{i,j\}}(\hat{\ell}_{k,n})}{D(\hat{\ell}_{k,n})}$ *where*

$$D(\hat{\ell}_{k,n}) = \frac{1}{k^2}\sum_{s=1}^{n}\sum_{s'=1}^{n}\left\{\prod_{t=1}^{d}\mu_t\left(\tilde{R}_s^{(t)}\wedge\tilde{R}_{s'}^{(t)}\right) - \prod_{t=1}^{d}\mu_t\left(\tilde{R}_s^{(t)}\right)\mu_t\left(\tilde{R}_{s'}^{(t)}\right)\right\}.$$

Proposition 2 provides accurate versions of these coefficients. These rank-based expressions avoid the additional Monte Carlo error estimation that we obtain when replacing f by the empirical estimator (13) in Equation (5).

4.2 Complete empirical Hoeffding–Sobol decomposition

The main difference between the previous sections and the current one is based on the fact that all subsets I from $\{1,\ldots,d\}$ are considered here. This paper could have begun by defining the new summaries associated with all of the subsets and not just the pairs. However, due to our strong desire to introduce a new graph for extreme value analysis, we have to deal with weights associated with pairs. The coefficient $D_I^{\text{super}}(\ell)$ generalizes the pair version and stands for $\sum_{J\supseteq I} D_J(\ell)$. The global sensitivity indices are also extended to any subset I by $S_I(\ell) = D_I(\ell)/D(\ell)$ and $S_I^{\text{super}}(\ell) = D_I^{\text{super}}(\ell)/D(\ell)$. These two ratios satisfy $0 \leq S_I(\ell) \leq S_I^{\text{super}}(\ell) \leq 1$ with extreme behavior $S_I^{\text{super}}(\ell) = 0$ or $S_I(\ell) = 1$ that characterizes the situation where $\ell(\mathbf{u})$ does not depend on $\mathbf{u_I}$ or where $\ell(\mathbf{u})$ only depends on $\mathbf{u_I}$.

The Hoeffding–Sobol decomposition of the function $\hat{\ell}_{k,n}$ is given by

$$\hat{\ell}_{k,n}(u_1,\ldots,u_d) = \hat{\ell}_{k,n;\emptyset} + \sum_{i=1}^{d} \hat{\ell}_{k,n;\{i\}}(u_i) + \sum_{i<j} \hat{\ell}_{k,n;\{i,j\}}(u_i,u_j) + \ldots + \hat{\ell}_{k,n;\{1,\ldots,d\}}(u_1,\ldots,u_d)$$

where the subfunctions are centered and orthogonal as explained in Section 2.1. The framework under study is fairly unusual in global sensitivity analysis since most of the empirical computation can be done and all of the terms have an explicit expression as stated below.

Theorem 2 *Let $I \subseteq \{1,\ldots,d\}$ be non-empty. The associated subfunction in the decomposition is*

$$\hat{\ell}_{k,n;I}(\mathbf{u}_I) = -\frac{1}{k}\sum_{s=1}^{n}\left\{\prod_{t\in I}\left(\mathbf{1}\{u_t < \tilde{R}_s^{(t)}\} - \mu_t\left(\tilde{R}_s^{(t)}\right)\right)\prod_{t\notin I}\mu_t\left(\tilde{R}_s^{(t)}\right)\right\},$$

and its variance, which corresponds to an estimate of $D_I(\ell)$, has the following form

$$D_I(\hat{\ell}_{k,n}) = \frac{1}{k^2}\sum_{s=1}^{n}\sum_{s'=1}^{n}\left(\prod_{t\in I}\left\{\mu_t\left(\tilde{R}_s^{(t)}\wedge\tilde{R}_{s'}^{(t)}\right) - \mu_t\left(\tilde{R}_s^{(t)}\right)\mu_t\left(\tilde{R}_{s'}^{(t)}\right)\right\}\prod_{t\notin I}\mu_t\left(\tilde{R}_s^{(t)}\right)\mu_t\left(\tilde{R}_{s'}^{(t)}\right)\right).$$

The superset part of $\hat{\ell}_{n,k}(\mathbf{u})$ associated with subset I, which is the part of Hoeffding–Sobol decomposition that contains the variables $\{u_t, t\in I\}$, is given by

$$\hat{\ell}^{super}_{k,n;I}(\mathbf{u}) = -\frac{1}{k}\sum_{s=1}^{n}\prod_{t\in I}\left(\mathbf{1}\{u_t < \tilde{R}_s^{(t)}\} - \mu_t\left(\tilde{R}_s^{(t)}\right)\right)\prod_{t\notin I}\mathbf{1}\{u_t < \tilde{R}_s^{(t)}\},$$

with a variance of

$$D_I^{super}(\hat{\ell}_{k,n}) = \frac{1}{k^2}\sum_{s=1}^{n}\sum_{s'=1}^{n}\left\{\prod_{t\in I}\left(\mu_t\left(\tilde{R}_s^{(t)}\wedge\tilde{R}_{s'}^{(t)}\right) - \mu_t\left(\tilde{R}_s^{(t)}\right)\mu_t\left(\tilde{R}_{s'}^{(t)}\right)\right)\prod_{t\notin I}\mu_t\left(\tilde{R}_s^{(t)}\wedge\tilde{R}_{s'}^{(t)}\right)\right\}.$$

Explicit expressions are then immediately deduced for the estimators of $S_I(\ell)$ and $S_I^{super}(\ell)$ defined as

$$\hat{S}_{k,n;I} := \frac{D_I(\hat{\ell}_{k,n})}{D(\hat{\ell}_{k,n})} \qquad \textit{and} \qquad \hat{S}^{super}_{k,n;I} := \frac{D_I^{super}(\hat{\ell}_{k,n})}{D(\hat{\ell}_{k,n})},$$

by plugging the expression of $D(\hat{\ell}_{k,n})$ given in Proposition 2. Additionally, the constant term in the Hoeffding–Sobol decomposition can be written as

$$\hat{\ell}_{k,n;\emptyset} = \frac{n}{k} - \frac{1}{k}\sum_{s=1}^{n}\prod_{t=1}^{d}\mu_t\left(\tilde{R}_s^{(t)}\right).$$

By construction, the inequality $D_I(\ell) \le D_I^{\text{super}}(\ell)$ is verified by the empirical estimates whose expressions are stated above. This can be directly checked by observing that

$$\mu_t\left(\tilde{R}_s^{(t)}\right)\mu_t\left(\tilde{R}_{s'}^{(t)}\right) = \mu_t\left(\tilde{R}_s^{(t)}\wedge\tilde{R}_{s'}^{(t)}\right)\mu_t\left(\tilde{R}_s^{(t)}\vee\tilde{R}_{s'}^{(t)}\right) \le \mu_t\left(\tilde{R}_s^{(t)}\wedge\tilde{R}_{s'}^{(t)}\right).$$

4.3 Under standard uniform distributions

The issue of the optimal choice for the marginal distribution of $\mathbf{U}$ will not be raised in this paper. However, it is natural to consider the uniform distribution. By requiring that $\mathbf{u}$ be in the unit cube, introducing $\bar{R}_s^{(t)} := \tilde{R}_s^{(t)} \wedge 1$ makes it possible to write

$$\hat{\ell}_{k,n}(\mathbf{u}) = \frac{1}{k}\sum_{s=1}^{n} \mathbf{1}\Big\{u_1 \geq \bar{R}_s^{(1)} \text{ or } \ldots \text{ or } u_d \geq \bar{R}_s^{(d)}\Big\} .$$

All preceding expressions can be simplified thanks to the equalities $\mu_t(\tilde{R}_s^{(t)}) = \bar{R}_s^{(t)}$ and $\mu_t(\tilde{R}_s^{(t)} \wedge \tilde{R}_{s'}^{(t)}) = \bar{R}_s^{(t)} \wedge \bar{R}_{s'}^{(t)}$, when each probability measure μ_t is the standard uniform distribution. In particular, the empirical tail superset importance coefficient is

$$D^{\text{super}}_{\{i,j\}}(\hat{\ell}_{k,n}) = \frac{1}{k^2}\sum_{s=1}^{n}\sum_{s'=1}^{n}\left\{\prod_{t\in\{i,j\}}\left(\bar{R}_s^{(t)} \wedge \bar{R}_{s'}^{(t)} - \bar{R}_s^{(t)}\bar{R}_{s'}^{(t))}\right)\prod_{t\notin\{i,j\}} \bar{R}_s^{(t)} \wedge \bar{R}_{s'}^{(t)}\right\}$$

and its scaled version becomes

$$\hat{S}^{\text{super}}_{k,n;\{i,j\}} = \frac{\sum_{s=1}^{n}\sum_{s'=1}^{n}\left\{\prod_{t\in\{i,j\}}\left(\bar{R}_s^{(t)} \wedge \bar{R}_{s'}^{(t)} - \bar{R}_s^{(t)}\bar{R}_{s'}^{(t))}\right)\prod_{t\notin\{i,j\}} \bar{R}_s^{(t)} \wedge \bar{R}_{s'}^{(t)}\right\}}{\sum_{s=1}^{n}\sum_{s'=1}^{n}\left\{\prod_{t=1}^{d} \bar{R}_s^{(t)} \wedge \bar{R}_{s'}^{(t)} - \prod_{t=1}^{d} \bar{R}_s^{(t)}\bar{R}_{s'}^{(t)}\right\}} .$$

4.4 Consistency

The large sample behavior of the new empirical indices is now stated. In order to prove consistency for the empirical tail superset importance coefficients, we only need to focus on settings where the estimator $\hat{\ell}_{k,n}$ is uniformly consistent for the estimation of ℓ.

(H_1) **The first-order assumption**
The multivariate distribution function F with continuous marginal distribution functions F_t for $t = 1, \ldots, d$ is such that the limit (6) exists for all $\mathbf{u}$ in the positive orthant.

Theorem 3 (Theorem 7.2.1 of de Haan and Ferreira (2006)) *Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be n independent copies of the random vector $\mathbf{X}$. Assume (H_1) holds and let k be an intermediate sequence: $k = k(n)$ such that $k \to \infty$ as $n \to \infty$ but with $k = o(n)$. Then, as n tends to infinity,*

$$\sup_{\mathbf{u}\in[0,1]^d} \left|\hat{\ell}_{k,n}(\mathbf{u}) - \ell(\mathbf{u})\right| \xrightarrow{\mathbb{P}} 0 .$$

This uniform convergence is valid for any subfunction of the Hoeffding–Sobol decomposition, as stated in the next result.

Corollary 1 *Under conditions of Theorem 3, all the empirical subfunctions defined in Theorem 2 are uniformly consistent*

$$\sup_{\mathbf{u}_I\in[0,1]^{|I|}} |\hat{\ell}_{k,n;I}(\mathbf{u}_I) - \ell_I(\mathbf{u}_I)| \xrightarrow{\mathbb{P}} 0 .$$

As a consequence, the empirical variances $D_I(\hat{\ell}_{k,n})$ *or* $D_I^{super}(\hat{\ell}_{k,n})$ *and the empirical ratios* $\hat{S}_{k,n;I}$ *or* $\hat{S}_{k,n;I}^{super}$ *from Theorem 2 are consistent for their equivalent theoretical forms. In particular, the (scaled) tail superset coefficients, whose empirical versions are detailed in Proposition 2, satisfy* $D_{\{i,j\}}^{super}(\hat{\ell}_{k,n}) \xrightarrow{\mathbb{P}} D_{\{i,j\}}^{super}(\ell)$ *and* $\hat{S}_{k,n;\{i,j\}}^{super} \xrightarrow{\mathbb{P}} S_{\{i,j\}}^{super}(\ell)$ *so that the empirical (scaled) dependograph is consistent.*

4.5 Asymptotic distribution

We need to assume some additional conditions in order to establish some asymptotics for the distribution of the empirical tail superset importance coefficients.

(H_2) **The second-order condition**
There exists a function a of constant sign and tending to zero and a function m that is not proportional to ℓ such that

$$\lim_{z\to\infty} \frac{z\left\{1 - F\left(F_1^{-1}(1-u_1/z), \ldots, F_d^{-1}(1-u_d/z)\right)\right\} - \ell(\mathbf{u})}{a(z)} = m(\mathbf{u}) \tag{14}$$

uniformly on $[0,1]^d$. Note that this condition quantifies the speed of convergence in the first-order assumption.

The next result provides a general setting to describe the asymptotic distribution of the function $\hat{\ell}_{k,n}(\cdot)$.

Lemma 1 *Let* $\mathbf{X}_1, \ldots, \mathbf{X}_n$ *be* n *independent copies of* $\mathbf{X}$. *Assume* (H_2) *holds. Suppose further that, for any* $t = 1, \ldots, d$, *the first-order partial derivative of* ℓ, *denoted by* $\partial_t \ell$, *exists and is continuous on the set of points* $\{\mathbf{u} = (u_1, \ldots, u_d) \in \mathbb{R}_+^d : u_t > 0\}$.
Let k *be an intermediate sequence such that* $\sqrt{k}a(n/k) \to a$. *Then as* n *tends to infinity,*

$$\sqrt{k}\left\{\hat{\ell}_{k,n}(\mathbf{u}) - \ell(\mathbf{u})\right\} \xrightarrow{d} Z_\ell(\mathbf{u}) + am(\mathbf{u}) \,, \tag{15}$$

in the space of functions in $[0,1]^d$ *that are right-continuous and have left-hand limits. The process* Z_ℓ *above can be stochastically represented as follows*

$$Z_\ell(\mathbf{u}) := W_\ell(\mathbf{u}) - \sum_{t=1}^{d} W_\ell(u_t \mathbf{e}_t)\partial_t \ell(\mathbf{u}) \,. \tag{16}$$

The process W_ℓ *is a continuous centered Gaussian process with* $\mathbb{E}[W_\ell(\mathbf{u})W_\ell(\tilde{\mathbf{u}})] = \nu\{H(\mathbf{u}) \cap H(\tilde{\mathbf{u}})\}$ *given in terms of the intensity measure* ν *defined by* $\nu\{\tilde{\mathbf{u}} \in \mathbb{R}_+^d : \exists t$ *such that* $\tilde{u}_t > u_t\} := \ell(\mathbf{u})$ *and* $H(\mathbf{u}) = \{\tilde{\mathbf{u}} \in \mathbb{R}_+^d : \exists t$ *such that* $0 \le \tilde{u}_t \le u_t\}$.

The asymptotic distributional expansion (16) is inherited by subfunctions of the ANOVA decomposition, as stated in the next result.

Corollary 2 *Suppose that the assumptions of Lemma 1 hold. Assume moreover that the function* m *in (14) is square integrable with respect to* $\otimes_{t=1}^d d\mu_t$. *Then, as* n *tends to infinity,*

$$\sqrt{k}\left\{\hat{\ell}_{k,n;I}(\mathbf{u}_I) - \ell_I(\mathbf{u}_I)\right\} \xrightarrow{d} Z_{\ell;I}(\mathbf{u}_I) + am_I(\mathbf{u}_I)$$

where the stochastic process $Z_{\ell;I} : \omega \in \Omega \mapsto \{Z_\ell(\omega)\}_I$ *corresponds to the path of the subfunction associated with subset* I *in the Hoeffding–Sobol decomposition of* $Z_\ell(\omega)$.
Moreover,

$$\sqrt{k}\left(D_I(\hat{\ell}_{k,n}) - D_I(\ell)\right) \xrightarrow{d} 2\mathbb{E}_{\mathbf{U}_I}[\ell_I(\mathbf{U}_I)\{Z_{\ell;I}(\mathbf{U}_I) + am_I(\mathbf{U}_I)\}]$$

where the expectation $\mathbb{E}_{\mathbf{U}_I}[\cdot]$ *refers to the integral with respect to* $\otimes_{t\in I} d\mu_t$. *Finally, if* $\ell_I \equiv 0$ *that is equivalent to* $D_I(\ell) = 0$, *we obtain*

$$kD_I(\hat{\ell}_{k,n}) \xrightarrow{d} D_I(Z_\ell + am) \ .$$

The previous result states an asymptotic normality in most cases. When the true coefficient is null, the asymptotic distribution has a χ^2 type. Let us point out as well that by adding terms from Corollary 2, we also obtain, if $D_I^{\text{super}}(\ell) > 0$,

$$\sqrt{k}\left(D_I^{\text{super}}(\hat{\ell}_{k,n}) - D_I^{\text{super}}(\ell)\right) \xrightarrow{d} \sum_{J\supseteq I} 2\mathbb{E}_{\mathbf{U}_J}[\ell_J(\mathbf{U}_J)\{Z_{\ell;J}(\mathbf{U}_J) + am_J(\mathbf{U}_J)\}] \ .$$

The next proposition once again states the asymptotic distribution for the empirical tail superset coefficients. Another limiting expression is thus obtained, depending only on function ℓ and not on the subfunctions $\{\ell_J, J \supseteq I\}$ as above.

Proposition 3 *Under the assumptions of Lemma 1, as* n *tends to infinity,*

$$\sqrt{k}\left(D_I^{super}(\hat{\ell}_{k,n}) - D_I^{super}(\ell)\right) \xrightarrow{d} \frac{1}{2^{|I|-1}} \int \Delta_I[\ell, Z_\ell + am](\mathbf{u}_I, \mathbf{v}) d\mathbf{u}_I d\mathbf{v}$$

where $\Delta_I[\ell, Z](\mathbf{u}_I, \mathbf{v}) = \left(\sum_{J\subseteq I}(-1)^{|I\setminus J|}\ell(\mathbf{u}_J, \mathbf{v}_{-J})\right)\left(\sum_{K\subseteq I}(-1)^{|I\setminus K|}Z(\mathbf{u}_K, \mathbf{v}_{-K})\right)$.
If $D_I^{super}(\ell) = 0$ *then*

$$kD_I^{super}(\hat{\ell}_{k,n}) \to D_I^{super}(Z_\ell + am) \ .$$

In particular, when I stands for the pair $\{i, j\}$, the asymptotic distribution for the indices involved in the empirical tail dependograph can be derived. More precisely, the convergence of $D_{\{i,j\}}^{\text{super}}(\hat{\ell}_{k,n})$ to $D_{\{i,j\}}^{\text{super}}(\ell)$ leads to two different regimes. In general, the classical root square rate of convergence is obtained. However, when the true tail superset importance coefficient vanishes, then the rate is faster.

5 Additional illustrations

We first exhibit a setting where classical tail measures fail to find extreme dependence. We study the performance of the non-parametric estimators on simulation. In particular, the order of tail superset importance between pairs is rather good for reasonable sample sizes. Finally, the last section illustrates the new indices and graphs on real data.

5.1 Revealing asymptotic dependence under asymmetric models

We now construct an example where δ and λ do not provide the same information. Let $d = 4$. Consider the parametric asymmetric model with s.t.d.f. given as

$$\ell(u_1, u_2, u_3, u_4) = (1-\psi)(u_1+u_2) + \psi\left(u_1^{1/\alpha} + u_2^{1/\alpha}\right)^{\alpha} + (1-\tilde{\psi})(u_3+u_4) + \tilde{\psi}\left(u_3^{1/\tilde{\alpha}} + u_4^{1/\tilde{\alpha}}\right)^{\tilde{\alpha}} \tag{17}$$

Each bivariate component is a mixture of AI with the logistic model of Gumbel introduced by Tawn (1990). Fix $(\psi, \alpha) = (0.2, 0.2)$ and $(\tilde{\psi}, \tilde{\alpha}) = (0.8, 0.83)$. The pairplots (X_1, X_2) and (X_3, X_4) are given on the left and right of Figure 3, respectively, under Gumbel margins.

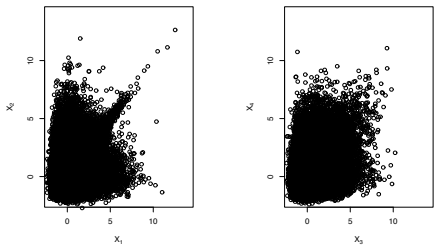


Fig. 3: Two pair plots for $\mathbf{X}$ a four-dimensional extreme value random vector with Gumbel margins and s.t.d.f. (17). Left: (X_1, X_2). Right: (X_3, X_4).

We have $\lambda_{12} \simeq 0.17 < \lambda_{34} \simeq 0.18$ whereas $D_{12}^{\text{super}} \simeq 3.09 \times 10^{-4} > D_{34}^{\text{super}} \simeq 1.81 \times 10^{-4}$. The new index better detects the strongest tail dependent pair even in a small proportion. In terms of graphs, this example leads to Figure 4, where the difference between the messages of the two graphs speaks for itself.

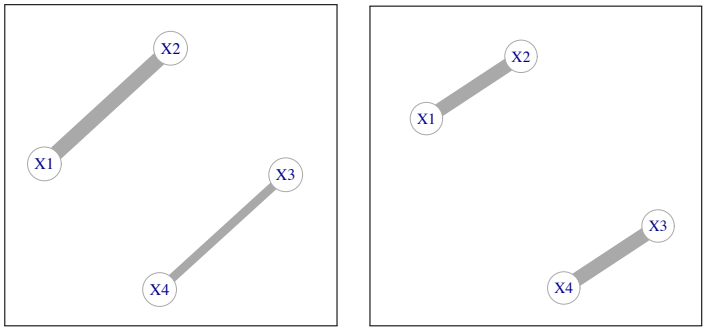


Fig. 4: The tail dependograph (left) and the λ-graph (right) for s.t.d.f. (17).

5.2 Finite sample performance in estimating the tail superset indices

We once again consider the s.t.d.f. ℓ as given in (11) from Section 3.5. Recall that the numerical values of several indices were collected in Table 1. Our goal now is to estimate the pairwise coefficients $\lambda_{\{i,j\}}$ and $D^{\text{super}}_{\{i,j\}}$ on trivariate samples. We assume the distribution to be trivariate max-stable with Fréchet margins, $F(x_1, x_2, x_3) = \exp\left(-\ell(x_1^{-1}, x_2^{-1}, x_3^{-1})\right)$. Computed on several repetitions, we first record the proportion of trials that correctly ordered a triplet of theoretical coefficients. This means that a trial is a success when $D^{\text{super}}_{\{1,2\}}(\hat{\ell}_{k,n}) > D^{\text{super}}_{\{2,3\}}(\hat{\ell}_{k,n}) > D^{\text{super}}_{\{1,3\}}(\hat{\ell}_{k,n})$ or when $\lambda_{\{1,2\}}(\hat{\ell}_{k,n}) > \lambda_{\{2,3\}}(\hat{\ell}_{k,n}) > \lambda_{\{1,3\}}(\hat{\ell}_{k,n})$ respectively, depending on the criterion we are examining. Varying the sample size n, we obtain the curves presented in Figure 5. The values for the experiment are: $n \in \{100, 200, \ldots, 1000\}$, $N = 1000$ repetitions and several choices for k have been taken. We only plot two of them: $k = 50$ (plain lines) and $k = n/2$ (dotted lines), which, roughly speaking, correspond to the worst and best cases.

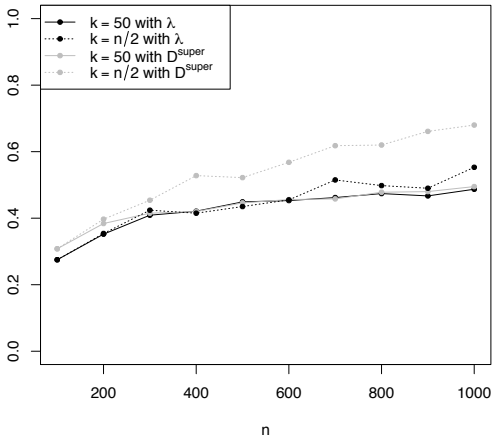
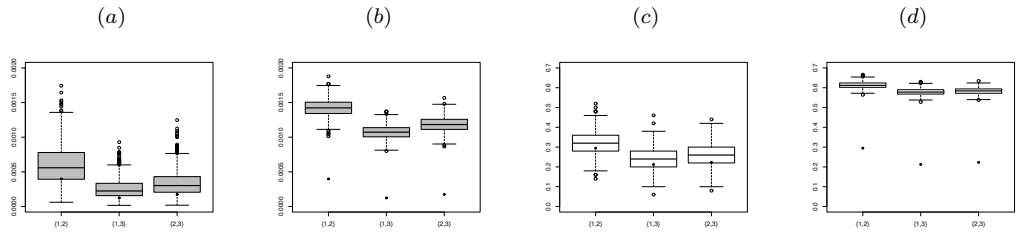


Fig. 5: Proportion of well-ordered triplets of theoretical coefficients. For the first (resp. second) line of Table 1, results are in gray (resp. black).

To furnish most of the details, we also provide the boxplots associated with the estimation of each triplet in both situations: $k = 50$ and $k = n/2$ when $n = 1000$ in Figure 6. It is worth noting that even if the coefficients are not as well estimated when k is large (see Figure 6 (b) and (d)), a classical observation in extreme value analysis, the proportion of well-ordered triplets remains good and even increases compared to the value obtained for a small k (see Figure 6 (a) and (c)).



$D^{\text{super}}_{\{i,j\}}$ with $k = 50$ | $D^{\text{super}}_{\{i,j\}}$ with $k = n/2$ | $\lambda_{\{i,j\}}$ with $k = 50$ | $\lambda_{\{i,j\}}$ with $k = n/2$

Fig. 6: Boxplots when $n = 1000$ for the estimation of $D^{\text{super}}_{\{i,j\}}$ (left) and $\lambda_{\{i,j\}}$ (right). For both type of indices, the first triplet is obtained with $k = 50$ and the second with $k = n/2$. Black points represent true values, already given in Table 1.

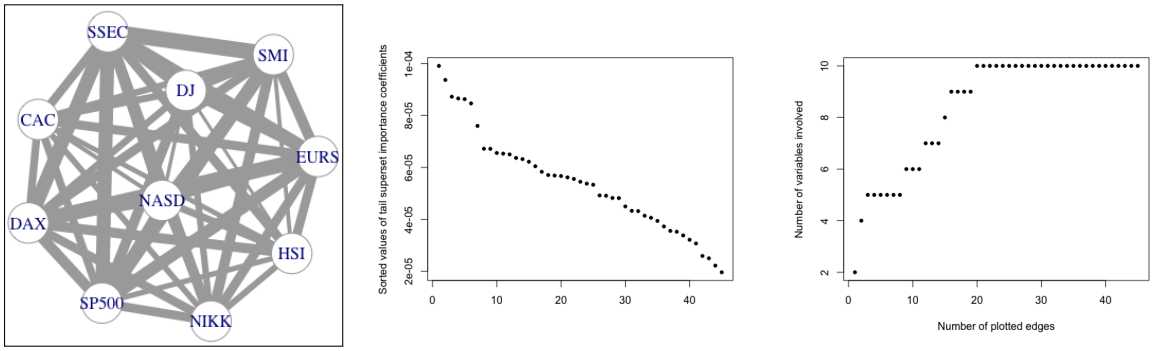
5.3 Selection of the strongest tail dependences in large dimension

In this case, we want to use the tail superset importance coefficients to select the pairs with the strongest asymptotic dependence. This should not be understood in the sense of the strongest extreme bivariate structure of dependence. Indeed, it should be recalled that the tail superset importance coefficient of a pair measures the contribution in the global tail dependence not only through the pair itself, but also through all its supersets.
In both Section 5.3.1 and 5.3.2, we focus on the tail dependence. For that reason, we have decided to work on preprocessed data (renormalized yearly maxima). In that way, we avoid the discussion about the choice of k and the marginal distributions. Note that working with yearly maxima also has the benefit of avoiding temporal dependence issues.

5.3.1 Log-return contagion

Let us consider the close prices of ten stock indices using the `qrmdata` package (Hofert and Hornik, 2016): S&P 500 (SP500), Dow Jones (DJ), NASDAQ 100 (NASD), Swiss Market Index (SMI), Euro Stoxx 50 (EURS), Cotation Assiste en Continu (CAC), Deutscher Aktienindex (DAX), Hang Seng Index (HSI), Shanghai Stock Exchange Composite Index (SSEC), and the NIKKEI (NIKK). Focusing on log-returns for the period 1990 to 2015, we compute renormalized yearly maxima. We thus obtain a ten-dimensional vector with unit Fréchet marginal distributions and a joint distribution characterized by an unknown s.t.d.f. ℓ.
The tail dependograph is provided in Figure 7(a). It contains 45 edges ($d = 10$). We also furnish two additional plots. Figure 7(b) shows the values of the tail superset importance coefficients $D^{\text{super}}_{\{i,j\}}$ in a decreasing order, whereas Figure 7(c) represents the number of variables involved by at least one edge with respect to the number of the largest tail indices that are plotted (starting by the largest).

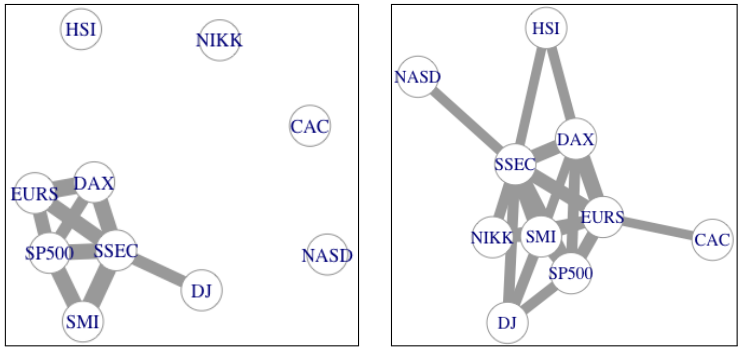
The tail dependences are not all of the same importance, as shown in Figure 7(b). It is possible to only focus on a partial dependograph (with the intention of dimension reduction) that is a tail dependence graph where only the largest tail superset importance coefficients are plotted.



(a) (b) (c)

Fig. 7: For ten log-returns over the period 1990-2015: (a) Complete empirical tail dependograph (b) Sorted estimated values of $D^{\text{super}}_{\{i,j\}}$ (c) Number of variables involved in a partial tail dependograph with respect to the number of plotted edges (starting by the largest).

The largest empirical tail superset coefficient is obtained for the pair SSEC-SMI. The second largest links EURS and DAX. The eight largest empirical tail superset coefficients concern the group {SMI, SSEC, SP500, EURS, DAX}. DJ then joins the graph through its tail superset dependence with SSEC, as illustrated in Figure 8(a). We now have to wait for the 12th largest (resp. 15th, 16th) value of these empirical measures to see a connection of NIKK (resp. HSI, NASD) with the group through SSEC (resp. DAX, SSEC). Finally the 20th largest value of the empirical tail superset coefficients links CAC to the rest of the group via EURS, as plotted in Figure 8(b).



(a) (b)

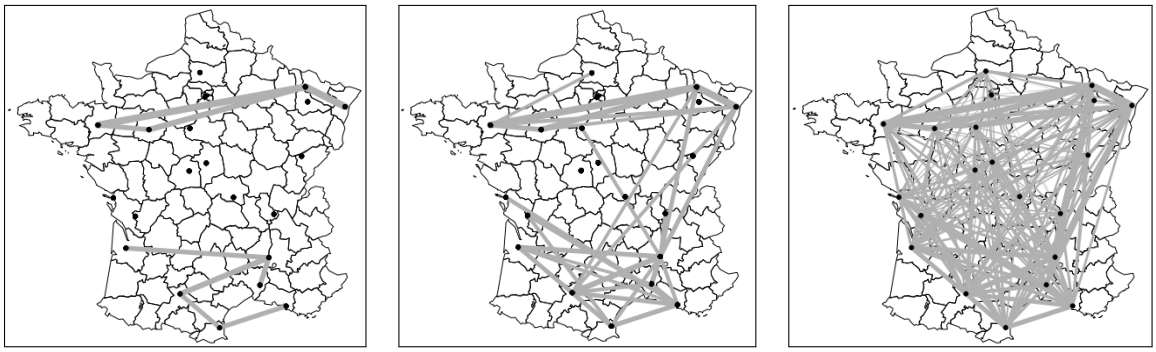
Fig. 8: Partial empirical tail dependographs for ten log-returns over the period 1990-2015: (a) The nine largest empirical tail superset coefficients (b) the twenty largest.

We do not want to favor one methodology over another to select variables. Our intention is to provide new indices, to compute or estimate them, and to share the interest for a number of related issues based on the assessment of these values. Selection with respect to a criterion

is a very general question that is beyond the scope of the underlying study. For that reason, we do not conclude by measuring the effective dimension of the tail dependence structure.

5.3.2 Temperatures in France

The yearly maxima temperatures of 21 continental cities in France were recorded during the period 1946-2000. In Figure 9, we plot two partial dependographs and the complete one. Of course, looking at all the empirical tail superset information simultaneously on Figure 9(c) is uninformative because of the dimension of the problem here. However, the strongest connections can be observed in the north, and after that, a strong dependence structure appears in the south, as can be seen in Figure 9(a). In the partial tail dependograph of Figure 9(b), links between the north and south are already present.



(a) (b) (c)

Fig. 9: Partial and complete empirical tail dependographs for temperatures of 21 French continental cities over the period 1946-2000: (a) The nine largest empirical tail superset coefficients (b) the 30 largest (c) all of them.

Concluding remarks and perspectives

We have introduced new measures to explore the asymptotic dependence structure in extreme value theory. These indices, referred to as tail superset importance coefficients, are based on the Hoeffding–Sobol (ANOVA) decomposition of the stable tail dependence function (s.t.d.f.). They quantify the (proportion of) variance of the s.t.d.f. explained by all supersets of pairs of variables. In other words, for a given pair, the tail superset importance index informs not only on the asymptotic bivariate structure but also on any tail dependence structure that contains this pair.

In a d-variate setting, it is remarkable that such information is available at a quadratic cost with respect to the number of variables ($d(d-1)/2$ pairs), compared to the exponential number of subsets of $\{1,\ldots,d\}$ in the case of the well-known extremal coefficients θ_I. Furthermore, these coefficients can be visualized on a graph, referred to as a tail dependograph. On the other hand, the powerful probabilistic interpretation (12) of the family of θ_I's coefficients is lost, whereas a comparison via a global variance decomposition is gained. Our aim is not to say

if the new family is preferable to the old one. In reality, they are both only necessarily partial summaries of an infinite-dimensional measure. Consequently, and as been illustrated through several examples in this paper, these two families of indices are complementary.

On the theoretical level, we derived several results for statistical inference. First, analytical expressions are available when the s.t.d.f. is replaced with its empirical version. Second, we proved consistency of the tail superset importance coefficient estimates under usual assumptions ensuring uniform consistency for the empirical s.t.d.f. Asymptotic distribution is also obtained, with the unusual feature that the speed of convergence depends on the nullity of the (true) value: $O(k^{-1/2})$ for a non-zero index and $O(k^{-1})$ for a null value.

We further investigated the behavior of the new measures compared to other competitors based on standard extremal coefficients. We observed that some features are better detected by the new coefficients, such as asymmetry, in dependent models for extremes. On the other hand, initial trials indicate that tail superset importance indices do not improve asymptotic independence testing. Thus, they may be more useful to describe the dependence structure of extremes via the tail dependograph.

Several questions could perhaps be answered with these new measures: First, the effective dimension of the asymptotic distribution under the assumption of max-domain of attraction; Second, a multivariate summary of the asymptotic dependence, similar to the multivariate tau, for example; and third, a real guide for the construction of nested models. Other topics related to the estimation of the tail superset importance coefficients could be dealt with as well. We restricted our study to the empirical estimator of the stable tail dependence function. Other choices are possible, such as the asymptotically unbiased version of Fougères et al. (2015) or the M-estimator of Einmahl et al. (2016). Finally, the inference of the tail superset coefficients inside parametric models could be the focus of future research.

Acknowledgments

The authors would like to thank Christian Genest (McGill University, Montréal, Canada) for fruitful discussions. The first author would also like to thank Roland Denis and Benoît Fabrèges (Institut Camille Jordan, Université de Lyon, France) whose discussions and workshop led to great progress in the codes associated with this paper.

Proofs

Details for Section 3.4.
We prove here the combinatorial formulas used in the definition of the λ-supergraph. First, starting from the identity $(1+x)^d = \sum_{I\subseteq\{1,\dots,d\}} x^{|I|}$, and computing the derivative at 1, we obtain $\sum_{I\subseteq\{1,\dots,d\}} |I| = d2^{d-1}$. Thus, by parameterizing the supersets I of a pair $\{i,j\}$ by $J = I \setminus \{i,j\}$, we have

$$\sum_{I\supseteq\{i,j\}} |I| = \sum_{J\subseteq\{1,\dots,d-2\}} \{|J|+2\} = (d-2)2^{d-3} + 2\times 2^{d-2} = (d+2)2^{d-3}.$$

Consequently, the upper bound of the weight edges is given by $\sum_{I\supseteq\{i,j\}} \{|I| - 1\} = d2^{d-3}$.

Proof of Proposition 1.
For the sake of simplicity, let us identify $\ell(\mathbf{u}_A, \mathbf{u}_B, \mathbf{0}_C)$ with $\ell(\mathbf{u}_A, \mathbf{u}_B)$. Assume first that $\ell(\mathbf{u}) = \ell(\mathbf{u}_A, \mathbf{0}_B) + \ell(\mathbf{0}_A, \mathbf{u}_B)$ and introduce the notation $f(\mathbf{u}_A) = \ell(\mathbf{u}_A, \mathbf{0}_B)$ and $g(\mathbf{u}_B) = \ell(\mathbf{0}_A, \mathbf{u}_B)$. The functions f and g admit the Hoeffing–Sobol decomposition $f(\mathbf{u}_A) = \sum_{I \subseteq A} f_I(\mathbf{u}_I)$ and $g(\mathbf{u}_B) = \sum_{I \subseteq B} g_I(\mathbf{u}_I)$ so that, from the uniqueness of the decomposition,

$$\ell(\mathbf{u}) = \sum_{I \subseteq \{1,\ldots,d\}} \ell_I(\mathbf{u}_I)$$

with $\ell_I = f_I$ if $I \subseteq A$, $\ell_I = g_I$ if $I \subseteq B$ and $\ell_I \equiv 0$ if $I \cap A \neq \emptyset$ and $I \cap B \neq \emptyset$. It yields

$$D^{\text{super}}_{\{i,j\}}(\ell) = \begin{cases} D^{\text{super}}_{\{i,j\}}(f) & \text{if } \{i,j\} \subseteq A \\ D^{\text{super}}_{\{i,j\}}(g) & \text{if } \{i,j\} \subseteq B \\ 0 & \text{if } (i,j) \in A \times B \text{ or } (i,j) \in B \times A\,. \end{cases}$$

No edge goes from any vertex in A to any vertex in B.
Now, assume that $D^{\text{super}}_{\{i,j\}}(\ell) = 0$ for all $i \in A$ and $j \in B$. Then, since it is a sum of positive terms, all terms vanish: $D_K = 0$ and $\ell_K \equiv 0$ $\forall K$ that contains $\{i,j\}$. Thus $\ell(\mathbf{u}) = f(\mathbf{u}_A) + g(\mathbf{u}_B)$. In particular $\ell(\mathbf{u}_A, \mathbf{0}_B) = f(\mathbf{u}_A) + g(\mathbf{0}_B)$ and $\ell(\mathbf{0}_A, \mathbf{u}_B) = f(\mathbf{0}_A) + g(\mathbf{u}_B)$. By adding these terms, we obtain

$$\ell(\mathbf{u}_A, \mathbf{0}_B) + \ell(\mathbf{0}_A, \mathbf{u}_B) = f(\mathbf{u}_A) + g(\mathbf{0}_B) + f(\mathbf{0}_A) + g(\mathbf{u}_B) = \ell(\mathbf{u}) + \ell(\mathbf{0}) = \ell(\mathbf{u})\,.$$

For the proofs, we first need to provide a technical preliminary lemma in sensitivity analysis.

Lemma 2 *Let f be a tensor-product function $f(\mathbf{u}) = \prod_{t=1}^{d} f_t(u_t)$. Assume, moreover, the framework of a Hoeffding–Sobol decomposition: $u_1, \ldots, u_d$ are viewed as independent random variables, and $f(\mathbf{u})$ is square integrable. For any subset $I \subseteq \{1, \ldots, d\}$, denote the term indexed by I in the Hoeffding–Sobol decomposition of $f(\mathbf{u})$ by $f_I(\mathbf{u}_I)$. Finally, set $m_t = \mathbb{E}[f_t(u_t)]$. Then*

$$f_I(\mathbf{u}_I) = \prod_{t \in I} \{f_t(u_t) - m_t\} \prod_{t \notin I} m_t$$

and its variance is $D_I(f) = \prod_{t \in I} \text{Var}(f_t(u_t)) \prod_{t \notin I} m_t^2$. Furthermore, for a given set $I \subseteq \{1, \ldots, d\}$, the superset part of $f(\mathbf{u})$, i.e. the part in the Hoeffding–Sobol decomposition that contains the variables $u_t, t \in I$ is given by

$$f_I^{super}(\mathbf{u}) = \prod_{t \in I} \{f_t(u_t) - m_t\} \prod_{t \notin I} f_t(u_t) \tag{18}$$

and its variance is $D_I^{super}(f) = \prod_{t \in I} \text{Var}(f_t(u_t)) \prod_{t \notin I} \mathbb{E}\left(f_t(u_t)^2\right)$.

Proof of Lemma 2.
For the first part of the lemma, observe that the Hoeffding–Sobol decomposition of $f(\mathbf{u})$ is obtained by expanding the product

$$f(\mathbf{u}) = \prod_{t=1}^{d} \{(f_t(u_t) - m_t) + m_t\} = \sum_{I \subseteq \{1,\ldots,d\}} f_I(\mathbf{u}_I),$$

with $f_I(\mathbf{u}_I) = \prod_{t\in I}\{f_t(u_t) - m_t\}\prod_{t\notin I} m_t$. Such f_Is satisfy the condition $\mathbb{E}[f_I(\mathbf{u}_I)|\mathbf{u}_J] = 0$ for each strict subset $J \subsetneq I$ so that the result follows from unicity of the decomposition.
Let us now fix a set $I \subseteq \{1,\dots,d\}$. By expanding the product

$$f(\mathbf{u}) = \prod_{t\in I}\{(f_t(u_t) - m_t) + m_t\}\prod_{t\notin I} f_t(u_t),$$

the term $\prod_{t\in I}(f_t(u_t) - m_t)\prod_{t\notin I} f_t(u_t)$ corresponds to all supersets J of I in the Hoeffding–Sobol decomposition of $f(\mathbf{u})$. Indeed the f_Js are obtained through the same expansion mechanism by picking the first term $(f_t(u_t) - m_t)$ in the parenthesis $\{(f_t(u_t) - m_t) + m_t\}$ when t belongs to $J \supseteq I$. This can also be checked using the analytical expression of f_J proved above:

$$\begin{aligned} f_I^{\text{super}}(\mathbf{u}) &= \prod_{t\in I}(f_t(u_t) - m_t)\left\{\sum_{J\supseteq I}\prod_{t\in J\setminus I}(f_t(u_t) - m_t)\prod_{t\notin J} m_t\right\} \\ &= \prod_{t\in I}(f_t(u_t) - m_t)\left\{\sum_{K\subseteq\{1,\dots,d\}\setminus I}\prod_{t\in K}(f_t(u_t) - m_t)\prod_{t\notin K} m_t\right\} = \prod_{t\in I}(f_t(u_t) - m_t)\prod_{t\notin I} f_t(u_t). \end{aligned}$$

Finally the variance expressions are direct, recalling that each f_I is centered and all the u_t are assumed to be independent.

Proof of Proposition 2.
We have:

$$\hat{\ell}_{k,n}(\mathbf{u}) = \frac{1}{k}\sum_{s=1}^{n}\left(1 - \prod_{t=1}^{d}\mathbf{1}\{u_t < \tilde{R}_s^{(t)}\}\right) = \frac{n}{k} - \frac{1}{k}\sum_{s=1}^{n}\prod_{t=1}^{d}\mathbf{1}\{u_t < \tilde{R}_s^{(t)}\}. \tag{19}$$

Let us now fix a pair $\{i,j\}$. The constant term n/k does not play a role in the superset importance. The second term is a sum of tensor-product functions. Thus, taking Lemma 2, its superset part is obtained by centering the terms in the products that involve the considered pair. Namely, from (18), we obtain

$$\hat{\ell}^{\text{super}}_{k,n;\{i,j\}}(\mathbf{u}) = -\frac{1}{k}\sum_{s=1}^{n}\prod_{t=1}^{d}\left(\mathbf{1}\{u_t < \tilde{R}_s^{(t)}\} - \mathbf{1}\{t\in\{i,j\}\}\mu_t\left(\tilde{R}_s^{(t)}\right)\right).$$

Observing that the term is centered, the variance reduces to the expectation of its square

$$\begin{aligned} \left(\hat{\ell}^{\text{super}}_{k,n;\{i,j\}}(\mathbf{u})\right)^2 = \frac{1}{k^2}\sum_{s=1}^{n}\sum_{s'=1}^{n}\prod_{t=1}^{d}&\left(\mathbf{1}\{u_t < \tilde{R}_s^{(t)}\} - \mathbf{1}\{t\in\{i,j\}\}\mu_t\left(\tilde{R}_s^{(t)}\right)\right) \\ &\times\left(\mathbf{1}\{u_t < \tilde{R}_{s'}^{(t)}\} - \mathbf{1}\{t\in\{i,j\}\}\mu_t\left(\tilde{R}_{s'}^{(t)}\right)\right). \end{aligned}$$

A direct computation of the expectation then gives the announced expression

$$D^{\text{super}}_{\{i,j\}}(\hat{\ell}_{k,n}) = \frac{1}{k^2}\sum_{s=1}^{n}\sum_{s'=1}^{n}\prod_{t=1}^{d}\left(\mu_t\left(\tilde{R}_s^{(t)}\wedge\tilde{R}_{s'}^{(t)}\right) - \mathbf{1}\{t\in\{i,j\}\}\mu_t\left(\tilde{R}_s^{(t)}\right)\mu_t\left(\tilde{R}_{s'}^{(t)}\right)\right).$$

Using Equation (19), we obtain the global variance as follows

$$D(\hat{\ell}_{k,n}) = \frac{1}{k^2}\sum_{s=1}^{n}\sum_{s'=1}^{n}\left\{\mathbb{E}\left(\prod_{t=1}^{d}\mathbf{1}\{u_t < \tilde{R}_s^{(t)}\}\,\mathbf{1}\{u_t < \tilde{R}_{s'}^{(t)}\}\right) - \mathbb{E}\left(\prod_{t=1}^{d}\mathbf{1}\{u_t < \tilde{R}_s^{(t)}\}\right)\mathbb{E}\left(\prod_{t=1}^{d}\mathbf{1}\{u_t < \tilde{R}_{s'}^{(t)}\}\right)\right\},$$

which yields the result by using the independence of the u_t's.

Proof of Theorem 2.
Lemma 2 can be applied to $\hat{\ell}_{k,n}(\cdot) - n/k$ since it is a sum of tensor-product functions. The results lead directly to $\hat{\ell}_{k,n;I}$ for any non-empty subset I, and the term n/k has to be added in order to recover the constant term in the decomposition.

Proof of Corollary 1.
Recall that the Hoeffding–Sobol decomposition relies on the recursive construction, written below in its empirical version,

$$\hat{\ell}_{k,n;I}(\mathbf{u}_I) = \mathbb{E}[\hat{\ell}_{k,n}(\mathbf{U})|\mathbf{U}_I = \mathbf{u}_I] - \sum_{I'\subsetneq I}\hat{\ell}_{k,n;I'}(\mathbf{u}_{I'})\ .$$

The proof is by induction on I. Observe first that

$$|\hat{\ell}_{k,n;\emptyset} - \ell_\emptyset| = |\mathbb{E}[\hat{\ell}_{k,n}(\mathbf{U}) - \ell(\mathbf{U})]| \le \sup_{u\in[0,1]^d}|\hat{\ell}_{k,n}(\mathbf{u}) - \ell(\mathbf{u})| \xrightarrow{\mathbb{P}} 0\ .$$

Now, let I be a non-empty subset of $\{1,\ldots,d\}$ and assume that for any subset $I' \subsetneq I$

$$\sup_{\mathbf{u}_{I'}\in[0,1]^{|I'|}}|\hat{\ell}_{k,n;I'}(\mathbf{u}_{I'}) - \ell_{I'}(\mathbf{u}_{I'})| \xrightarrow{\mathbb{P}} 0\ .$$

Applying the recursive formula twice, we obtain

$$\sup_{\mathbf{u}_I\in[0,1]^{|I|}}|\hat{\ell}_{k,n;I}(\mathbf{u}_I) - \ell_I(\mathbf{u}_I)| \le \sup_{\mathbf{u}_I\in[0,1]^{|I|}}\left|\mathbb{E}[\hat{\ell}_{k,n}(\mathbf{U}) - \ell(\mathbf{U})|\mathbf{U}_I = \mathbf{u}_I]\right| + \sum_{I'\subsetneq I}\sup_{\mathbf{u}_{I'}\in[0,1]^{|I'|}}|\hat{\ell}_{k,n;I'}(\mathbf{u}_{I'}) - \ell_{I'}(\mathbf{u}_{I'})|$$

that both converge to zero in probability. The consistency of the global variance or that of the variances associated with subsets is a direct consequence of what preceeds. Nevertheless, note that the existence and the consistency of the empirical global sensitivity indices $\hat{S}_{k,n;I}$ and $\hat{S}^{\text{super}}_{k,n;I}$ rely on the fact that the empirical global variance $D(\hat{\ell}_{k,n})$ and the limiting global variance $D(\ell)$ never vanish in the extreme value theory setting.

Proof of Lemma 1.
Lemma 1 can be seen as the concatenation of several statements already established in the

literature. See, for example, de Haan and Ferreira (2006; Theorem 7.2.2), Einmahl et al. (2012; Theorem 4.6) or Fougères et al. (2015; Proposition 1) for some very similar statements. The difference between the results is concentrated on the specification of the intermediate sequence k.

Let $Y_s^{(t)}$ denote the uniform random variables $Y_s^{(t)} = 1 - F_t(X_s^{(t)})$ for $t = 1, \dots, d$ and $s = 1, \dots, n$. One can write

$$\hat{\ell}_{k,n}(\mathbf{u}) = V_k \left(\frac{n}{k} Y^{(1)}_{[ku_1],n}, \dots, \frac{n}{k} Y^{(d)}_{[ku_d],n} \right)$$

and the difference

$$\hat{\ell}_{k,n}(\mathbf{u}) - \ell(\mathbf{u}) = A_{k,n}(\mathbf{u}) + B_{k,n}(\mathbf{u}) + C_{k,n}(\mathbf{u})$$

with

$$V_k(\mathbf{u}) = \frac{1}{k} \sum_{s=1}^{n} \mathbf{1}_{\left\{ Y_s^{(1)} \le ku_1/n \text{ or } \dots \text{ or } Y_s^{(d)} \le ku_d/n \right\}}$$

$$A_{k,n}(\mathbf{u}) = V_k \left(\frac{n}{k} Y^{(1)}_{[ku_1],n}, \dots, \frac{n}{k} Y^{(d)}_{[ku_d],n} \right) - \frac{n}{k} \left[1 - F\{F_1^{-1}(1 - Y^{(1)}_{[ku_1],n}), \dots, F_d^{-1}(1 - Y^{(d)}_{[ku_d],n})\} \right]$$

$$B_{k,n}(\mathbf{u}) = \frac{n}{k} \left[1 - F\{F_1^{-1}(1 - Y^{(1)}_{[ku_1],n}), \dots, F_d^{-1}(1 - Y^{(d)}_{[ku_d],n})\} \right] - \ell \left(\frac{n}{k} Y^{(1)}_{[ku_1],n}, \dots, \frac{n}{k} Y^{(d)}_{[ku_d],n} \right)$$

and

$$C_{k,n}(\mathbf{u}) = \ell \left(\frac{n}{k} Y^{(1)}_{[ku_1],n}, \dots, \frac{n}{k} Y^{(d)}_{[ku_d],n} \right) - \ell(\mathbf{u}) \, .$$

We thus have

$$\begin{aligned}
&\sup_{0 \le u_1, \dots, u_d \le 1} \left| \sqrt{k}(\hat{\ell}_{k,n}(\mathbf{u}) - \ell(\mathbf{u})) - W_\ell(\mathbf{u}) + \sum_{t=1}^{d} W_\ell(u_t \mathbf{e}_t) \partial_t \ell(\mathbf{u}) - am(\mathbf{u}) \right| \\
&\quad \le \sup_{0 \le u_1, \dots, u_d \le 1} \left| \sqrt{k} A_{k,n}(\mathbf{u}) - W_\ell(\mathbf{u}) \right| \\
&\quad + \sup_{0 \le u_1, \dots, u_d \le 1} \left| \sqrt{k} C_{k,n}(\mathbf{u}) + \sum_{t=1}^{d} W_\ell(u_t \mathbf{e}_t) \partial_t \ell(\mathbf{u}) \right| \\
&\quad + \sup_{0 \le u_1, \dots, u_d \le 1} \left| \sqrt{k} B_{k,n}(\mathbf{u}) - am(\mathbf{u}) \right| \, .
\end{aligned}$$

Applying de Haan and Ferreira (2006; Proposition 7.2.3) leads in $D([0,1]^d)$ to

$$\sqrt{k} A_{k,n}(\mathbf{u}) \xrightarrow{d} W_\ell(\mathbf{u}) \, .$$

Due to the Skorohod construction we can write

$$\sup_{0 \le u_1, \dots, u_d \le 1} \left| \sqrt{k} A_{k,n}(\mathbf{u}) - W_\ell(\mathbf{u}) \right| \to 0 \quad \text{a.s.} \, . \tag{20}$$

For any intermediate sequence, the proof of Einmahl et al. (2012; Theorem 4.6) ensures the convergence for $t = 1, \ldots, d$

$$\sup_{u\in[0,1]} |\sqrt{k}\left(\frac{n}{k}Y^{(t)}_{[ku],n} - u\right) + W_\ell(u\mathbf{e}_t)| \to 0 \quad \text{a.s.}\,, \tag{21}$$

and finally

$$\sup_{0\le u_1,\ldots,u_d\le 1} \left|\sqrt{k}C_{k,n}(\mathbf{u}) + \sum_{t=1}^{d} W_\ell(u_t\mathbf{e}_t)\partial_t\ell(\mathbf{u})\right| \to 0 \quad \text{a.s.}\,. \tag{22}$$

It thus remains to prove that

$$\sup_{0\le u_1,\ldots,u_d\le 1} \left|\sqrt{k}B_{k,n}(\mathbf{u}) - am(\mathbf{u})\right| \to 0 \quad \text{a.s.}\,.$$

The second order condition that holds uniformly on $[0,1]^d$ in (14) yields

$$\sup_{0\le u_1,\ldots,u_d\le 1} \left|\frac{B_{k,n}(\mathbf{u})}{a(n/k)} - m\left(\frac{n}{k}Y^{(1)}_{[ku_1],n},\ldots,\frac{n}{k}Y^{(d)}_{[ku_d],n}\right)\right| \to 0 \quad \text{a.s.}\,.$$

Then the result follows from $\sqrt{k}a(n/k) \to a$ and

$$\sup_{0\le u_1,\ldots,u_d\le 1} \left|m(\mathbf{u}) - m\left(\frac{n}{k}Y^{(1)}_{[ku_1],n},\ldots,\frac{n}{k}Y^{(d)}_{[ku_d],n}\right)\right| \to 0 \quad \text{a.s.}\,,$$

that is obtained combining (21) and the continuity of the function m.
Finally, we proved that

$$\sup_{0\le u_1,\ldots,u_d\le 1} \left|\sqrt{k}(\hat{\ell}_{k,n}(\mathbf{u}) - \ell(\mathbf{u})) - W_\ell(\mathbf{u}) + \sum_{t=1}^{d} W_\ell(u_t\mathbf{e}_t)\partial_t\ell(\mathbf{u}) - am(\mathbf{u})\right| \to 0 \quad \text{a.s.}\,.$$

Proof of Corollary 2.
From Lemma 1 and due to the Skorohod construction, almost surely

$$\sqrt{k}(\hat{\ell}_{k,n}(\mathbf{u}) - \ell(\mathbf{u})) - W_\ell(\mathbf{u}) + \sum_{t=1}^{d} W_\ell(u_t\mathbf{e}_t)\partial_t\ell(\mathbf{u}) - am(\mathbf{u}) = o(1)$$

uniformly over $[0,1]^d$. Now, making use again of the recursive definition $\hat{\ell}_{k,n;I}(\mathbf{u}_I) = \mathbb{E}[\hat{\ell}_{k,n}(\mathbf{U})|\mathbf{U}_I = \mathbf{u}_I] - \sum_{I'\subsetneq I} \hat{\ell}_{k,n;I'}(\mathbf{u}_{I'})$ there is no difficulty to make $\hat{\ell}_{k,n;I}$ inherit the asymptotic distribution property from $\hat{\ell}_{k,n}$. The Hoeffding–Sobol decomposition of the process $Z_\ell + am$ makes sense. Indeed, the sample paths of the process in (16) are bounded on $[0,1]^d$ and consequently square integrable because $0 \le \partial_j\ell \le 1$ and the Gaussian process W_ℓ has continuous trajectories almost surely.
Proof of Proposition 3.

This result is obtained by the fact that any superset importance coefficient has an integral representation. Indeed, from (9) in Liu and Owen (2006), we know that

$$D_I^{\text{super}}(\ell) = \frac{1}{2^{|I|}} \int \left(\sum_{J \subseteq I} (-1)^{|I \setminus J|} \ell(\mathbf{u}_J, \mathbf{v}_{-J}) \right)^2 d\mu_J(\mathbf{u}_J) d\mu_{-J}(\mathbf{v}_{-J}) \ .$$

Applying this formula to the empirical version $\hat{\ell}_{k,n}$ and making use of the distributional expansion $\hat{\ell}_{k,n} \simeq \ell + Z_\ell/\sqrt{k} + a(n/k)m$, provided by Lemma 1, allows us to write

$$\begin{aligned} D_I^{\text{super}}(\hat{\ell}_{k,n}) &= D_I^{\text{super}}(\ell) + \frac{1}{k} D_I^{\text{super}}(Z_\ell) + a^2(n/k) D_I^{\text{super}}(m) \\ &+ \frac{1}{2^{|I|}-1} \int \Delta_I[\ell, Z_\ell/\sqrt{k}](\mathbf{u}_I, \mathbf{v}) + \Delta_I[\ell, a(n/k)m](\mathbf{u}_I, \mathbf{v}) + \Delta_I[Z_\ell/\sqrt{k}, a(n/k)m](\mathbf{u}_I, \mathbf{v}) \\ &+ o(1/\sqrt{k}) + o(a^2(n/k)) + o(a(n/k)/\sqrt{k}) \ , \end{aligned}$$

where the integral is with respect to the measure $d\mu_I(\mathbf{u}_I) d\mu(\mathbf{v})$. The result follows from the asymptotic behavior of a, i.e. $a(n/k) \sim a/\sqrt{k}$.

References

T. Browne, J.-C. Fort, B. Iooss, and L. Le Gratiet. Estimate of quantile-oriented sensitivity indices. working paper or preprint, 2017.

J. J. Cai, A.-L. Fougères, and C. Mercadier. Environmental data: multivariate extreme value theory in practice. *Journal de la Société Française de Statistique*, 154(2):178–199, 2013.

G. Chastaing, F. Gamboa, and C. Prieur. Generalized Hoeffding-Sobol decomposition for dependent variables. *Electronic Journal of Statistics*, 6:2420–2448, 2012.

L. de Haan and A. Ferreira. *Extreme value theory. An introduction.* Springer Series in Operations Research and Financial Engineering. New York, NY: Springer., 2006.

B. Efron and C. Stein. The jackknife estimate of variance. *The Annals of Statistics*, 9(3): 586–596, 1981.

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

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

A.-L. Fougères. *Multivariate Extremes, In: Extreme Values in Finance, Telecommunications, and the Environment, ed. Finkenstädt, B. and Rootzén, H.* Monographs on Stat. and Appl. Prob. 99, Chapter 7, pp. 373-388. Chapman and Hall/CRC, 2004.

A.-L. Fougères, L. de Haan, and C. Mercadier. Bias correction in multivariate extremes. *The Annals of Statistics*, 43(2):903–934, 2015.

J. Fruth, O. Roustant, and S. Kuhnt. Total interaction index: A variance-based sensitivity index for interaction screening. *Journal of statistical planning and inference*, 147:212–223, 2014.

E. J. Gumbel. Multivariate distributions with given margins and analytical examples. *Bulletin de l'Institut International de Statistique*, 37(3):363–373, 1960.

W. Hoeffding. A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics*, 19:293–325, 1948.

M. Hofert and K. Hornik. *qrmdata: Data Sets for Quantitative Risk Management Practice*, 2016. R package version 2016-01-03-1.

G. Hooker. Discovering additive structure in black box functions. In *Proceedings of KDD 2004*, pages 575–580. ACM DL, 2004.

X. Huang. Statistics of bivariate extremes. PhD Thesis, Erasmus University Rotterdam, Tinbergen Institute Research series No. 22, 1992.

M. Kereszturi, J. Tawn, and Ph. Jonathan. Assessing extremal dependence of north sea storm severity. *Ocean Engineering*, 118:242 – 259, 2016.

S. Kucherenko and S. Song. Quantile based global sensitivity measures. *ArXiv e-prints*, 2016.

R. Liu and A. B. Owen. Estimating mean dimensionality of analysis of variance decompositions. *Journal of the American Statistical Association*, 101(474):712–721, 2006.

V. Maume-Deschamps and I. Niang. Estimation of quantile oriented sensitivity indices. *Statistics & Probability Letters*, 134:122 – 127, 2018.

L. Mhalla, V. Chavez-Demoulin, and Ph. Naveau. Non-linear models for extremal dependence. *Journal of Multivariate Analysis*, 159:49 – 66, 2017.

T. Muehlenstaedt, O. Roustant, L. Carraro, and S. Kuhnt. Data-driven kriging models based on FANOVA-decomposition. *Statistics and Computing*, 22(3):723–738, 2012.

A. Owen, J. Dick, and S. Chen. Higher order sobol' indices. *Information and Inference: A Journal of the IMA*, 3(1):59–81, 2014.

Y. Peng and W. Ng. Analysing financial contagion and asymmetric market dependence with volatility indices via copulas. *Annals of Finance*, 8(1):49–74, 2012.

G. Pujol, B. Iooss, and A. Janon. *sensitivity: Global Sensitivity Analysis of Model Outputs*, 2017. R package version 1.14.0.

P. Ressel. Homogeneous distributions, and a spectral representation of classical mean values and stable tail dependence functions. *Journal of Multivariate Analysis*, 117:246 – 256, 2013.

J. Segers. Max-stable models for multivariate extremes. *REVSTAT – Statistical Journal*, 10 (1):61–82, 2012.

R. L. Smith. Max-stable processes and spatial extremes. Dept. of Math., Univ. of Surrey, Guildford GU2 5XH, England, 1990.

I. M. Sobol′. Sensitivity estimates for nonlinear mathematical models. *Mathematical Modeling and Computational Experiment. Model, Algorithm, Code*, 1(4):407–414 (1995), 1993.

A. G. Stephenson. evd: Extreme value distributions. *R News*, 2(2), 2002.

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

R. Sueur, B. Iooss, and Th. Delage. Sensitivity analysis using perturbed-law based indices for quantiles and application to an industrial case. working paper or preprint, 2017.

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

J. Tiago de Oliveira. Structure theory of bivariate extremes, extensions. *Estudos de Matematica, Estatistica, e Economicos*, 7:165–195, 1962/63.

A. W. van der Vaart. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998.