Partial Tail-Correlation Coefficient Applied to Extremal-Network Learning

Yan Gong[1], Peng Zhong[1], Thomas Opitz[2], Raphaël Huser[1]

October 13, 2022

Abstract

We propose a novel extremal dependence measure called the partial tail-correlation coefficient (PTCC), which is an analogy of the partial correlation coefficient in the non-extreme setting of multivariate analysis. The construction of the new coefficient is based on the framework of multivariate regular variation and transformed-linear algebra operations. We show how this coefficient allows identifying pairs of variables that have partially uncorrelated tails given the other variables in a random vector. Unlike other recently introduced asymptotic independence frameworks for extremes, our approach requires only minimal modeling assumptions and can thus be used generally in exploratory analyses to learn the structure of extremal graphical models. Thanks to representations similar to traditional graphical models where edges correspond to the non-zero entries of a precision matrix, we can exploit classical inference methods for high-dimensional data, such as the graphical LASSO with Laplacian spectral constraints, to efficiently learn the extremal network structure via the PTCC. The application of our new tools to study extreme risk networks for two datasets extracts meaningful extremal structures and allows for relevant interpretations. Specifically, our analysis of extreme river discharges observed at a set of monitoring stations in the upper Danube basin shows that our proposed method is able to recover the true river flow network quite accurately, and our analysis of historical global currency exchange rate data reveals interesting insights into the dynamical interactions between major economies during critical periods of stress.

Keywords: Graphical LASSO; Multivariate regular variation; Network structure learning; Partial tail dependence; Tail dependence.

[1]Statistics Program, Computer, Electrical and Mathematical Sciences and Engineering (CEMSE) Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia. E-mails: yan.gong@kaust.edu.sa; peng.zhong@kaust.edu.sa; raphael.huser@kaust.edu.sa.

[2]INRAE, Biostatistics and Spatial Processes, Avignon, France. E-mail: thomas.opitz@inrae.fr

1 Introduction

Characterizing the extremal dependence of complex stochastic processes (e.g., in spatial, temporal, spatio-temporal settings) is fundamental for both statistical modeling and applications, such as risk assessment in environmental and financial contexts. Important applications include the modeling of precipitation extremes (Huser and Davison, 2014; Opitz *et al.*, 2018; Bacro *et al.*, 2020; Saunders *et al.*, 2021), heatwaves (Winter and Tawn, 2016; Richards and Wadsworth, 2021; Zhong *et al.*, 2022), and air pollution (Vettori *et al.*, 2019, 2020), as well as financial risk assessment (Bassamboo *et al.*, 2008; Ferro and Stephenson, 2011; Larsson and Resnick, 2012; Marcon *et al.*, 2016; Bekiros and Uddin, 2017; Yan *et al.*, 2019; Gong and Huser, 2021).

Models for extremal dependence traditionally rely on asymptotic frameworks, such as max-stable processes for block maxima or r-Pareto processes for threshold exceedances of a summary functional of the process over a high threshold. Recently, more advanced models have been proposed to further improve flexibility, especially towards modeling of asymptotically independent data with depencence vanishing at the most extreme levels, such as inverted max-stable processes (Wadsworth and Tawn, 2012), max-mixture models (Ahmed *et al.*, 2020), random scale mixtures (Engelke *et al.*, 2019; Wadsworth *et al.*, 2017; Huser and Wadsworth, 2019), max-infinitely divisible processes (Bopp *et al.*, 2021), and conditional spatial extremes models (Wadsworth and Tawn, 2019); for a comprehensive review, see Huser and Wadsworth (2022). Specifics of serial extremal dependence have been studied by Davis *et al.* (2013), among others.

In the study of stochastic dependence structures, networks and graphs are natural tools to represent dependence relationships in multivariate data. Conditional independence, sparsity, and parsimonious representations are key concepts in graph-based approaches for random vectors. Recently, graph-based tools have also been developed for extremal dependence, where variants of conditional independence apply for variables not directly connected by the edges of the graph. For example, Huang *et al.* (2019) provide an exploratory tool, called the χ-*network,* for modeling extremal dependence, and they use it to analyze maximum precipitation during the hurricane season at the United States (US) Gulf Coast and in surrounding areas; Engelke and Hitz (2020) introduce a notion of *conditional independence* adapted to multivariate Pareto distributions

arising for limiting multivariate threshold exceedances, and use it to develop parametric graphical models for extremes based on the Hüsler–Reiss distribution; Gissibl (2018) and Klüppelberg and Krali (2021) propose max-linear model-based frameworks for modeling maxima on tree-like supports; Tran *et al.* (2021) propose QTree, a simple and efficient algorithm to solve the "Latent River Problem" for the important case of extremes on trees. In a similar setting, Engelke and Volgushev (2020) develop a data-driven methodology for learning the graphical structure in the setting of Engelke and Hitz (2020), whereas Röttger *et al.* (2021) further propose Hüsler–Reiss graphical models under the assumption of multivariate total positivity of order two (MTP_2), which allows estimating sparse graphical structures; finally, Engelke and Ivanovs (2021) review the recent developments on sparse representations, dimension reduction approaches, and graphical models for extremes. Overall, existing graphical representations for extremes from the literature often rely on rather stringent asymptotically justified models, sometimes leading to issues when dealing with relatively high-dimensional problems or when specific graph structure assumptions (e.g., trees) are required.

By contrast, rather than restricting ourselves to a strict modeling framework, we adopt a more pragmatic and empirical approach. Specifically, our goal is to extend and enrich existing approaches by defining the new concept of *partial tail correlation* as an extreme-value analog of the notion of partial correlation widely used in classical multivariate analysis, and by introducing a new coefficient that enables the estimation of general extremal networks under minimal modeling assumptions. In the same way that correlation does not imply independence in general, our concept of *partial tail-uncorrelatedness* is a weaker assumption than *conditional tail independence*. However, we shall show that it still provides relevant insights into various forms of extremal dependence structures and helps in guiding modeling choices at a data exploratory stage.

As a novel extremal dependence measure, we propose the *partial tail correlation coefficient (PTCC)* as an equivalent of the partial correlation coefficient in the non-extreme setting. In the classical setting, the Pearson correlation coefficient between two random variables can give misleading interpretations when there are confounding variables that influence both variables, whereas the partial correlation coefficient measures the residual degree of association between two variables after the linear effects of a set of other variables

have been removed. To compute the partial correlation between two variables of interest, we regress each of these variables onto the set of covariates given by all the other variables in the multivariate random vector, and then compute the correlation between the residuals from the two fitted linear regressions. In the Gaussian setting, a partial correlation of zero is equivalent to conditional independence between two variables (Lawrance, 1976; Baba *et al.*, 2004), and the elements of the inverse of the covariance matrix (i.e., the *precision matrix*) are known to fully characterize this conditional dependence structure.

In the extreme-value setting, Cooley and Thibaud (2019) propose the tail pairwise dependence matrix (TPDM) based on transformed-linear operations, which preserves multivariate regular variation in the positive orthant, a fundamental characterization of extremal dependence. The TPDM can be thought of as an analogy of the classical covariance matrix but tailored for multivariate extremes. It summarizes information about extremal dependence and has suitable properties for decompositions such as principal component analysis. In some follow-up work, Mhatre and Cooley (2020) developed non-negative regularly varying time series models with autoregressive moving average (ARMA) structure using the transformed-linear operations for time series extremes. For spatial extremes, Fix *et al.* (2021) extended the simultaneous autoregressive (SAR) model under the transformed-linear framework and developed an estimation method to minimize the discrepancy between the TPDM of the fitted model and an empirically estimated TPDM. Furthermore, Lee and Cooley (2021) introduced transformed-linear prediction methods for extremes. Similar to covariances, the entries of the TPDM are tail dependence measures suitable to show the direct extremal dependence structures without removing the influence of other confounding variables. However, just as the covariance matrix does not reflect partial correlations, the TPDM does not directly inform us about partial associations among extremes.

Given the inverse of a TPDM matrix, denoted by $Q = \Sigma^{-1}$ where Σ is the TPDM of random vector $\boldsymbol{X}$, we define our new partial tail correlation coefficient, the PTCC, in such a way that the (i,k)th entry of Q, denoted by Q_{ik}, equals zero if and only if the PTCC between the two components X_i and X_k from $\boldsymbol{X}$ is also equal to zero. In other words, *partial tail-uncorrelatedness* can be read off from the zero elements of the

matrix Q, similarly to classical Gaussian graphical models. We then exploit this property to define a new class of extremal graphical models based on the PTCC, and then use efficient inference methods to learn the extremal network structure from high-dimensional data based on state-of-the-art techniques from graph theory (e.g., the graphical LASSO, etc.). Here, our focus is on studying undirected graph structures, which is different from causal inference, where causal relationships can be encoded using directed graph edges.

The remainder of this article is organized as follows. In Section 2, we first review necessary background on multivariate regular variation and transformed-linear algebra, as introduced in Cooley and Thibaud (2019). Then, we define the new PTCC and the related notion of partial tail-uncorrelatedness. The properties of our new tool are illustrated through a small numerical example. In Section 3, we present methods for learning general extremal network structures from the PTCC in a high-dimensional data setting, and we discuss two particularly appealing approaches, namely the graphical LASSO and Laplacian spectral constraint-based methods. Section §4 presents a simulation study for general structured undirected graphs using the above two inference methods. In Section 5, we apply these new tools to explore the risk networks formed by river discharges observed at a collection of monitoring stations in the upper Danube basin, and by historical global currency exchange rate data from different historical periods, covering different economic cycles, the COVID-19 pandemic, and the 2022 military conflict in Ukraine.

2 Transformed-linear algebra for multivariate extremes

To introduce the partial tail correlation coefficient (PTCC) and the related notion of partial-tail uncorrelatedness, we first briefly review the multivariate regular variation framework and the foundations of transformed-linear algebra.

2.1 Regular variation framework and transformed linear algebra

A random vector is multivariate regularly varying (Resnick, 2007) (i.e., jointly heavy-tailed) if its joint tail decays as a power function. Precisely, we say that a p-dimensional random vector $\boldsymbol{X} \in \mathbb{R}_+^p = [0, \infty)^p$ with

$p \in \mathbb{N}$ is *regularly varying* if there exists a sequence $b_n \to \infty$ such that

$$n \Pr(b_n^{-1} \boldsymbol{X} \in \cdot) \xrightarrow{v} \nu_{\boldsymbol{X}}(\cdot), \quad n \to \infty, \tag{1}$$

where $\xrightarrow{v}$ denotes vague convergence to the non-null limit measure $\nu_{\boldsymbol{X}}$, a Radon measure defined on the space $[0,\infty]^p \setminus \{0\}$. This measure has the scaling property $r^\alpha \nu_{\boldsymbol{X}}(rB) = \nu_{\boldsymbol{X}}(B)$ for $r > 0$ and Borel sets $B \subset [0,\infty]^p \setminus \{0\}$, where $\alpha > 0$ is the tail index. For this reason, the measure can be further decomposed into a radial measure and an angular measure $H_{\boldsymbol{X}}$ on the unit sphere $\mathbb{S}_{p-1}^+ = \{\boldsymbol{x} \in \mathbb{R}_+^p : ||\boldsymbol{x}||_2 = 1\}$, such that $\nu_{\boldsymbol{X}}(\{\boldsymbol{x} \in [0,\infty]^p \setminus \{0\} : \|\boldsymbol{x}\|_2 \geq r,\ \boldsymbol{x}/\|\boldsymbol{x}\|_2 \in B_H) = r^{-\alpha} \times H_{\boldsymbol{X}}(B_H)$ for $r > 0$ and Borel subsets B_H of $\mathbb{S}_{p-1}^+$. The normalizing sequence b_n is not uniquely determined but must satisfy $b_n = L(n) n^{1/\alpha}$, where $L(n)$ is a slowly varying function (at infinity), i.e., $L(n) > 0$ and $L(rn)/L(n) \to 1$ for any $r > 0$, as $n \to \infty$. We use short notation $\boldsymbol{X} \in \mathrm{RV}_+^p(\alpha)$ for a regularly varying vector $\boldsymbol{X}$ with tail index α.

Cooley and Thibaud (2019) introduced the *transformed-linear algebra framework* to construct an inner product space on an open set (the so-called *target space*) via a transformation, where the distribution of the random vector $\boldsymbol{X}$ has support within this set. Our main use will concern transformation towards the target space $\mathbb{R}_+^p$ from the space $\mathbb{R}^p$, but we first present the general approach. Let t be a bijective transformation from $\mathbb{R}$ onto some open set $\mathbb{X} \subset \mathbb{R}$, and let t^{-1} be its inverse. For a p-dimensional vector $\boldsymbol{y} \in \mathbb{R}^p$, we define $\boldsymbol{x} = t(\boldsymbol{y}) \in \mathbb{X}^p$ componentwise. Then, arithmetic operations among elements of the target space are carried out in the space $\mathbb{R}^p$ before transforming back to the target space. We define vector addition in $\mathbb{X}^p$ as $\boldsymbol{x}_1 \oplus \boldsymbol{x}_2 = t\{t^{-1}(\boldsymbol{x}_1) + t^{-1}(\boldsymbol{x}_2)\}$; we define scalar multiplication with a factor $a \in \mathbb{R}$ as $a \circ \boldsymbol{x} = t\{a t^{-1}(\boldsymbol{x})\}$. The additive identity in $\mathbb{X}^p$ is set to $\mathbf{0}_{\mathbb{X}^p} = t(\mathbf{0})$, and the additive inverse of $\boldsymbol{x} \in \mathbb{X}^p$ is given as $\ominus \boldsymbol{x} = t\{-t^{-1}(\boldsymbol{x})\}$. A valid inner product between two elements $\boldsymbol{x}_1 = (x_{1,1}, \ldots, x_{1,p})^T, \boldsymbol{x}_2 = (x_{2,1}, \ldots, x_{2,p})^T \in \mathbb{X}^p$ from the target space is then obtained by applying the usual scalar product in $\mathbb{R}^p$, i.e., we set $\langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle = \sum_{j=1}^p t^{-1}(x_{1,j}) t^{-1}(x_{2,j})$. To obtain an inner product space on the positive orthant for which arithmetic operations have a negligible effect on large values, and therefore preserve multivariate regular variation, we follow Cooley and Thibaud (2019) and define the specific transformation $t : \mathbb{R} \mapsto (0,\infty)$ given by

$$t(y) = \log\{1 + \exp(y)\}.$$

We have $y/t(y) \to 1$ as $y \to \infty$, such that the upper tail behavior of a random vector $\boldsymbol{Y} = t(\boldsymbol{X})$ is preserved through t. For lower tails, we have $\exp(y)/t(y) \to 1$ as $y \to -\infty$. The inverse transformation is $t^{-1}(x) = \log(\exp(x) - 1)$, $x > 0$. Algebraic operations done in the vector space induced the above transformation t are commonly called *transformed-linear operations*, and we can exploit this framework to extend classical linear algebra methods (e.g., principal component analysis, etc.) to the multivariate extremes setting, where vectors and models are often conveniently expressed on the positive orthant, $\mathbb{R}_+^p$.

2.2 Inner product space of regularly varying random variables

With transformed-linear operations, we can use a vector of independent and identically distributed (i.i.d.) random variables to construct new regularly varying random vectors on the positive orthant that possess tail dependence. Suppose that $\boldsymbol{Z} = (Z_1, \ldots, Z_q)^T \geq 0$ is a vector of $q \in \mathbb{N}$ i.i.d. regularly varying random variables with tail index α, such that there exists a sequence $\{b_n\}$ that yields

$$n \Pr(Z_j > b_n z) \to z^{-\alpha}, \quad n \Pr\{Z_j \leq \exp(-k b_n)\} \to 0, \quad k > 0, \quad j = 1, \ldots, q,$$

where the first condition is equivalent to regular variation (1) in dimension $p = 1$. The random vector $\boldsymbol{Z}$ with independent components has limit measure of multivariate regular variation characterized by $\nu_{\boldsymbol{Z}}\left\{[\boldsymbol{0}, \boldsymbol{z}]^C\right\} = \sum_{j=1}^q z_j^{-\alpha}$ for $\boldsymbol{z} = (z_1, \ldots, z_q)^T > \boldsymbol{0}$. Then, we can construct new regularly varying p-dimensional random vectors $\boldsymbol{X} = (X_1, \ldots, X_p)^T$ via a transformed-linear matrix product with using a deterministic matrix $A = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_q) \in \mathbb{R}_+^{p \times q}$, with columns $\boldsymbol{a}_j \in \mathbb{R}_+^p$, as follows:

$$\boldsymbol{X} = \bigoplus_{j=1}^{q} \boldsymbol{a}_j \circ Z_j. \tag{2}$$

We write $\boldsymbol{X} = A \circ \boldsymbol{Z} \in \mathrm{RV}_+^p(\alpha)$. This construction ensures that the multivariate regular variation property is preserved with the same index α (Corollary 1, Cooley and Thibaud, 2019). Furthermore, we require A to have full row-rank. Based on the construction (2), it is possible to define a (different) inner product space spanned by the random variables obtained by transformed-linear operations on $\boldsymbol{Z}$, where some but not all of the components of $\boldsymbol{a}_j$ are further allowed to be non-positive. Following Lee and Cooley (2021), an inner product of $\langle X_i, X_k \rangle$ on the space spanned by all possible transformed-linear combinations of the elements of

the random vector $\boldsymbol{X}$ constructed as in (2), may be defined as follows:

$$\langle X_i, X_k \rangle = \sum_{j=1}^{q} a_{ij} a_{kj},$$

where a_{ij} refers to the entry in row i of the column j of the matrix A for $i \in \{1, \ldots, p\}$ and $j \in \{1, \ldots, q\}$, and the corresponding norm becomes $||X|| = \sqrt{\langle X, X \rangle}$. The metric induced by the inner product is $d(X_i, X_k) = ||X_i \ominus X_k|| = [\sum_{j=1}^{q} (a_{ij} - a_{kj})^2]^{1/2}$, for $i, k = 1, \ldots, p$.

2.3 Generality of the framework

In practice, given a random vector $\boldsymbol{X}$ for which we assume $\boldsymbol{X} \in \mathrm{RV}_+^p(\alpha)$, we will further assume that it allows for a stochastic representation as in (2). Since the constructions of type (2) form a dense subclass of the class of multivariate regularly varying vectors (if q is not fixed but $q \to \infty$), this assumption is not restrictive; see Fougères *et al.* (2013); Cooley and Thibaud (2019).

Thanks to their flexibility, representations akin to the transformed-linear random vectors in (2) have recently found widespread interest in statistical learning for extremes. The fundamental model structure used in the causal discovery framework for extremes developed by Gnecco *et al.* (2021) is essentially based on a variant of (2). In the setting of max-linear models, in particular the graphical models of Gissibl (2018), we can use (2) to construct random vectors $\boldsymbol{X}$ possessing the same limit measure $\nu_{\boldsymbol{X}}$ as the max-linear vectors. Finally, low-dimensional representations of extremal dependence in random vectors obtained through variants of the k-means algorithm can be shown to be equivalent to the extremal dependence induced by construction (2); see Janßen and Wan (2020).

2.4 Tail pairwise dependence matrix

The tail pairwise dependence matrix (TPDM, Cooley and Thibaud, 2019) is defined to summarize the pairwise extremal dependence of a regularly varying random vector using the second-order properties of its angular measure. Let $\alpha = 2$, which ensures desirable properties; in practice, this condition can be ensured through appropriate marginal pretransformation of data (Cooley and Thibaud, 2019). Then, the TPDM

$\Sigma_{\boldsymbol{X}}$ of $\boldsymbol{X} \in \mathrm{RV}_+^p(2)$ is defined as follows:

$$\sigma_{ik} := \int_{\Theta_{p-1}^+} w_i w_k \mathrm{d}H_{\boldsymbol{X}}(w), \quad \Sigma_{\boldsymbol{X}} = (\sigma_{\boldsymbol{X}_{ik}})_{i,k=1,\dots,p},$$

where $H_{\boldsymbol{X}}$ is the angular measure on $\Theta_{p-1}^+ = \{\boldsymbol{w} \in \mathbb{R}_+^p : ||\boldsymbol{w}||_2 = 1\}$ as introduced before. The matrix $\Sigma_{\boldsymbol{X}}$ is an extreme-value analogue of the covariance matrix, and it has similar useful properties. It is positive semi-definite and completely positive, i.e., for any $q \leq p$ there exists a finite $p \times q$ matrix A such that $\Sigma_{\boldsymbol{X}} = AA^T$. The matrix A is not unique, and it is always possible to choose A to have only non-negative entries. Specifically, for random vectors $\boldsymbol{X}$ obtained by the transformed-linear construction (2), the entries of $\Sigma_{\boldsymbol{X}}$ correspond to the values of the inner product $\sigma_{ij} = \langle X_i, X_j \rangle$. In the following, we further assume that $\Sigma_{\boldsymbol{X}}$ is positive definite, which guarantees the existence of the inverse matrix of the TPDM.

The special case where $\sigma_{ik} = 0$ is equivalent to asymptotically independent components X_i and X_k, meaning that the conditional exceedance probability $\Pr(F_{X_i}(X_i) > u \mid F_{X_k}(X_k) > u)$ tends to 0 as the quantile u tends to 1 (Sibuya, 1960; Ledford and Tawn, 1996).

We can construct new transformed-linear random vectors that have the same TPDM as a given random vector $\boldsymbol{X}$. Therefore, we use the construction (2) by applying A (with $AA^T = \Sigma_{\boldsymbol{X}}$) to a random vector $\boldsymbol{Z} \in \mathrm{RV}_+^p(2)$ with independent components. Depending on the choice of q, the angular measure of the new random vector can be arbitrarily close to that of $\boldsymbol{X}$ thanks to the denseness property of discrete angular measures. While A is not unique, we emphasize that it is very important that the inner product depends only on the entries of the matrix $\Sigma_{\boldsymbol{X}} = AA^T$, such that the specific choice of A does not matter.

An estimator of the TPDM was proposed by Cooley and Thibaud (2019, §7.1). For an i.i.d. sequence of vectors $\boldsymbol{x}_t$, $t = 1, \dots, n_{\text{samp}}$, i.e., of observations from a random vector $\boldsymbol{X} \in \mathrm{RV}_+^p(2)$, define

$$\widehat{\sigma}_{ik} = \widehat{m} \int_{\Theta_{p-1}^+} w_i w_k \mathrm{d}\widehat{N}_{\boldsymbol{X}}(w) = \widehat{m}\, n_{\text{ext}}^{-1} \sum_{t=1}^{n_{\text{samp}}} w_{ti} w_{tk} \mathbb{1}(r_t > r_0), \tag{3}$$

where $r_t = ||\boldsymbol{x}_t||_2$, $w_t = \boldsymbol{x}_t / r_t$, r_0 is a high threshold for the radial component, $n_{\text{ext}} = \sum_{t=1}^{n_{\text{samp}}} \mathbb{1}(r_t > r_0)$, and the probability measure $N_{\boldsymbol{X}}(\cdot) = m^{-1} H_{\boldsymbol{X}}(\cdot)$ (with $m = H_{\boldsymbol{X}}(\Theta_{p-1}^+)$) obtained by normalizing $H_{\boldsymbol{X}}$. Here, $\widehat{m}$ is an estimate of $H_{\boldsymbol{X}}(\Theta_{p-1}^+)$; when the data is preprocessed to have a common unit scale, we can set $m = p$

and there is no need to estimate it. The estimator (3) was discussed by Larsson and Resnick (2012) in the bivariate case.

2.5 Partial tail correlation coefficient (PTCC)

When we remove two components X_i and X_k from the random vector $\boldsymbol{X}$, the remaining $(p-2)$-dimensional vector is denoted as $\boldsymbol{X}_{\text{-}ik}$, and we write $A_{\text{-}ik}$ for the matrix without its i-th and k-th columns. In addition, we write vectors in the form $\boldsymbol{X}_{ik} = (X_i, X_k)^T$ and re-order the columns of $\boldsymbol{X}$ as $\boldsymbol{X}' = (\boldsymbol{X}_{ik}^T, \boldsymbol{X}_{\text{-}ik}^T)^T = (A_{ik}, A_{\text{-}ik}) \circ \boldsymbol{Z}$. The best transformed-linear predictor of $\boldsymbol{X}_{ik}$ given $\boldsymbol{X}_{\text{-}ik}$ is given as $\widehat{\boldsymbol{X}}_{ik} = \boldsymbol{B} \circ \boldsymbol{X}_{\text{-}ik} = (\widehat{X}_{ij}^i, \widehat{X}_{ij}^j)^T$, where $\boldsymbol{B} = (\boldsymbol{b}_1, \boldsymbol{b}_2)^T$ is a $2 \times (p-2)$ matrix with $\boldsymbol{b}_1, \boldsymbol{b}_2$ chosen such that $d(X_i, \widehat{X}_{ij}^i)$ and $d(X_j, \widehat{X}_{ij}^j)$ attain their minimum, respectively. Suppose that the TPDM of $\boldsymbol{X}'$ is of the form

$$\Sigma_{\boldsymbol{X}'} = \begin{bmatrix} \Sigma_{hh} & \Sigma_{hl} \\ \Sigma_{lh} & \Sigma_{ll} \end{bmatrix}, \text{ and } \Sigma_{hh} = \text{TPDM}(\boldsymbol{X}_{ik}) = \begin{bmatrix} \Sigma_{ii} & \Sigma_{ik} \\ \Sigma_{ki} & \Sigma_{kk} \end{bmatrix},$$

where $\Sigma_{hl} = \Sigma_{lh}^T = (\Sigma_{il}, \Sigma_{kl})^T$, $i, k, l = 1, \ldots, p, l \neq i, l \neq k$. Then, based on the projection theorem for vector spaces with inner products, we have

$$\widehat{\boldsymbol{B}} = \Sigma_{hl}\Sigma_{ll}^{-1}.$$

Straightforward calculations show that the prediction error $\boldsymbol{e} = \boldsymbol{X}_{ik} \ominus \widehat{\boldsymbol{X}}_{ik}$ has the following TPDM:

$$\text{TPDM}(\boldsymbol{e}) = \Sigma_{hh} - \Sigma_{hl}\Sigma_{ll}^{-1}\Sigma_{lh} = \begin{bmatrix} \Sigma_{ii} - \Sigma_{il}\Sigma_{ll}^{-1}\Sigma_{li} & \Sigma_{ik} - \Sigma_{il}\Sigma_{ll}^{-1}\Sigma_{lk} \\ \Sigma_{ki} - \Sigma_{kl}\Sigma_{ll}^{-1}\Sigma_{li} & \Sigma_{kk} - \Sigma_{kl}\Sigma_{ll}^{-1}\Sigma_{lk} \end{bmatrix}. \tag{4}$$

Definition 1. *The* partial tail correlation coefficient (PTCC) *of two random variables X_i and X_k is defined as the off-diagonal TPDM coefficient of the bivariate residual vector $\boldsymbol{e}$ in (4), such that transformed-linear dependence with respect to all other random variables is removed.*

Definition 2. *Let $\boldsymbol{X}_{ik} = (X_i, X_k)^T$ and $\boldsymbol{X}_{\text{-}ik}$ be defined as above. Given $\boldsymbol{X}_{\text{-}ik}$, we say that X_i and X_k are* ***partially tail-uncorrelated*** *if the PTCC of X_i and X_k (given $\boldsymbol{X}_{\text{-}ik}$) is equal to zero, i.e., if $\Sigma_{ki} - \Sigma_{kl}\Sigma_{ll}^{-1}\Sigma_{li} = 0$ as defined in (4).*

Remark 1: Thanks to the properties of TPDMs, the residuals of two partially tail-uncorrelated random variables are necessarily asymptotically independent.

The following proposition links tail-uncorrelatedness to the entries of the inverse of the TPDM of $\boldsymbol{X}$.

Proposition 3. *Given the representation of the TPDM of $\boldsymbol{X}'$ as a 3×3 block matrix as follows,*

$$\Sigma_{\boldsymbol{X}'} = TPDM\{(X_i, X_k, \boldsymbol{X}_{-ik}^T)^T\} = \begin{bmatrix} \Sigma_{ii} & \Sigma_{ik} & \Sigma_{il} \\ \Sigma_{ki} & \Sigma_{kk} & \Sigma_{kl} \\ \Sigma_{li} & \Sigma_{lk} & \Sigma_{ll} \end{bmatrix}, \tag{5}$$

where the dimensions of the submatrices are as follows: Σ_{il} and Σ_{kl} are $(p-2) \times 1$, Σ_{ll} is $(p-2) \times (p-2)$, then the following two statements are equivalent:

$$(1)\ \Sigma_{ki} - \Sigma_{kl}\Sigma_{ll}^{-1}\Sigma_{li} = 0, \qquad (2)\ (\Sigma^{-1})_{ik} = 0.$$

This result is given in Proposition 1 in Speed and Kiiveri (1986), see the Appendix A for more details. The following corollary is a direct consequence.

Corollary 4. *Denote the inverse matrix of the TPDM of a random vector $\boldsymbol{X}$ by $Q = \Sigma^{-1}$. Then,*

$$Q_{ik} = 0 \quad \textit{iff} \quad PTCC_{ik} = 0,$$

where $PTCC_{ik}$ is the PTCC of components X_i and X_k. Recall that a PTCC equal to zero corresponds to partial tail-uncorrelatedness.

3 Learning extremal networks for high-dimensional extremes

Using PTCCs, we now explore the (partial) extremal dependence structure of multivariate random variables under the framework of multivariate regular variation. In this section, we define new graphical models to represent extremal dependence for extremes. Thanks to the transformed-linear framework exposed in the previous section, we can proceed as for classical graphical models by replacing the classical covariance matrix with the TPDM.

3.1 Graphical models for extremes

Let $G = (V, E)$ be a graph, where $V = \{1, \ldots, d\}$ represents the node set and $E \in V \times V$ the edge set. We call G an undirected graph if for two nodes $i, j \in V$ the edge (i, j) is in E if and only if the edge (j, i) is

also in E. We show how to estimate graphical structures for extremes for any type of undirected graph in which we have no edge (i, j) if and only if the variables X_i and X_j are partially tail-uncorrelated given all the other variables in the graph, written as

$$X_i \perp_p X_j \mid \boldsymbol{X}_{\setminus\{i,j\}}.$$

Our methods work for general undirected graph structures, including trees, decomposable graphs and non-decomposable graphs, see the example illustrations in Figure 1. Note, however, that our general method cannot restrict the estimated graph to be of a specific type, such as a tree.

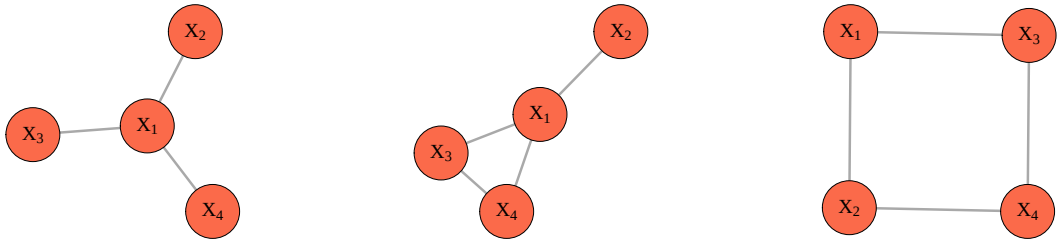

Figure 1: Examples of undirected graph structures: a tree, a decomposable graph, and a non-decomposable graph (from the left to the right).

3.2 Sparse representation of high-dimensional extreme networks

For high-dimensional extremes with a relatively large number of components p (e.g., up to tens and hundreds of variables), a graphical representation of the extremal dependence structure is desirable for reasons of parsimony and interpretability. We now introduce two efficient inference methods to learn extremal networks from high-dimensional data via PTCCs based on two state-of-the-art graphical methods: the extremal graphical Lasso, and structured graph learning method via Laplacian spectral constraints. These two methods both work efficiently in high-dimensional settings and return an estimation of the underlying extremal dependence with sparse structure, i.e., with the cardinality of E being of the same order as the one of V.

3.2.1 Extremal graphical Lasso

Given an empirical estimator $\widehat{\Sigma}$ of the TPDM, and a tuning parameter $\lambda \geq 0$, the optimization carried out by the method of extremal graphical Lasso is expressed as follows (Friedman *et al.*, 2008):

$$\widehat{\Theta}_\lambda = \arg\max_{\Theta \succeq 0} \left[\log\det\Theta - \mathrm{tr}(\widehat{\Sigma}\Theta) - \lambda \sum_{i \neq j} |\Theta_{ij}| \right]$$

where $\succeq$ indicates positive-semidefiniteness and $\widehat{\Theta}_\lambda$ contains the information on the extremal graph structure. Larger λ enforces a larger proportion of zeros in $\widehat{\Theta}_\lambda$ and hence fewer edges in the graph. Choosing an appropriate value for λ is important. On the one hand, we want to enforce sparsity in the graph, where only significant connections are maintained in the network. On the other hand, $\widehat{\Theta}_\lambda$ should be well-defined, with estimation being stable, and with meaningful dependence structures in the estimated model.

3.2.2 Structured graph learning via Laplacian spectral constraints

As an alternative to the Lasso approach, we can seek to include more structural information into the graph by using the structured graph learning method (Kumar *et al.*, 2019), known as structured graph Laplacian (SGL). It allows us to precisely control the sparsity and to preserve the connectedness thanks to constraints on the eigenvalues of the graph Laplacian operator that encodes the graph structure. For instance, if exactly one eigenvalue is zero and all other eigenvalues are positive, then the graph is connected. Laplacian matrix estimation can be formulated as the estimation problem for a precision matrix $\boldsymbol{Q}$, thus connecting to our framework of the PTDM and its inverse. For any vector of eigenvalues $\boldsymbol{\lambda} \in S_{\boldsymbol{\lambda}}$ with appropriate a priori constraints for the desired graph structure defined through the set of admissible eigenvalues $S_{\boldsymbol{\lambda}}$, we set $Q = \mathcal{L}\boldsymbol{w}$ with $\mathcal{L}$ the linear operator that maps a non-negative set of edge weights $\boldsymbol{w}$ into the Laplacian matrix Q. The Laplacian matrix $\boldsymbol{Q}$ can be factorized as $\boldsymbol{Q} = U\boldsymbol{\lambda}U^T$ (with an orthogonal matrix U) to enforce the constraints on $\boldsymbol{\lambda}$. Then the optimization problem can be formulated as follows:

$$(\widehat{\boldsymbol{\lambda}}, \widehat{U}) = \arg\max_{\boldsymbol{\lambda}, U} \max_{\boldsymbol{w}} \left(\log \mathrm{gdet}(\mathrm{Diag}(\boldsymbol{\lambda})) - \mathrm{tr}(\widehat{\Sigma}\mathcal{L}\boldsymbol{w}) + \alpha||\mathcal{L}\boldsymbol{w}||_1 + \frac{\beta}{2}||\mathcal{L}\boldsymbol{w} - U\mathrm{Diag}(\lambda)U^T||_F^2 \right),$$

$$\text{subject to } \boldsymbol{w} \geq 0, \boldsymbol{\lambda} \in S_{\boldsymbol{\lambda}}, \text{and } U^T U = I,$$

where $S_{\boldsymbol{\lambda}}$ denotes the set of constrained eigenvalues, $||\cdot||_F$ is the Frobenius norm, and gdet is a generalized determinant defined as the product of all positive values in $\boldsymbol{\lambda}$. The optimization problem can be viewed as penalized likelihood if data have been generated from a Gauss–Markov random field; in more general cases such as ours, it still provides meaningful graphical structures since it can be viewed as a so-called penalized log-determinant Bregman divergence problem. Therefore, this method can be seen as an extension of the graphical Lasso that allows us to set useful spectral constraints with respect to the structure of the graph. A larger value of α increases the sparsity level of the graph. The hyperparameter $\beta \geq 0$ additionally controls the level of connectedness, and a larger value of β enforces a higher level of the connectedness of the estimated graph structure.

4 Simulation study

We present a simulation study with three examples where the corresponding true structures of the extremal dependence graphs are as in Figure 1, i.e., a tree (Case 1), a decomposable graph (Case 2), and a non-decomposable graph (Case 3). The simulated models are constructed as follows. We simulate i.i.d. random variables $R_1, R_2, R_3, R_4 \sim$ Frechét(2) of size $n = 10^5$ for each. We consider $\boldsymbol{X} = (X_1, X_2, X_3, X_4)^T$ according to the following three cases.

Case 1:

$$\begin{cases} X_1 &= R_1 \\ X_2 &= R_1 \oplus R_2 \\ X_3 &= R_1 \oplus R_3 \\ X_4 &= R_1 \oplus R_4 \end{cases}, \quad TPDM(\boldsymbol{X}) = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 \end{bmatrix}, \quad Q(\boldsymbol{X}) = \begin{bmatrix} 4 & -1 & -1 & -1 \\ -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}.$$

Case 2:

$$\begin{cases} X_1 = R_1 \\ X_2 = R_1 \oplus R_2 \\ X_3 = R_1 \oplus R_3 \\ X_4 = R_1 \oplus 2R_3 \oplus R_4 \end{cases}, \quad TPDM(\boldsymbol{X}) = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 3 \\ 1 & 1 & 3 & 6 \end{bmatrix}, \quad Q(\boldsymbol{X}) = \begin{bmatrix} -4 & -1 & 1-3 & 1 \\ -1 & 1 & 0 & 0 \\ -3 & 0 & 5 & -2 \\ 1 & 0 & -2 & 1 \end{bmatrix}.$$

Case 3:

$$\begin{cases} X_1 = R_1 \\ X_2 = R_1 \oplus 3/\sqrt{6}R_2 \\ X_3 = R_1 \oplus 1/\sqrt{6}R_2 \oplus 2/\sqrt{3}R_3 \\ X_4 = R_1 \oplus \sqrt{6}/3R_2 \oplus 1/\sqrt{3}R_3 \oplus R_4 \end{cases}, \quad TPDM(\boldsymbol{X}) = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2.5 & 1.5 & 2 \\ 1 & 1.5 & 2.5 & 2 \\ 1 & 2 & 2 & 3 \end{bmatrix}, \quad Q(\boldsymbol{X}) = \begin{bmatrix} 2 & -0.5 & -0.5 & 0 \\ -0.5 & 1 & 0 & -0.5 \\ -0.5 & 0 & 1 & -0.5 \\ 0 & -0.5 & -0.5 & 1 \end{bmatrix}.$$

We proceed as follows to infer the extremal graph structure in each case. First, we estimate the TPDM of $\boldsymbol{X}$, $\widehat{\Sigma}_{\boldsymbol{X}}$. Then, we apply the extremal graphical Lasso and SGL methods. In each setting, we test $m_1 = 300$ different values for λ when using extremal graphical Lasso and $m_2 = 400$ different settings for the combination of α and β for the SGL method.

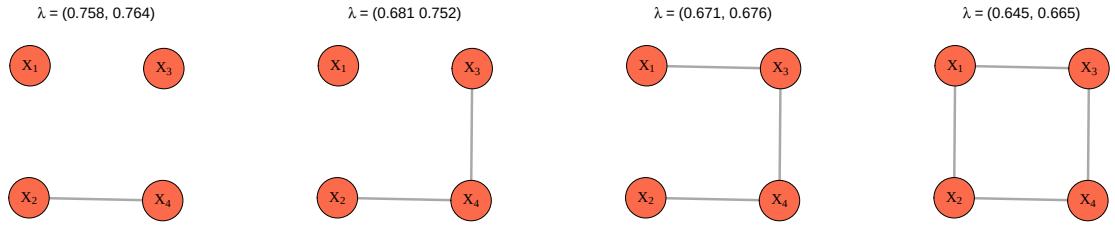

Figure 2: Extremal graph learning using the extremal graphical Lasso method for Case 3.

In all experimental results, when we specify the true number of edges, both methods can retrieve 100% of the true extremal graph structures. There is no universally optimal way of tuning these parameters; however, we can use problem-specific criteria to achieve the desired outcome (and therefore to set the penalty parameters). In Figure 2, we display the extremal graph learning results for Case 3 when using the extremal graphical Lasso method. The heading of each display shows the range of λ, which controls the number of edges in the graph, i.e., the sparsity level. When λ decreases, the number of edges increase, and the method consistently retrieves the true set of edges with 100% accuracy.

5 Applications

Risk networks are useful in quantitative risk management to elucidate complex extremal dependence structures in collections of random variables. We show two examples of both environmental and financial risk analysis. First, we study river discharge data of the upper Danube basin, which has become a benchmark dataset for learning extremal networks in the recent literature. The true underlying physical river flow network is available, which allows comparing the performance of our method with other existing approaches. Second, we apply our method to historical global currency exchange rate data from different historical periods, including different economic cycles, the COVID-19 period, and the period of the 2022 military conflict

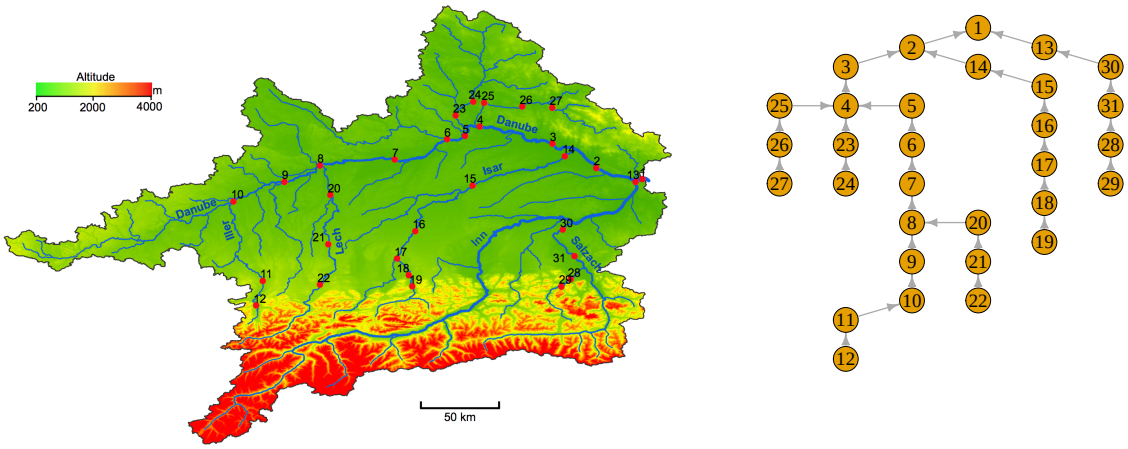

Figure 3: (a) Topographic map of the upper Danube basin, showing 31 sites of gauging stations (red circles) and the altitudes of the region. (b) The true physical river flow connections; the arrows show the flow directions.

opposing Russia and Ukraine (2022.02.24–2022.04.15).

5.1 Extremal network estimation for a river network

We apply our method to study the extreme discharges on the river network of the upper Danube basin (see Figure 3 (a) for the topographic map). This region is historically affected regularly by severe flooding that causes losses of human lives and damage to material goods. The original daily discharge data from 1960 to 2009 were provided by the Bavarian Environmental Agency (http://www.gkd.bayern.de), and Asadi *et al.* (2015) preprocessed the data, which now include $n = 428$ supposedly independent events $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \in \mathbb{R}^d$ on $d = 31$ gauging stations from three summer months (June, July, and August), using declustering methods. The data were also studied by Engelke and Hitz (2020) using graphical models for extremes based on conditional independence notations for multivariate Pareto distributions. The true physical river flow connections and directions can be represented by a directed graph shown in Figure 3 (b) where the arrows show the flow directions.

5.1.1 Graph structure learning using the extremal graphical Lasso

To learn the extremal dependence structure of the river network, we first perform a nonparametric empirical transformation to satisfy Fréchet($\alpha = 2$) margins for each station. In addition, we estimate the TPDM,

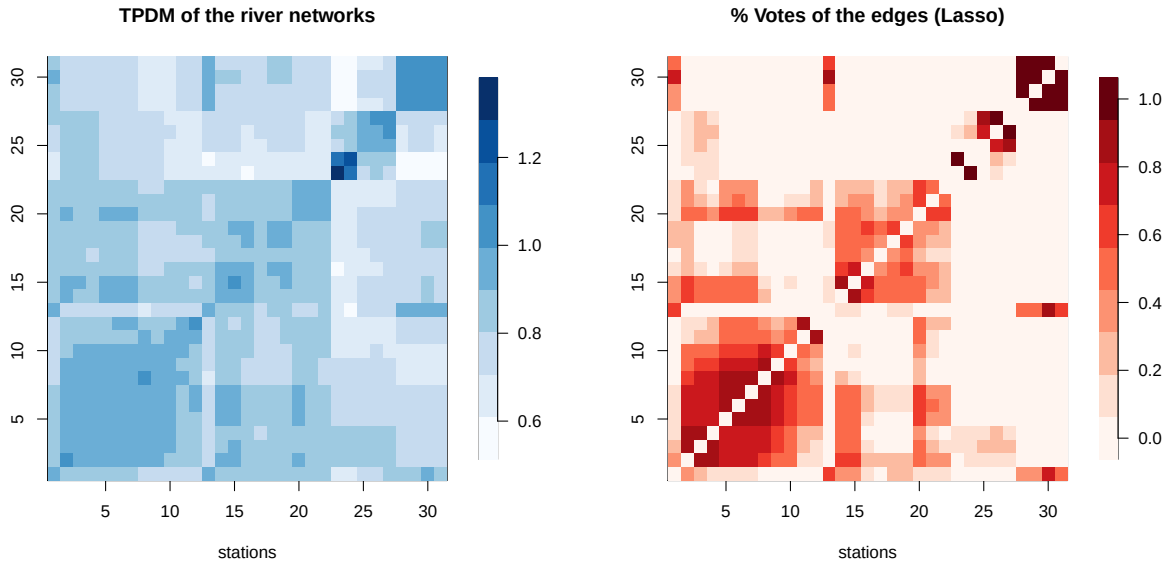

Figure 4: (a) Estimated TPDM of the river discharge from the upper Danube basin. (b) Votes (%) of the edges based on the extremal graphical Lasso method.

$\widehat{\Sigma}_{\boldsymbol{X}}$, using the proposed estimator in (3). In particular, we choose $m = d = 31$ because the margins are preprocessed to have a common unit scale and $r_0 = 19$, which corresponds to the empirical 90% quantile. Therefore, $n_{\text{exc}} = 43$ is the number of large observations which are used to estimate $\widehat{\Sigma}_{\boldsymbol{X}}$. Figure 4 (a) displays the estimated TPDM of the river discharge from the upper Danube basin, and Figure 4 (b) displays the votes (in percentage) of the edges based on the extremal graphical Lasso method.

5.1.2 Graph structure learning using the SGL method

To enhance connectedness, we further explore the estimation of sparse graphs using graph learning methods via spectral constraints. In particular, as we described in §3.2.2, we control the estimation of the graph structure by tuning the sparsity parameter α and the connectedness parameter β. As shown in Figure 5 (a), the sparsity level changes for different combinations of α and β. Figure 5 (b) displays the votes (in percentage) of the edges based on the SGL method. Essentially, we would like to obtain a sparse graph structure while keeping the graph connected. However, there is a tradeoff between these two requirements.

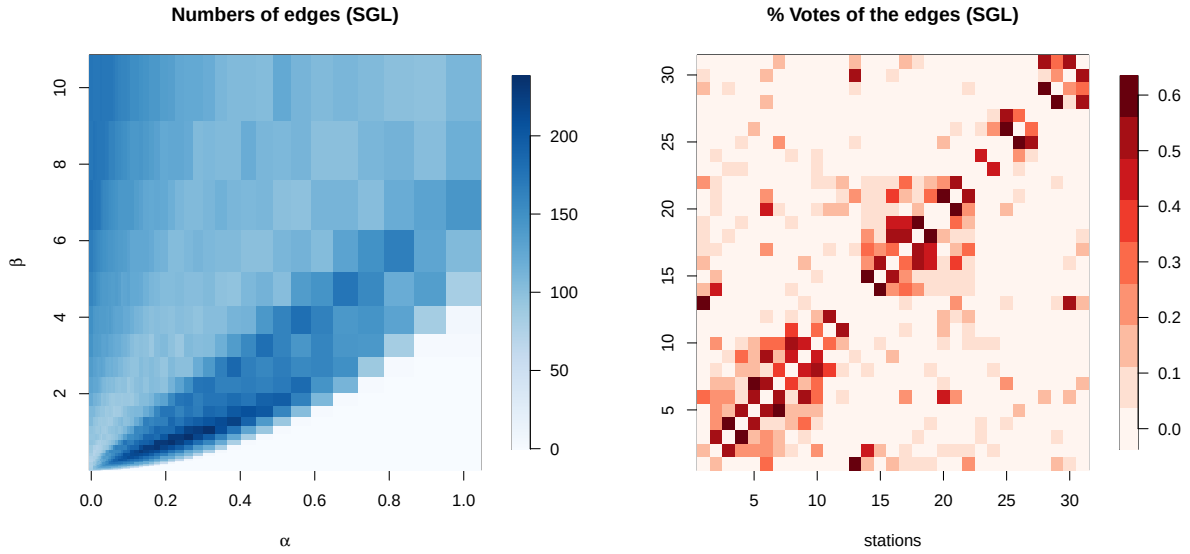

Figure 5: (a) Number of edges under different parameter settings. (b) Votes (%) of the edges based on the SGL method.

5.1.3 Estimated extremal river discharge networks

Figures 6 (a) and (b) display the estimated extremal river discharge network using the extremal graphical Lasso and SGL methods, respectively. We add the connections between nodes based on the rankings of the voting for the edges shown in Figures 4 (b) and 5 (b). The displayed thickness of the edges is proportional to the votes (in percentage). Edges are added until no nodes are left isolated.

When two nodes are not connected, it means they are partially tail-uncorrelated. The estimation based on the extremal graphical Lasso method has some more edges than the true physical river flow networks. The extra links could be interpreted as being due to extremal dependence induced by regional weather events. In addition, the estimation result based on the SGL method matches most of the true river flow networks while showing the strength of the extremal dependence connections. Compared with existing methods (Engelke and Hitz, 2020; Klüppelberg and Krali, 2021), our estimated extremal networks look more easily interpretable since they are closer to the actual flow structure of the river network.

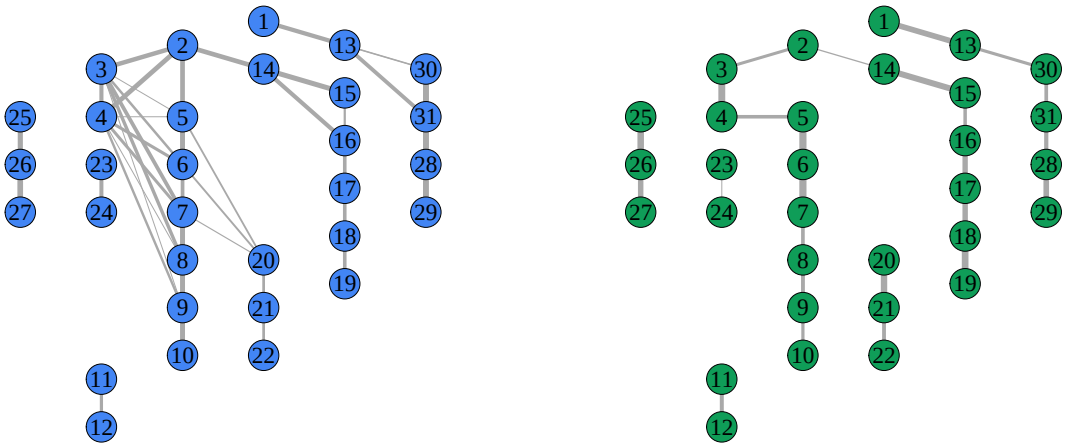

Figure 6: Estimated extremal river discharge network using the (a) extremal graphical Lasso and (b) SGL methods.

5.2 Extremal network estimation for global currency exchange rate network

We apply our method to explore historical global currency exchange rate data for 20 currencies from different historical periods, including two different global economic cycles (2009–2014 and 2015–2019, where the segmentation is based on the world GDP cycles, shown in Figure 9 from Appendix B), COVID-19 (2020.01.01–2022.02.23), and since the beginning of the 2022 military conflict between Russia and Ukraine until a most recent date when we downloaded the data (2022.02.24–2022.09.26). Historical data were downloaded from Yahoo Finance. We chose the currencies from all G20-countries, and in addition also the Ukrainian and the Kazakhstani currencies. The list of the currencies and their corresponding symbols can be found in Table 1 in the Appendix B. The unit of the currencies is US Dollar (USD), so USD is not on the list of considerations.

First, we preprocess the historical daily closing prices of the currencies. An ARMA(1,1)-GARCH(1,1) time series model is fitted to the log return time series of each currency, and then the standardized residuals $z_i, i = 1, \ldots, 20$, are extracted. Thereafter, we flip the sign of z_is such that exchange rate decreases correspond to the upper tail, and we transform them marginally to Frechét margins with shape parameter 2. Therefore,

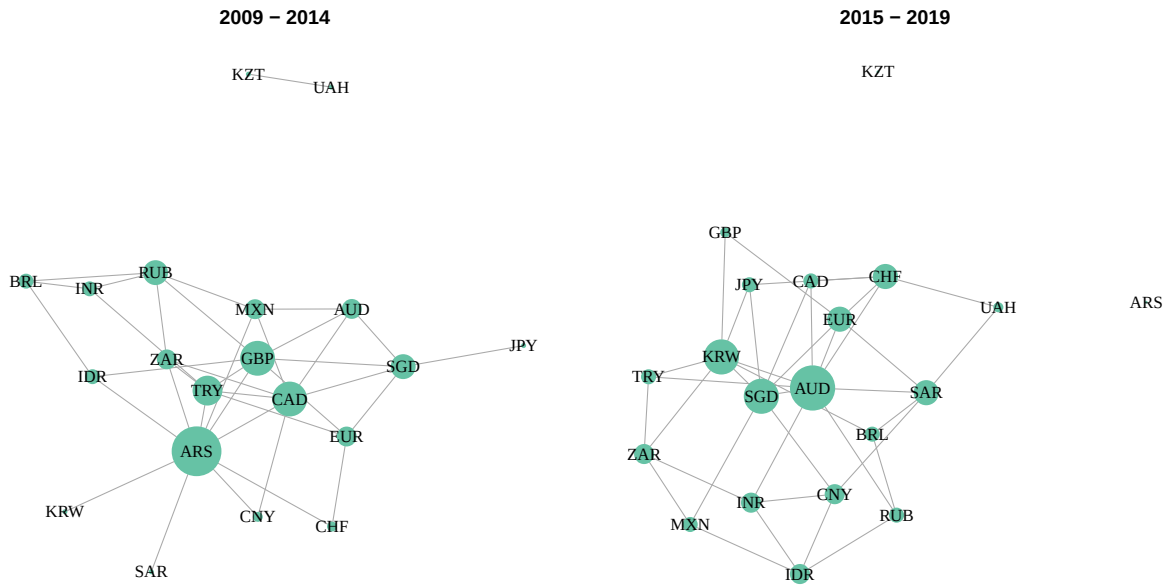

Figure 7: Estimated extremal currency exchange rate risk network (a) (2009–2014) and (b) (2015–2019).

the estimated network represents a risk network by displaying the extremal dependence of the losses of log returns for currencies. We follow the same procedure as before to estimate the corresponding TPDM.

In this analysis, we apply only the SGL method, which includes the extremal graphical Lasso method as a special case when $\beta = 0$. Further, for comparability among different historical periods, we choose the number of edges to be around 38, presenting a sparsity level of 80%. In addition, the size of the node is proportional to the number of connections for each node in Figures 7 and 8. The bigger a node, the stronger connected it is.

Based on the results, we provide some heuristic interpretations of the risk networks, where more connected nodes can be understood to be more vulnerable/exposed to network risks. We can back up some of our findings using historical events and economic-development regimes for different countries. From Figure 7 (a), during the economic cycle 2009–2014, the strongest-connected currencies within the estimated risk network are ARS (Argentina) and CAD (Canada), whereas KZT (Kazakstan) and UAH (Ukraine) are isolated from the major clusters. In Figure 7 (b), during the economic cycle 2015–2019, the strongest-connected currencies from the estimated risk network are AUD, SGD and KRW, whereas ARS and KZT are relatively isolated.

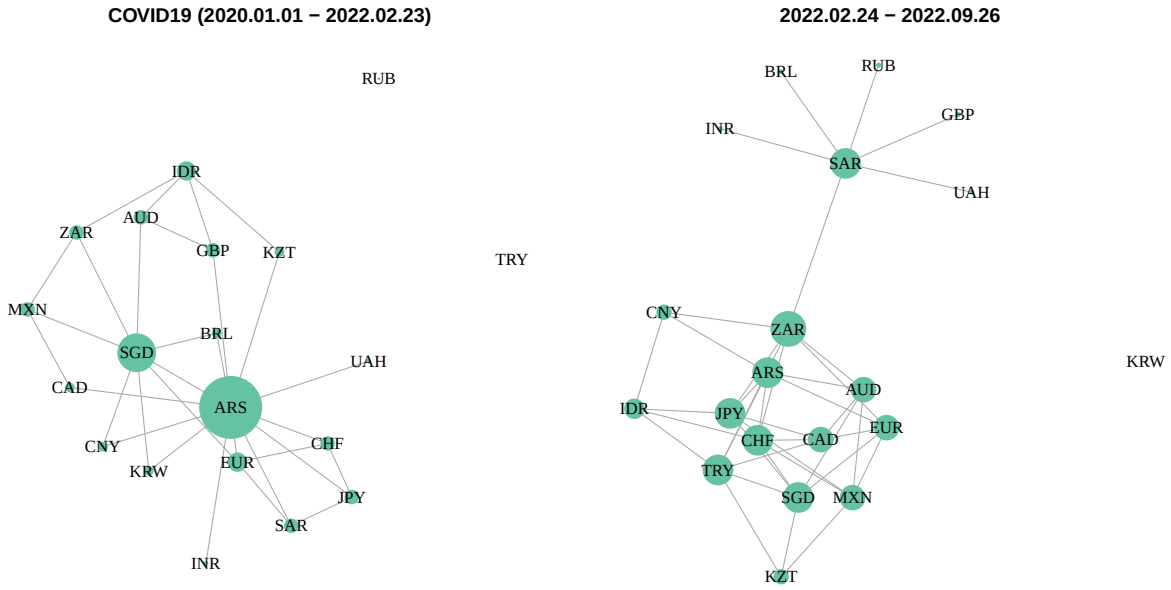

Figure 8: Estimated extremal currency exchange rate risk network (a) COVID-19 and (b) the 2022 military conflict between Russia and Ukraine.

Figure 8 (a) shows that during the COVID-19 pandemics the currency ARS is the strongest-connected one in the risk network, whereas TRY and RUB are isolated and most of the other currencies are also less interconnected. Historical economic data show that the Argentinian economy went through a crisis, and COVID-19 has further deteriorated its economic situation. Finally, Figure 8 (b) shows the risk network for currency exchange rates from the beginning of the 2022 military conflict between Russia and Ukraine until 2022.09.26. Interestingly, SAR is in the center of a local cluster which indicates its important role in this global energy-related disturbance.

6 Conclusions

We have proposed a novel extremal dependence measure – PTCC – and have further defined related concepts of partial tail-uncorrelatedness and extreme graphical models. This flexible modeling framework strongly draws from discrete representations of extremal correlation based on linear combinations of independent random variables and an appropriately defined inner product. It enables using tools for linear prediction and for matrix-based multivariate analysis similar to the classical statistical setting of Gaussian random vectors.

Efficient inference methods are proposed to learn extremal networks from high-dimensional data via PTCCs based on state-of-the-art methods from graph theory.

Our graph-inference approach is flexible, can be applied to general undirected graphs, and easily scales to the high-dimensional setting. We demonstrate the effectiveness of our method as an exploratory tool for interpretable extremal network estimation. In an application to river discharge data from the upper Danube basin, our method outperforms existing methods in capturing physically (mostly) consistent connection flows with the level of connection strength demonstrated. In another application to historical currency exchange rate data, we obtain interesting findings based on the estimated risk network, which can be backed up by real historical evidence. Our interpretations are still heuristic for this applications, and economic expertise may be needed for further clarification.

From the theoretical perspective, our research has allowed us to gain insights into remaining challenges. By analogy with the widely used Gaussian Markov random field framework, one could construct Markov random fields for extremes where the distribution for connected variables can be fitted with an asymptotically justified model for extremes, whereas the factorization of the likelihood could be specified by the extremal networks learned from the data. Another direction to investigate concerns the geometric representation of multivariate extremes (Nolde and Wadsworth, 2020; Simpson *et al.*, 2021). It looks promising to unite different modeling approaches for extremes that heavily rely on geometric representations.

Acknowledgments

We mention that there has been independent and parallel work by Lee & Cooley, which also investigates partial tail correlation for extremes. Our understanding is that inference in their work focuses rather on hypothesis testing (with the goal of checking if the partial tail correlation for a fixed pair of variables is significantly different from zero), and not on learning network structures with shape constraints. This publication is based upon work supported by the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under Award No. OSR-CRG2020-4394.

A Proof

Suppose a block matrix $X = \begin{bmatrix} E & F & G \\ H & J & K \\ L & M & N \end{bmatrix}$ which is nonsingular, then we have

$$X^{-1} = \begin{bmatrix} E^{-1} + E^{-1}[FA^{-1}H + US^{-1}V]E^{-1} & -E^{-1}[F - US^{-1}C]A^{-1} & -E^{-1}US^{-1} \\ -A^{-1}[H - BS^{-1}V]E^{-1} & A^{-1} + A^{-1}BS^{-1}CA^{-1} & -A^{-1}BS^{-1} \\ -S^{-1}VE^{-1} & -S^{-1}CA^{-1} & S^{-1} \end{bmatrix},$$

where

$$U = G - FA^{-1}B, \quad V = L - CA^{-1}H,$$

$$A = J - HE^{-1}F, \quad B = K - HE^{-1}G,$$

$$C = M - LE^{-1}F, \quad D = N - LE^{-1}G, \quad S = D - CA^{-1}B.$$

In the case when X is symmetric and positive definite, we can rewrite

$$X = \begin{bmatrix} E & F & G \\ F^T & J & K \\ G^T & K^T & N \end{bmatrix},$$

and $C = B^T$, E, J, N are symmetric and positive definite. Then the problem becomes proving $-S^{-1}VE^{-1} =$

$D \iff G = FJ^{-1}K$.

Proof:

First, $\Rightarrow$

$$\begin{aligned} S &= N - LE^{-1}G - (M - LE^{-1}F)(J - HE^{-1}F)^{-1}(K - HE^{-1}G) \\ &= N - G^TE^{-1}G - (K^T - G^TE^{-1}F)(J - F^TE^{-1}F)(K - F^TE^{-1}G), \end{aligned}$$

Since $GJ^{-1}K = G$,

$$\begin{aligned} V &= L - (M - LE^{-1}F)(J - HE^{-1}F)^{-1}H \\ &= L - (K^TG^TE^{-1}F)(J - F^TE^{-1}F)^{-1}F^T \\ &= G^T - (K^T - G^TE^{-1}F)(J - F^TE^{-1}F)^{-1}F^T \\ &= K^TJ^{-1}F^T - (K^T - K^TJ^{-1}F^TE^{-1}F)(J - F^TE^{-1}F)^{-1}F^T \\ &= [K^TJ^{-1} - K^T(I - J^{-1}F^TE^{-1}F)(J - F^TE^{-1}F)^{-1}]F^T \end{aligned}$$

$$= K^T J^{-1}[I - (J - F^T E^{-1} F)(J - F^T E^{-1} F)^{-1}]F^T$$

$$= K^T J^{-1}[I - I]F^T$$

$$= 0.$$

Second, since E is positive definite, E^{-1} is also positive definite.

$$S = D - CA^{-1}B$$

$$= (N - LE{-}1G) - (M - LE{-}1F)(J - HE{-}1F){-}1(K - HE{-}1G)$$

$$= (N - LE{-}1G) - (K^T - G^T E(N - LE{-}1G)F)(J - F^T E(N - LE{-}1G)F)^{-1}(K - FE^{-1}G)$$

$$= (N - LE{-}1G) - (K^T - G^T E(N - LE{-}1G)F)(J - F^T E(N - LE{-}1G)F)^{-1}(K - HE^{-1}G).$$

To ensure $-S^{-1}VE^{-1} = 0$, we need to let $V = 0$. Assume $G^T - K^T J^{-1} F^T = Q, Q \neq 0$. Since $V = 0$, we have

$$V = G^T - (K^T - G^T E^{-1} F)(J - F^T E^{-1} F)^{-1} F^T = 0$$

$$\iff Q + K^T J^{-1} F^T - [K^T - (Q + K^T J^{-1} F^T)E^{-1} F](J - F^T E^{-1} F)^{-1} F^T = 0$$

$$\iff Q + K^T J^{-1} F^T - (K^T - K^T J^{-1} F^T E^{-1} F)(J - F^T E^{-1} F)^{-1} F^T + QE^{-1} F(J - F^T E^{-1} F)^{-1} F^T = 0$$

$$\iff Q = -QE^{-1} F(J - F^T E^{-1} F)^{-1} F^T, \text{Denote } E^{-1} F(J - F^T E^{-1} F)^{-1} F^T = P,$$

$$\iff P^T Q^T = -Q^T.$$

Each column of Q^T is either 0 or the eigenvector of P^T that is corresponding to the eigenvalue -1, such that $P^T x = -x$. However, E^{-1} is positive definite and $F(J - F^T E^{-1} F)^{-1} F^T$ is at least semi-positive definite, since $J - F^T E^{-1} F$ should be positive definite. Therefore, it is impossible for P^T to have negative eigenvalues. Therefore, $Q^T = 0$.

Remark 1: For two semi-positive definite matrices A_1, B_1, $A_1 B_1 x = \lambda x, x \neq 0$, and λ is the eigenvalue of $A_1 B_1$, we have if $x^T B_1^T A_1 B_1 x = \lambda x^T B_1^T x$, then $\lambda \geq 0$. (easy to prove).

Remark 2: Inverse a 2×1 symmetric block matrix $X = \begin{bmatrix} A & B \\ B^T & D \end{bmatrix}$,
$X^{-1} = \begin{bmatrix} A^{-1} + A^{-1}B(D - B(D - B^T A^{-1} B)^{-1} B^T A^{-1}) & -A^{-1}B(D - B^T A^{-1} B)^{-1} \\ -(D - B^T A^{-1} B)^{-1} B^T A^{-1} & (D - B^T A^{-1} B)^{-1} \end{bmatrix}$.

B Data details

Countries	Currency symbol	Currency name
EU	EUR	EURO
United Kingdom	GBP	Pound Sterling
India	INR	Indian Rupee
Australia	AUD	Australian Dollar
Canada	CAD	Canadian Dollar
South Africa	ZAR	Rand
Japan	JPY	Yen
Singapore	SGD	Singapore Dollar
China	CNY	Yuan
Switzerland	CHF	Swiss Franc
Republic of Korea	KRW	Won
Turkey	TRY	Turkish Lira
Mexico	MXN	Mexican Peso
Brazil	BRL	Real
Indonesia	IDR	Rupiah
Saudi Arabia	SAR	Saudi Riyal
Russia	RUB	Ruble
Argentina	ARS	Argentine Peso
Ukraine	UAH	Ukrainian Hryvnia
Kazakstan	KZT	Kazakhstani Tenge

Table 1: Currency symbol list.

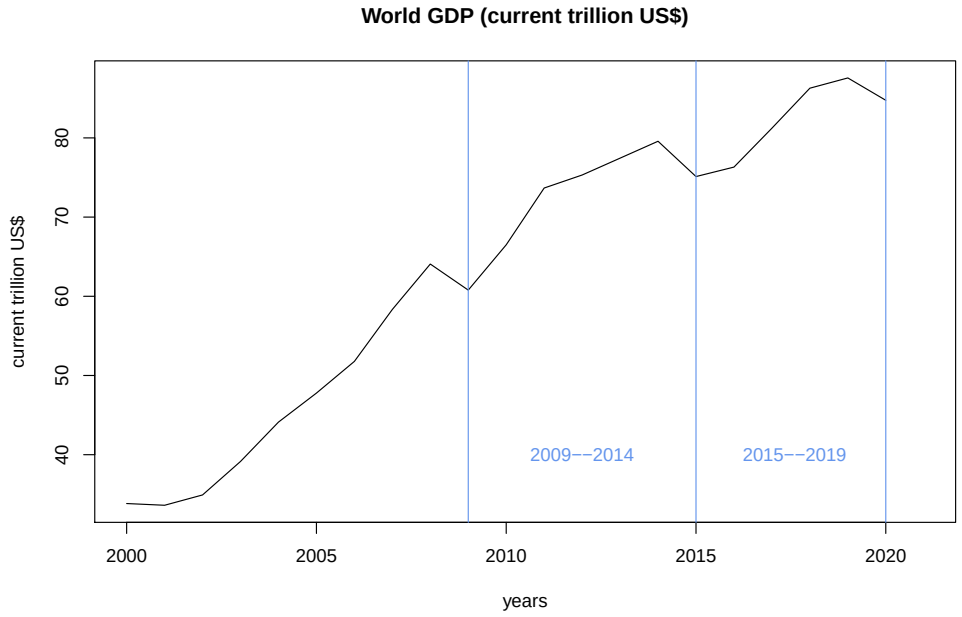

Figure 9: World GDP (current trillion US$) from 2000 to 2020. Data source: The World Bank.

References

Ahmed, M., Maume-Deschamps, V., Ribereau, P. and Vial, C. (2020) Spatial risk measures for max-stable and max-mixture processes. *Stochastics* **92**(7), 1005–1020.

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

Baba, K., Shibata, R. and Sibuya, M. (2004) Partial correlation and conditional correlation as measures of conditional independence. *Australian & New Zealand Journal of Statistics* **46**(4), 657–664.

Bacro, J.-N., Gaetan, C., Opitz, T. and Toulemonde, G. (2020) Hierarchical space-time modeling of asymptotically independent exceedances with an application to precipitation data. *Journal of the American Statistical Association* **115**(530), 555–569.

Bassamboo, A., Juneja, S. and Zeevi, A. (2008) Portfolio credit risk with extremal dependence: Asymptotic analysis and efficient simulation. *Operations Research* **56**(3), 593–606.

Bekiros, S. and Uddin, G. S. (2017) Extreme dependence under uncertainty: an application to stock, currency and oil markets. *International Review of Finance* **17**(1), 155–162.

Bopp, G. P., Shaby, B. A. and Huser, R. (2021) A hierarchical max-infinitely divisible spatial model for extreme precipitation. *Journal of the American Statistical Association* **116**(533), 93–106.

Cooley, D. and Thibaud, E. (2019) Decompositions of dependence for high-dimensional extremes. *Biometrika* **106**(3), 587–604.

Davis, R. A., Mikosch, T. and Zhao, Y. (2013) Measures of serial extremal dependence and their estimation. *Stochastic Processes and their Applications* **123**(7), 2575–2602.

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

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

Engelke, S., Opitz, T. and Wadsworth, J. (2019) Extremal dependence of random scale constructions. *Extremes* **22**(4), 623–666.

Engelke, S. and Volgushev, S. (2020) Structure learning for extremal tree models. *arXiv preprint arXiv:2012.06179* .

Ferro, C. A. and Stephenson, D. B. (2011) Extremal dependence indices: Improved verification measures for deterministic forecasts of rare binary events. *Weather and Forecasting* **26**(5), 699–713.

Fix, M. J., Cooley, D. S. and Thibaud, E. (2021) Simultaneous autoregressive models for spatial extremes. *Environmetrics* **32**(2), e2656.

Fougères, A.-L., Mercadier, C. and Nolan, J. P. (2013) Dense classes of multivariate extreme value distributions. *Journal of Multivariate Analysis* **116**, 109–129.

Friedman, J., Hastie, T. and Tibshirani, R. (2008) Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9**(3), 432–441.

Gissibl, N. (2018) *Graphical modeling of extremes: Max-linear models on directed acyclic graphs*. Ph.D. thesis, Universitätsbibliothek der TU München.

Gnecco, N., Meinshausen, N., Peters, J. and Engelke, S. (2021) Causal discovery in heavy-tailed models. *The Annals of Statistics* **49**(3), 1755–1778.

Gong, Y. and Huser, R. (2021) Asymmetric tail dependence modeling, with application to cryptocurrency market data. *The Annals of Applied Statistics, to appear* .

Huang, W. K., Cooley, D. S., Ebert-Uphoff, I., Chen, C. and Chatterjee, S. (2019) New Exploratory Tools for Extremal Dependence: χ Networks and Annual Extremal Networks. *Journal of Agricultural, Biological and Environmental Statistics* **24**(3), 484–501.

Huser, R. and Davison, A. C. (2014) Space-time modelling of extreme events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **76**, 439–461.

Huser, R. and Wadsworth, J. L. (2019) Modeling spatial processes with unknown extremal dependence class. *Journal of the American Statistical Association* **114**(525), 434–444.

Huser, R. and Wadsworth, J. L. (2022) Advances in statistical modeling of spatial extremes. *Wiley Interdisciplinary Reviews: Computational Statistics* **14**(1), e1537.

Janßen, A. and Wan, P. (2020) k-means clustering of extremes. *Electronic Journal of Statistics* **14**(1), 1211–1233.

Klüppelberg, C. and Krali, M. (2021) Estimating an extreme Bayesian network via scalings. *Journal of Multivariate Analysis* **181**, 104672.

Kumar, S., Ying, J., de Miranda Cardoso, J. V. and Palomar, D. (2019) Structured graph learning via Laplacian spectral constraints. *Advances in neural information processing systems* **32**.

Larsson, M. and Resnick, S. I. (2012) Extremal dependence measure and extremogram: the regularly varying case. *Extremes* **15**(2), 231–256.

Lawrance, A. (1976) On conditional and partial correlation. *The American Statistician* **30**(3), 146–149.

Ledford, A. W. and Tawn, J. A. (1996) Statistics for near independence in multivariate extreme values. *Biometrika* **83**(1), 169–187.

Lee, J. and Cooley, D. (2021) Transformed-linear prediction for extremes. *arXiv preprint arXiv:2111.03754* .

Marcon, G., Padoan, S. A. and Antoniano-Villalobos, I. (2016) Bayesian inference for the extremal dependence. *Electronic Journal of Statistics* **10**(2), 3310–3337.

Mhatre, N. and Cooley, D. (2020) Transformed-Linear Models for Time Series Extremes. *arXiv preprint arXiv:2012.06705* .

Nolde, N. and Wadsworth, J. L. (2020) Linking representations for multivariate extremes via a limit set. *arXiv preprint arXiv:2012.00990* .

Opitz, T., Huser, R., Bakka, H. and Rue, H. (2018) INLA goes extreme: Bayesian tail regression for the estimation of high spatio-temporal quantiles. *Extremes* **21**(3), 441–462.

Resnick, S. I. (2007) *Heavy-tail phenomena: probabilistic and statistical modeling*. Springer Science & Business Media.

Richards, J. and Wadsworth, J. L. (2021) Spatial deformation for nonstationary extremal dependence. *Environmetrics* **32**(5), e2671.

Röttger, F., Engelke, S. and Zwiernik, P. (2021) Total positivity in multivariate extremes. *arXiv preprint arXiv:2112.14727* .

Saunders, K., Stephenson, A. and Karoly, D. (2021) A regionalisation approach for rainfall based on extremal dependence. *Extremes* **24**(2), 215–240.

Sibuya, M. (1960) Bivariate extreme statistics, I. *Annals of the Institute of Statistical Mathematics* **11**(3), 195–210.

Simpson, E. S., Wadsworth, J. L. and Tawn, J. A. (2021) A geometric investigation into the tail dependence of vine copulas. *Journal of Multivariate Analysis* **184**, 104736.

Speed, T. P. and Kiiveri, H. T. (1986) Gaussian Markov distributions over finite graphs. *The Annals of Statistics* pp. 138–150.

Tran, N. M., Buck, J. and Klüppelberg, C. (2021) Estimating a Latent Tree for Extremes. *arXiv preprint arXiv:2102.06197* .

Vettori, S., Huser, R. and Genton, M. G. (2019) Bayesian modeling of air pollution extremes using nested multivariate max-stable processes. *Biometrics* **75**, 831–841.

Vettori, S., Huser, R., Segers, J. and Genton, M. G. (2020) Bayesian model averaging over tree-based dependence structures for multivariate extremes. *Journal of Computational and Graphical Statistics* **29**, 174–190.

Wadsworth, J., Tawn, J. A., Davison, A. and Elton, D. (2017) Modelling across extremal dependence classes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **79**(1), 149–175.

Wadsworth, J. L. and Tawn, J. (2019) Higher-dimensional spatial extremes via single-site conditioning. *arXiv preprint arXiv:1912.06560* .

Wadsworth, J. L. and Tawn, J. A. (2012) Dependence modelling for spatial extremes. *Biometrika* **99**(2), 253–272.

Winter, H. C. and Tawn, J. A. (2016) Modelling heatwaves in central France: a case-study in extremal dependence. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **65**(3), 345–365.

Yan, X., Wu, Q. and Zhang, W. (2019) Cross-sectional learning of extremal dependence among financial assets. *Advances in Neural Information Processing Systems* **32**.

Zhong, P., Huser, R. and Opitz, T. (2022) Modeling nonstationary temperature maxima based on extremal dependence changing with event magnitude. *The Annals of Applied Statistics* **16**(1), 272–299.