SFB
823

Regional extreme value index estimation and a test of homogeneity

Paul Kinsvater, Roland Fried,
Jona Lilienthal

Nr. 20/2015

Discussion Paper

Regional Extreme Value Index Estimation and a Test of Homogeneity

Paul Kinsvater, Roland Fried, Jona Lilienthal
TU Dortmund

Abstract

This paper deals with inference on extremes of heavy tailed distributions. We assume distribution functions F of Pareto-type, i.e. $1 - F(x) = x^{-1/\gamma}L(x)$ for some $\gamma > 0$ and a slowly varying function $L : \mathbb{R}_+ \to \mathbb{R}_+$. Here, the so called extreme value index (EVI) γ is of key importance. In some applications observations from closely related variables are available, with possibly identical EVI γ. If these variables are observed for the same time period, a procedure called BEAR estimator has already been proposed. We modify this approach allowing for different observation periods and pairwise extreme value dependence of the variables. In addition, we present a new test for equality of the extreme value index. As an application, we discuss regional flood frequency analysis, where we want to combine rather short sequences of observations with very different lengths measured at many gauges for joint inference.

Keywords: Hill estimator; extreme value index; homogeneity test; regional flood frequency analysis

1 Introduction

In environmental sciences we are interested in extreme realizations of a variable X following some distribution F in order to analyze the frequency of hazardous events such as floods (Dixon et al., 1998; Hosking and Wallis, 2005), extreme precipitations (Cooley et al., 2007) or extreme temperatures (Jarušková and Rencová, 2008; Fuentes et al., 2013). Measurements are collected at different locations, with observation lengths for each location being usually rather limited. The analysis is further complicated by the typical heavy tailed behavior of these quantities.

The class of Pareto-type distributions is used as a flexible model in heavy tail analysis. These distributions are characterized by polynomial decreasing right tail behavior. More precisely, F is called a Pareto-type distribution, if $\inf\{x : F(x) < 1\} = \infty$ (right-unlimited support) and for some $\gamma > 0$ and a slowly varying function L,

$$\bar{F}(x) = (1 - F)(x) = x^{-1/\gamma} \cdot L(x) \tag{1}$$

holds for all $x > 0$. The parameter γ is called extreme value index (EVI) and the function $L : \mathbb{R}_+ \to \mathbb{R}_+$ satisfies $L(tx)/L(t) \to 1$ for $t \to \infty$ and all $x > 0$. The popularity of this class can be explained by the fact that it coincides with the Fréchet maximum domain of attraction (de Haan and Ferreira, 2006, Theorem 1.2.1), which means that F^n is well approximated by a parametric extreme value distribution $GEV(\mu_n, \sigma_n, \gamma)$ for large n. The parameter γ is an indicator for the heaviness of the right tail, where e.g. $\int_0^\infty x^k dF(x)$ is finite for $k < \gamma^{-1}$ and infinite for $k > \gamma^{-1}$. Examples for distributions satisfying (1) are given by Student's t_ν with $\gamma = \frac{1}{\nu}$, Fisher's $F_{m,k}$ with $\gamma = \frac{2}{k}$, generalized extreme value $GEV(\mu, \sigma, \xi)$ and generalized Pareto $GP(\sigma, \xi)$ distributions with shape parameter $\xi = \gamma > 0$ and many others.

In practice γ is unknown and thus a key challenge is its adequate estimation. For a sample $X_1, \ldots, X_n$ of positive i.i.d. random variables with distribution function F satisfying (1) the popular Hill estimator (Hill, 1975) is given by

$$\widehat{\gamma} = H_{k,n} = \frac{1}{k} \sum_{i=1}^{k} \left(\log X_{n-i+1:n} - \log X_{n-k:n} \right), \tag{2}$$

where $X_{1:n} \leqslant \ldots \leqslant X_{n:n}$ are the order statistics and $k = k_n \leqslant n$ is a sequence of integers such that $k \to \infty$ and $k/n \to 0$. Since by (1)

$$P(X/u > x | X > u) = \bar{F}(ux)/\bar{F}(u) \approx x^{-1/\gamma} \tag{3}$$

for large u, the conditional distribution on the left hand side is approximated by the Pareto($1/\gamma$) and $\widehat{\gamma}$ in (2) can be interpreted as maximum likelihood approach for the parameter $\gamma > 0$, where we choose $u = Y_{n-k:n}$.

In environmental applications, where we observe the same variables at many sites j with site specific distributions F_j, regional frequency analysis provides methods for pooled estimation to overcome the problem of having only short sequences for each site available. There are a few different approaches. So called Index Flood procedures (Hosking and Wallis, 2005, Chapter 1.3) are very popular in hydrology. The Index Flood procedure is built on the assumption that

$$H_{0,IF}: \; F_j^{-1}(p) = \mu_j \cdot F_\theta^{-1}(p), \; j = 1, \ldots, d, \tag{4}$$

holds for a group of d distributions, where $\{F_\theta : \; \theta \in \Theta\}$ is a predetermined parametric family of distributions and θ, $\mu_j = \mu(F_j)$, $j = 1, \ldots, d$, are unknown parameters. Lettenmaier et al. (1987) show by simulation that such an regional approach is preferable compared to marginal estimation, even under moderate deviations from assumption $H_{0,IF}$.

Here theory is developed under weaker assumptions than stated in (4). Essentially, we suppose that a group of similar distributions shares the same EVI γ, i.e. we assume

$$H_{0,evi}: \; \gamma_1 = \ldots = \gamma_d = \gamma, \tag{5}$$

where γ_j is the EVI of F_j. If $H_{0,evi}$ holds and in the context of regional frequency analysis, γ indicates occurrence and amount (up to local scale) of extreme events in a whole region and therefore γ is called regional extreme value index. For the theory we do not impose any parametric assumptions concerning the margins F_j or the spatial dependence structure

modeled by a copula C. An additional assumption of extreme value dependence greatly improves the efficiency of the estimation of the limiting covariance between marginal Hill estimators. This turns out to be indispensable if only short data sequences are available. Our approach generalizes the BEAR procedure given in Clémençon and Dematteo (2014) to the practically relevant situation where the marginal data sequences are of very different lengths. The original BEAR procedure from the latter reference is based on an asymptotically optimal weighting scheme that allows to decrease the variability but does not tackle the bias of the joint Hill estimator. As opposed to these authors we additionally take also the dimension d into account in order to reduce the bias and we propose a test for the basic hypothesis $H_{0,evi}$.
The Hill estimator and related methodology is suited within a non-parametric framework, if rather long data sequences are available. We are particularly interested in the applicability of the new methods, the generalized BEAR procedure and the new test of $H_{0,evi}$, to estimate the EVI γ from a group of d jointly extreme value dependent variables fulfilling $H_{0,evi}$. This is advantageous particularly if only a small to moderate number n_j of observations is available for each variable $j = 1, \ldots, d$. The main results of this paper can be summarized as follows:

- We derive the asymptotic distribution of the vector of Hill estimators in case of very different lengths of the marginal samples. This allows us to formulate a joint estimator of γ with an arbitrary weighting of the individual estimators and an asymptotic test for $H_{0,evi}$.
- For reasonable settings from hydrology with large dimension d, small to moderate marginal sample sizes n_j, $j = 1, \ldots, d$, and under extreme value dependence, the estimation procedure proposed here significantly reduces the estimation error. In particular, taking into account the dimension d for the choice of upper order statistics, i.e. setting $k_j = k_j^{(d)}$ in (2), is important to reduce a typically dominant bias.
- Under assumption $H_{0,IF}$ stated in (4), the bias issue of Hill's estimator is much less present when the proposed test is applied. The nominal level is preserved well in reasonable settings from hydrology. Moreover, when variables are spatially dependent, the new test turns out to be much more powerful against certain alternatives than competing methods known from the literature.

The rest of the paper is organized as follows. Section 2 provides asymptotic results and Section 3 discusses statistical methodology for joint estimation of γ from several data sets and testing H_0. Section 4 reports a simulation study and in Section 5 we analyze seasonal maxima from a number of river gauges located in Saxony, Germany. We conclude in Section 6. Proofs are given in an Appendix.

2 Asymptotic results

Let $\mathbf{X} = (X_1, \ldots, X_d)^T$ be a random vector with support in $\mathbb{R}_+^d$ and continuous marginal c.d.f.'s $F_j(x) = P(X_j \leqslant x)$, $j = 1, \ldots, d$. By Sklar's representation theorem (Sklar, 1959), the joint c.d.f. F of $\mathbf{X}$ is uniquely determined by

$$F(\mathbf{x}) = C(F_1(x_1), \ldots, F_d(x_d)), \quad \mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d, \tag{6}$$

where $C : [0,1]^d \to [0,1]$ is the distribution function of the probability transform $\mathbf{U} = (F_1(X_1), \ldots, F_d(X_d))$ giving a margin-free characterization of the dependence structure of $\mathbf{X}$. Suppose that F_j satisfies (1) for some EVI $\gamma_j > 0$ for each $j = 1, \ldots, d$ and that $\gamma_1 = \ldots = \gamma_d = \gamma$ holds.

Let $\mathbf{X}_i = (X_{i,1}, \ldots, X_{i,d})^T$, $i = 1, \ldots, n$, be independent copies of $\mathbf{X}$ with i indicating time. The assumption that all d components $X_{i,j}$ are observed for the same time points $i = 1, \ldots, n$ is very restrictive in the context of regional frequency analysis. In our hydrological applications, where data is collected from many stations, this is rarely the case. Instead, we assume that we observe a scheme of variables

$$X_{i,j}, \ j = 1, \ldots, d \ \text{ and } \ i = a_j + 1, a_j + 2, \ldots, n, \tag{7}$$

where the integers $1 \leqslant a_j \leqslant n$ denote the starting point of measurement at station j and $n_j = n - a_j$ is the total number of observations for component j. In order to account for possibly very different numbers n_j in the asymptotics, we set $a_j = \lfloor n(1-\tau_j) \rfloor$ for some real $0 < \tau_j < 1$ such that $n_j/n \to \tau_j$ for $n \to \infty$. τ_j is interpreted as the relative proportion of time observed at location j.

Let $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_d) \in (0,1]^d$ be fixed and $\mathbf{H}_{\mathbf{k},\boldsymbol{\tau},n} = \left(H^{(1)}_{k_1,\tau_1,n}, \ldots, H^{(d)}_{k_d,\tau_d,n}\right)^T$, where the j-th component $H^{(j)}_{k_j,\tau_j,n}$ is Hill's estimator for the sample $X_{\lfloor n(1-\tau_j) \rfloor+1,j}, \ldots, X_{n,j}$. In addition, we assume the same technical assumptions as in Clémençon and Dematteo (2014):

1. For $j = 1, \ldots, d$, $k_j = k_j(n)$ is an intermediate sequence of integers, i.e. $k_j \to \infty$ and $k_j/n \to 0$ for $n \to \infty$. In addition, $\lim_{n\to\infty} \frac{k_1}{k_j} = c_j$ for some $c_j \in (0, \infty)$.

2. We assume that von Mises' condition holds for all $j = 1, \ldots, d$ and the same $\gamma > 0$. I.e. the derivatives $f_j = F_j'$ exist and satisfy

$$\lim_{x\to\infty} \frac{x f_j(x)}{1 - F_j(x)} = \frac{1}{\gamma_j}, \ j = 1, \ldots, d. \tag{8}$$

3. For $j = 1, \ldots, d$, $U_j(t) = F_j^{-1}(1 - 1/t)$ and $n \to \infty$ we have

$$\sqrt{k_j} \int_1^\infty \left\{ \frac{n}{k_j} \bar{F}_j \left(U_j(n/k_j)x \right) - x^{-1/\gamma_j} \right\} \frac{dx}{x} \to 0. \tag{9}$$

4. For $1 \leqslant \ell \neq m \leqslant d$ and $n \to \infty$ we have

$$\sup_{x,y>1} \left| \frac{n}{k_1} \bar{F}_{\ell,m} \left(U_\ell \left(\frac{n}{k_1 x} \right), U_m \left(\frac{n}{k_1 y} \right) \right) - \Lambda_{\ell,m}(x,y) \right| = o\left(\frac{1}{\log k_1} \right), \tag{10}$$

where $\bar{F}_{\ell,m}(x,y) = P\left(X_\ell > x, X_m > y\right)$ and

$$\Lambda_{\ell,m}(x,y) = \lim_{t\to\infty} t \cdot P\left(X_\ell > U_\ell(t/x), X_m > U_m(t/y)\right).$$

exists for all $1 \leqslant \ell, m \leqslant d$.

Remark: A) The assumption $k_j = o(n)$ is standard for the Hill estimator. To ensure that the joint distribution also converges to a non degenerate limit in $\mathbb{R}^d$ it is natural to demand that $\lim_{n\to\infty} k_1/k_j > 0$ exists for all $2 \leqslant j \leqslant d$.
B) The von Mises condition, which also implies (1) (de Haan and Ferreira, 2006, Theorem 1.1.11), together with assumption (9) guarantees weak convergence of $\sqrt{k_j}(H^{(j)}_{k_j,\tau_j,n} - \gamma_j)$ against a centered normal distribution (Resnick, 2007, Prop. 9.3). These assumptions can be weakened by various versions of so-called second order regular variation conditions, e.g. such that the asymptotic normality holds with a not necessarily centered limiting distribution (de Haan and Ferreira, 2006, Theorem 3.2.5). All these conditions, von Mises and second order regular variation, require the availability of detailed information on the tail of the distribution which is however not usually the case in practice.
C) $\Lambda_{l,m}$ is called the upper tail dependence copula between the components X_l and X_m (Schmidt and Stadtmüller, 2006). Since $\bar{F}\left(U_l(t/x), U_m(t/y)\right) = \bar{C}_{l,m}(x/t, y/t)$, where $\bar{C}_{l,m}$ is the survival copula of (X_l, X_m), $\Lambda_{l,m}$ is a margin-free characterization of the upper tail dependence between X_l and X_m. Clémençon and Dematteo (2014) use quantities $\nu_{l,m}$ instead, which provide an alternative upper tail dependence measure (Resnick, 2007, Chapter 6). In contrast to $\Lambda_{l,m}$, the latter characterization is not margin-free. The relation between these two measures is given by $\Lambda_{l,m}(x,y) = \nu_{l,m}(x^{-\gamma}, y^{-\gamma})$, provided they exist.

The following Proposition is an extension of Corollary 3.6 in Clémençon and Dematteo (2014) to scenarios described by (7) with $(n - a_j)/n \to \tau_j \in (0,1)$.

Proposition 1 (Joint weak convergence)
Assume that assumptions 1.-4. are met and let $\mathbf{1} = (1, \dots, 1)^T \in \mathbb{R}^d$. Then we have for $n \to \infty$ that

$$\sqrt{k_1}\left(\mathbf{H}_{\mathbf{k},\boldsymbol{\tau},n} - \gamma\mathbf{1}\right) \xrightarrow{D} N\left(0, \gamma^2 \cdot \Sigma\right) \tag{11}$$

holds, where $\Sigma \in \mathbb{R}^{d\times d}$ is given by

$$\Sigma_{l,m} = c_l \cdot c_m \cdot (\tau_l \wedge \tau_m) \cdot \Lambda_{l,m}\left((\tau_l c_l)^{-1}, (\tau_m c_m)^{-1}\right)$$

for $1 \leqslant l, m \leqslant d$. For $l = m$ this reduces to $\Sigma_{l,l} = c_l$.

Set $\hat{\gamma}_{\mathbf{k},\boldsymbol{\tau},n}(\boldsymbol{w}) = \boldsymbol{w}^T \cdot \mathbf{H}_{\mathbf{k},\boldsymbol{\tau},n}$, where $\boldsymbol{w} \in W = \{x \in \mathbb{R}^d : \sum_{i=1}^d x_i = 1\}$ is a vector of weights. As a direct consequence, we have

$$\sqrt{k_1}\left(\hat{\gamma}_{\mathbf{k},\boldsymbol{\tau},n}(\boldsymbol{w}) - \gamma\right) \xrightarrow{D} N(0, \gamma^2 \boldsymbol{w}^T \Sigma \boldsymbol{w}). \tag{12}$$

This result together with a weakly consistent estimator $\hat{\Sigma}$ of Σ (see Section 3) allows for the derivation of asymptotically valid confidence intervals. The drawback of this approach for finite sample applications is that a potentially apparent bias term of Hill's estimator is not taken into account. As a consequence of a possibly dominant bias in the overall estimation error, the true coverage probability can differ substantially from the nominal one.
A second consequence of Proposition 1 is the following:

Proposition 2 (Wald-type test statistic)
Assume (11) holds with Σ being positive definite. Then for $\boldsymbol{w} \in W$, a weakly consistent estimator $\widehat{\Sigma}$ of Σ and $n \to \infty$ we have

$$\begin{aligned}\tilde{W}_{\mathbf{k},\boldsymbol{\tau},n}(\boldsymbol{w}) &= \frac{k_1}{\widehat{\gamma}_{\mathbf{k},\boldsymbol{\tau},n}(\boldsymbol{w})^2} \left(\mathbf{H}_{\mathbf{k},\boldsymbol{\tau},n} - \widehat{\gamma}_{\mathbf{k},\boldsymbol{\tau},n}(\boldsymbol{w})\mathbf{1}\right)^T \widehat{\Sigma}^{-1} \left(\mathbf{H}_{\mathbf{k},\boldsymbol{\tau},n} - \widehat{\gamma}_{\mathbf{k},\boldsymbol{\tau},n}(\boldsymbol{w})\mathbf{1}\right) \\ &\xrightarrow{D} bZ_1^2 + \sum_{j=2}^{d-1} Z_j^2,\end{aligned}$$

where $Z_1, \ldots, Z_{d-1}$ are i.i.d. standard normal and $b = \mathbf{1}^T\Sigma^{-1}\mathbf{1} \cdot \boldsymbol{w}^T\Sigma\boldsymbol{w}$. In addition, let $W_{\mathbf{k},\boldsymbol{\tau},n} = \tilde{W}_{\mathbf{k},\boldsymbol{\tau},n}(\widehat{\boldsymbol{w}}_{opt})$ with $\widehat{\boldsymbol{w}}_{opt} = (\mathbf{1}^T\widehat{\Sigma}^{-1}\mathbf{1})^{-1} \cdot \widehat{\Sigma}^{-1}\mathbf{1}$. Then we have for $n \to \infty$ that

$$W_{\mathbf{k},\boldsymbol{\tau},n} \xrightarrow{D} \chi^2_{d-1}. \tag{13}$$

On the other hand, if F_j satisfies (9)*, $j = 1, \ldots, d$, with $\gamma_i \neq \gamma_j$ for some $1 \leqslant i \neq j \leqslant d$, we have $W_{\mathbf{k},\boldsymbol{\tau},n} \xrightarrow{P} \infty$.*

According to these results, $W_{\mathbf{k},\boldsymbol{\tau},n}$ provides an asymptotic significance test of $H_{0,evi}$ under assumptions 1.-4., which is consistent against arbitrary fixed alternatives.

3 Statistical methodology

We discuss two statistical applications of the theory presented in the previous section. First, the joint estimation of γ and second, a test for hypothesis $H_{0,evi}$ from (5). Although the theory is developed in a quite general framework, we are particularly interested in applications under additional distributional assumptions with only short data sequences available.

For both applications, the limiting covariance matrix Σ needs to be estimated. Recall that the (ℓ, m)-th component of Σ, $1 \leqslant l, m \leqslant d$, is given by

$$\Sigma_{\ell,m} = c_\ell c_m(\tau_\ell \wedge \tau_m) \cdot \Lambda_{\ell,m}\left((\tau_\ell c_\ell)^{-1}, (\tau_m c_m)^{-1}\right), \ 1 \leqslant \ell, m \leqslant d, \tag{14}$$

with $c_j = \lim_{n\to\infty} k_1/k_j$, $\tau_j = \lim_{n\to\infty} n_j/n$ and the upper tail dependence copula $\Lambda_{\ell,m}$ of (X_ℓ, X_m). For $\ell = m$ this equation simplifies to $\Sigma_{\ell,\ell} = c_\ell$. In order to estimate $\Sigma_{\ell,m}$ we replace c_j, τ_j and $\Lambda_{\ell,m}$ by k_1/k_j, n_j/n and a consistent estimator $\hat{\Lambda}_{\ell,m}$, respectively.

Let $(X_{i,\ell}, X_{i,m})$, $i = 1, \ldots, N = N(\ell, m)$, denote the independent copies of (X_ℓ, X_m) that are available for estimation. Then the empirical estimator of $\Lambda_{\ell,m}$ studied by Schmidt and Stadtmüller (2006) is given by

$$\widehat{\Lambda}_{\ell,m}(x, y) = \frac{1}{k}\sum_{i=1}^{N} \mathbb{1}\left(X_{i,\ell} > X_{\lceil N-xk\rceil:N,\ell}, X_{i,m} > X_{\lceil N-yk\rceil:N,m}\right), \tag{15}$$

where $k = o(N)$ is a tuning parameter. The disadvantage of this estimator lies in its slow convergence rate of only $\sqrt{k}$-consistence, since essentially only a sample fraction of k/N

observations is taken into account. When only moderate sample sizes N are available, like in our applications, the estimation error can be large.
However, we consider componentwise maxima in our application. It is well known that extreme value copulas are the only possible limits of copulas of componentwise maxima of i.i.d. vectors. For extreme value copulas $C_{\ell,m}$ we have a one-to-one correspondence between $C_{\ell,m}$ and its upper tail dependence copula $\Lambda_{\ell,m}$ given by

$$\Lambda_{\ell,m}(x,y) = (x+y)\cdot\left[1 - A_{\ell,m}\left(\frac{y}{x+y}\right)\right], \tag{16}$$

where $A_{\ell,m}(t) = -\log C_{\ell,m}(e^{-(1-t)}, e^{-t})$, $0 \leqslant t \leqslant 1$, is called Pickands dependence function (Pickands, 1981). Several estimators of A are known from the literature. In particular, the corrected CFG-estimator $A_{N,c}^{CFG}$ from Genest and Segers (2009) offers high efficiency. As opposed to (15), an estimator of $\Lambda_{\ell,m}$ based on the extreme value dependence assumption and the CFG-estimator is $\sqrt{N}$-consistent. This advantage over estimator (15) turns out to be crucial for an acceptable type-1 error of the proposed test in our simulation study.
In what follows, we denote the empirical estimator by $\hat{\Sigma}_{emp}$ and the CFG-based estimator by $\hat{\Sigma}_{ev}$.

3.1 Joint estimation of γ and choice of $\mathbf{k}$

Suppose for the moment that the numbers k_j, $j = 1,\dots,d$, are already available for estimation. In Clémençon and Dematteo (2014, Sec. 3.2) the joint estimator $\widehat{\gamma}(\boldsymbol{w}_{opt}) = \boldsymbol{w}_{opt}^T \cdot \mathbf{H}_{\mathbf{k},\boldsymbol{\tau},n}$ of γ has been studied in order to reduce the variability, where

$$\boldsymbol{w}_{opt} = \underset{\boldsymbol{w}\in W}{\arg\min}\, AVar(\boldsymbol{w}^T\cdot\mathbf{H}_{\mathbf{k},\boldsymbol{\tau},n}) = \underset{\boldsymbol{w}\in W}{\arg\min}\, \boldsymbol{w}^T\Sigma\boldsymbol{w}.$$

In the latter reference only non-negative weights $\boldsymbol{w}$ were considered for the minimization problem. Here, however, we do not apply this restriction.
Because in practice Σ is unknown, these asymptotically optimal weights are estimated by plugging in a consistent estimator $\hat{\Sigma}$ of Σ. Provided that $\hat{\Sigma}$ is nonsingular, this is solved by the Lagrange multipliers technique with solution

$$\widehat{\boldsymbol{w}}_{opt} = \left(\mathbf{1}^T\widehat{\Sigma}^{-1}\mathbf{1}\right)^{-1}\cdot\widehat{\Sigma}^{-1}\mathbf{1}. \tag{17}$$

Note that these are the same weights used in the test statistic $W_{\mathbf{k},\boldsymbol{\tau},n}$ in (13).
In order to study the gain in efficiency of the optimal weighting scheme, we also included the joint estimator with weights $\boldsymbol{w}_{ind} = \mathbf{k}/(\mathbf{1}^T\mathbf{k})$ in our simulations, where $\mathbf{k} = (k_1,\dots,k_d)^T$ are the integers used for the marginal Hill estimators. Note that these weights correspond to the assumption of upper tail independence.

Actually, more crucial than the choice of weights $\boldsymbol{w}$ is the choice of integers $\mathbf{k}$ for the upper order statistics. Several methods were proposed to solve this problem in the univariate setting (Drees et al., 2000). A difficulty for multivariate observations as considered here is that, in general, the optimal numbers k_j for marginal estimation and optimal $k_j^{(d)}$ for joint estimation do not coincide.

To motivate this finding, suppose that the observations follow a multivariate extreme value distribution given in (6) with identical marginal GEV distributions $F_j = GEV(\mu, \sigma, \gamma)$, $0 < \gamma < 1$, and independent components, i.e. $C(\mathbf{u}) = u_1 \cdots u_d$. Let n_j denote the number of observations of component $j = 1, \ldots, d$. The optimal k_j value that minimizes the asymptotic mean squared error (MSE) of the marginal Hill estimator based on the observations of component j alone is given by $k_j^{(1)} = \lfloor 2n_j^{2/3} \rfloor$ (Gomes and Pestana, 2007, Remark 3.1).

In fact, in this simple case, all observations are $N = \sum_{j=1}^d n_j$ realizations of the same GEV distribution, which implies that a total of $K = \lfloor 2N^{2/3} \rfloor$ upper observations out of the N observations should be used. Suppose that a fraction of n_j/N of the upper K values belongs to component j, $j = 1, \ldots, d$. Consequently, it is plausible to set $k_j^{(d)} = \lfloor 2n_j/N^{1/3} \rfloor < k_j^{(1)}$ for the joint estimation. For $n_1 = \ldots = n_d$ this simplifies to $k_j^{(d)} = \lfloor 2n_j^{2/3}/d^{1/3} \rfloor$, which means that optimal numbers $k_j^{(d)}$ should decrease with increasing dimension d. Indeed, from our simulation results presented in Section 4.1 we find that the performance of the joint Hill estimator with $k_j^{(d)} = \lfloor 2n_j^{2/3}/d^{1/3} \rfloor$ is superior to that with $k_j^{(1)} = \lfloor 2n_j^{2/3} \rfloor$ in most cases.

To be mathematically more precise, suppose that each marginal distribution F_j is a member of the Hall-Welsh class (Hall and Welsh, 1985; Gomes and Pestana, 2007) such that

$$F_j^{-1}\left(1 - \frac{1}{t}\right) = C_j t^\gamma \left(1 + \frac{\gamma \beta_j t^{\rho_j}}{\rho_j} + o(t^{\rho_j})\right) \tag{18}$$

holds for $t \to \infty$, extreme value index $\gamma > 0$, constants $C_j > 0$ and so-called second order parameters $\rho_j < 0$, $\beta_j \neq 0$, $j = 1, \ldots, d$. The Hall-Welsh class is a rich subset of the Pareto-type distributions. It contains, among others, the GEV and GP with positive shape and Student's t_ν distributions. Assume that the asymptotic variance of the joint Hill estimator derived in Section 2 is also valid for margins within the Hall-Welsh class. Together with the bias term obtained from Gomes and Pestana (2007, Sec. 3.1), we conclude that the mean squared error of $\widehat{\gamma}_{\mathbf{k},\boldsymbol{\tau},n}(\boldsymbol{w})$ is well approximated by

$$MSE(\widehat{\gamma}_{\mathbf{k},\boldsymbol{\tau},n}(\boldsymbol{w})) \approx \gamma^2 \boldsymbol{w}^T \Gamma(\mathbf{k}) \boldsymbol{w} + \gamma^2 \left(\sum_{j=1}^d w_j \beta_j \frac{(n_j/k_j)^{\rho_j}}{1 - \rho_j}\right)^2 \tag{19}$$

for large n, where the matrix $\Gamma(\mathbf{k}) \approx \frac{1}{k_1}\Sigma$ is given by

$$(\Gamma(\mathbf{k}))_{\ell,m} = \frac{\tau_\ell \wedge \tau_m}{k_\ell k_m} \Lambda_{\ell,m}\left(\frac{k_\ell}{\tau_\ell}, \frac{k_m}{\tau_m}\right), \ 1 \leqslant \ell, m \leqslant d.$$

From a theoretical point of view the optimal combination of weights $\boldsymbol{w}$ with $\sum_j w_j = 1$ and integers $\mathbf{k}$ with $1 \leqslant k_j < n_j$ is achieved by minimizing (19) with respect to both $\boldsymbol{w}$ and $\mathbf{k}$. However, this high dimensional and nonlinear minimization problem is computationally expensive and associated with the estimation of the second order parameters β_j

and ρ_j. Having our applications from hydrology in mind, we did not further pursue this optimization problem.

3.2 A new homogeneity test for regional frequency analysis

It is natural to consider the statistic $W_{\mathbf{k},\boldsymbol{\tau},n}$ to test the hypothesis $H_{0,evi}: \gamma_1 = \ldots = \gamma_d$.

Test procedure: Reject $H_{0,evi}$ at a significance level $\alpha \in (0,1)$, if $W_{\mathbf{k},\boldsymbol{\tau},n}$ exceeds the $(1-\alpha)$-quantile of the χ^2 distribution with $d-1$ degrees of freedom.

The asymptotic validity and consistency of this test follows from Proposition 2. Nevertheless, the performance of this test for finite samples can be very poor, even for large n. A reason for this is the bias of the Hill estimator, which depends on many different characteristics of the underlying marginal distributions. Very different marginal bias terms can lead to a rejection of $H_{0,evi}$, even if the null hypothesis is true.
It turns out that the bias issue for the test is much less present under classical assumptions from regional frequency analysis stated in (4). The latter means that all marginal variables are equal in distribution up to scale. Note that Hill's estimator is scale invariant. As a consequence, no matter from what marginal distribution a sample of size n is drawn, the exact distribution of $H_{k,n}$ remains the same. As long as the marginal sample lengths $(n_j)_{1\leqslant j\leqslant d}$ do not vary very much, we may expect that the approximation of statistic $W_{\mathbf{k},\boldsymbol{\tau},n}$ to its distributional limit will be acceptable.
To illustrate these considerations, we want to discuss a particular setting, which is of practical relevance in regional frequency analysis and which we also study in detail in simulations in Section 4.

Assumption: We assume that F from (6) is a d-variate extreme value distribution, which means that C is an extreme value copula and each margin F_j is an extreme value distribution $GEV(\mu_j, \sigma_j, \gamma_j)$ with location, scale and shape parameters μ_j, σ_j and γ_j, respectively.

Let $\delta_j = \mu_j/\sigma_j$ denote the location-scale ratios and assume that $\gamma_j > 0$ for all $j = 1, \ldots, d$. In spite of the asymptotic theory derived under $H_{0,evi}$, due to the bias problems mentioned above, it turns out that the proposed test is approximately valid in finite samples only under the stronger null hypothesis (4) applied e.g. in hydrology. In this particular setting the latter can be reformulated to

$$H_{0,IF} = H_{0,evi} \cap H_{0,delta}, \tag{20}$$

which means that $H_{0,delta}: \delta_1 = \ldots = \delta_d$ holds in addition to $H_{0,evi}$.
Many methods were proposed in order to test assumption (4) or, more specifically, $H_{0,IF}$. For an overview of the most popular procedures and a comparative simulation study, we refer to Viglione et al. (2007). The main drawback of all these methods is that they were designed for spatially independent observations, but this assumption is unlikely to hold in regional flood frequency analysis. Note also that this issue was not addressed in Viglione et al. (2007), i.e. all simulations there were carried out under spatial independence.

4 Simulation study

Motivated by our illustration presented in Section 5, we focus on simulations with multivariate extreme value distributed sequences. More precisely, we draw independent vector valued realizations from d dimensional distributions $F = C(F_1, \ldots, F_d)$ with (univariate) extreme value distributed margins $F_j = GEV(\mu_j, \sigma_j, \gamma_j)$, positive extreme value index $\gamma_j > 0$ and extreme value copula C from the family

$$C_{\boldsymbol{\theta},\mathbf{a}}(\mathbf{u}) = C_{\theta_1}(\mathbf{u}^{\mathbf{a}}) \cdot C_{\theta_2}(\mathbf{u}^{1-\mathbf{a}}), \tag{21}$$

where $\mathbf{u}^{\mathbf{a}} = (u_1^{a_1}, \ldots, u_d^{a_d})$, $1 - \mathbf{a} = (1 - a_1, \ldots, 1 - a_d)$, $\boldsymbol{\theta} = (\theta_1, \theta_2) \in [1, \infty)^2$, $\mathbf{a} = (a_1, \ldots, a_d) \in [0, 1]^d$ and C_θ is the d-dimensional Gumbel(θ) copula. The construction principle (21) is known as Khoudraji's device (Khoudraji, 1995; Durante and Salvadori, 2010). It is used in order to account for possible asymmetry in the dependence, which is also present in our illustration but not covered by common one-parameter copula families. Since all considered methods are scale invariant, we pay particular attention to the performance depending on the choice of $\delta_j = \mu_j/\sigma_j$. Recall that under the classical homogeneity assumption stated in (4) we have $\gamma_1 = \ldots = \gamma_d = \gamma$ and $\delta_1 = \ldots = \delta_d = \delta$.
Most simulations are carried out for dimension $d = 5$ and the following parameter values:

- $n \in \{50, 100\}$ (maximal sample length)
- $\boldsymbol{\tau} \in \{(1, 1, 1, 1, 1), (1, 0.9, 0.8, 0.7, 0.6)\}$ (relative sample lengths)
- $\gamma \in \{0.25, 0.5, 0.75\}$ (extreme value index)
- $\delta \in [1,\ 3]$ (location-scale ratio)
- $\boldsymbol{\theta} \in \{(1, 1), (1.5, 2.5)\}$ (strength of dependence)
- $\mathbf{a} = (0.9, 0.7, 0.5, 0.3, 0.1)$ (asymmetry of dependence)

These scenarios are supposed to cover many settings from regional flood frequency analysis. We also studied the performance for $d = 10$ and $d = 15$, but many results were qualitatively similar to those for $d = 5$ and are thus not reported in full detail. For $d = m \cdot 5$, $m \in \mathbb{N}$, and $\boldsymbol{\tau}, \mathbf{a} \in \mathbb{R}^5$ from above, the relative sample lengths and asymmetry coefficients were set to $\boldsymbol{\tau}_m = (\boldsymbol{\tau}, \boldsymbol{\tau}, \ldots, \boldsymbol{\tau}) \in \mathbb{R}^{md}$ and $\mathbf{a}_m = (\mathbf{a}, \mathbf{a}, \ldots, \mathbf{a}) \in \mathbb{R}^{md}$, respectively.
However, we found that the new test based on statistic $W_{\mathbf{k},\boldsymbol{\tau},n}$ tends to get liberal with increasing dimension d (at constant n). Based on our simulation results, we decided heuristically to multiply the statistic with an asymptotically negligible factor of $1 - d/(5N)$ with $N = min_{1 \leqslant j \leqslant d} n_j$ at the cost of a loss of power.
Simulations were carried out in R (R Core Team, 2013). In particular, we used code provided by the packages `copula` (Hofert et al., 2014), `fExtremes` (Würtz et al., 2013), `fgof` (Kojadinovic and Yan, 2012a) and `homtest` (Viglione, 2012) available on CRAN.

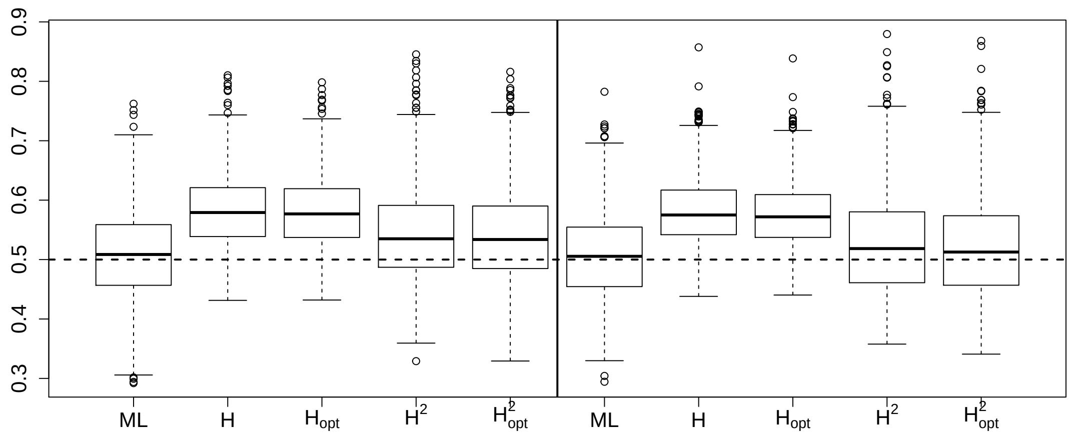

Figure 1: Each box plot is derived from 1000 independent realizations of a joint Hill estimator applied on multivariate data with distribution given in the beginning of Section 4 and $n = 100$, $\gamma = 0.5$, $\delta = 2$, $\boldsymbol{\theta} = (1.5, 2.5)$, $\mathbf{a} = (0.9, 0.7, 0.5, 0.3, 0.1)$, $\boldsymbol{\tau} = (1, 0.9, 0.8, 0.7, 0.6)$, $d = 5$ (left panel) and $d = 15$ (right panel).

4.1 Joint estimation of γ

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be independent copies with the distribution F given above and $\gamma_1 = \ldots = \gamma_d = \gamma$. In contrast to many other comparable studies from hydrology, where typically $\gamma < 0.3$ is used, we are particularly interested in more heavy-tailed scenarios with e.g. $\gamma = 0.5$ (see also our illustration in Section 5). In this case, the L-moment estimator of the shape γ of the GEV distribution is not advisable (**?**, Fig. 1 & 3). This is also confirmed by our simulation results and therefore, we decided to use only a maximum likelihood (ML) based approach $\hat{\gamma}_{ML}$ as a benchmark for the performance of several versions of estimator $\hat{\gamma}_H = \widehat{\gamma}_{\mathbf{k},\boldsymbol{\tau},n}(\boldsymbol{w})$ from (12).
Let $n_j = \lfloor n\tau_j \rfloor$ denote the number of observations available for component j. We consider the following joint estimators of γ:

- $\widehat{\gamma}_{ML} = \sum_{j=1}^{d} w_j \widehat{\gamma}_{ML}^{(j)}$ with $w_j = n_j / \sum_{\ell=1}^{d} n_\ell$ $\quad (ML)$
- $\widehat{\gamma}_H(\boldsymbol{w}_{ind})$ with $k_j = \lfloor 2n_j^{2/3} \rfloor$ $\quad (H)$
- $\widehat{\gamma}_H(\widehat{\boldsymbol{w}}_{opt})$ with $\hat{\Sigma} = \hat{\Sigma}_{ev}$ and $k_j = \lfloor 2n_j^{2/3} \rfloor$ $\quad (H_{opt})$
- $\widehat{\gamma}_H(\boldsymbol{w}_{ind})$ with $k_j^{(d)} = \lfloor 2n_j^{2/3}/d^{1/3} \rfloor$ $\quad (H^{(2)})$
- $\widehat{\gamma}_H(\widehat{\boldsymbol{w}}_{opt})$ with $\hat{\Sigma} = \hat{\Sigma}_{ev}$ and $k_j^{(d)} = \lfloor 2n_j^{2/3}/d^{1/3} \rfloor$ $\quad (H_{opt}^{(2)})$

$\hat{\gamma}_{ML}^{(j)}$ denotes the ML estimator of the GEV distribution applied on the j-th marginal series, $j = 1, \ldots, d$. A simple weighting scheme is applied, which is common practice in hydrology (Hosking and Wallis, 2005). Extensions that also take spatial dependence into account are computationally difficult, e.g. because of complicated likelihood equations. To the best of our knowledge, this problem has not been solved satisfactorily yet.

The performance of four versions of the joint Hill estimator is compared, using simple or asymptotically optimal weights and $k_j = k_j^{(1)} = \lfloor 2n_j^{2/3} \rfloor$ or $k_j = k_j^{(d)} = \lfloor 2n_j^{2/3}/d^{1/3} \rfloor$ (see Section 3.1). We also studied estimators (H_{opt}) and ($H_{opt}^{(2)}$) with $\hat{\Sigma} = \hat{\Sigma}_{emp}$ (not reported here). These, however, are not advisable when the sample lengths n_j are small and dimension d is large because of numerical problems.

We begin with a discussion of our main findings, which can be deduced from Figure 1. Each of the five boxplots on the left ($d = 5$) and on the right ($d = 15$) represents the estimation error of the above estimators, derived from 1000 repetitions with $n = 100$, $\gamma = 0.5$, $\delta = 2$, $\boldsymbol{\theta} = (1.5, 2.5)$ and $\boldsymbol{\tau} = (1, 0.9, 0.8, 0.7, 0.6)$. We want to emphasize the following conclusions that were also present for many other settings: First, the bias of the Hill estimator can be very dominant in the overall estimation error. Second, optimal weighting leads to a small reduction in variability while the bias remains the same as expected. Third, taking the dimension d into account in the choice of $\mathbf{k}$ is important to decrease a possibly strong bias.

Table 1 reports root mean squared errors $\left(E\left[\hat{\gamma} - \gamma\right]^2\right)^{1/2}$ of all five estimators estimated from 1000 independent repetitions for each of many different settings. Generally, the optimal weighting provides only little improvement. As opposed to this, the choice of $k_j^{(d)}$ instead of $k_j^{(1)}$ has a huge impact on the estimation error. In only a few cases, where the bias of Hill's estimator is very small (e.g. $\gamma = 0.5$ and $\delta = 3$), the error increases when using $k_j^{(d)}$ instead of $k_j^{(1)}$ because of an increase in variability. In "typical cases", where the bias is dominant, the incorporation of the dimension d into the choice of upper order statistics greatly improves the performance of the joint Hill estimator.

The observation that optimal weighting provides only a small decrease in estimation error is a little disappointing. Loosely speaking, joint estimation of γ benefits only a little from the asymptotic theory derived in Section 2 in case of small to moderately large samples. This, however, is not true for the test statistic from Proposition 2. In fact, the next subsection demonstrates that the established theory is of key importance in order to achieve an acceptable type 1 error rate.

4.2 Finite-sample performance of W as a test for the null hypothesis $H_{0,IF}$

We studied the finite sample performance of the statistic $W_{\mathbf{k},\boldsymbol{\tau},n}$ as a test for the null hypothesis $H_{0,IF}$ stated in (20). Other established tests for $H_{0,IF}$, which were already compared by simulations in Viglione et al. (2007), are also included in our experiments. The simulation setting used here differs from Viglione et al. (2007) mainly in the following aspects: First, we also take into account possible spatial extreme value dependence. In

Table 1: RMSE's estimated from 1000 independent realizations of five joint Hill estimators applied on extreme valued distributed data with distribution given at the beginning of Section 4 and with $n = 100$, $\boldsymbol{\tau} = (1, 0.9, 0.8, 0.7, 0.6)$ and $\boldsymbol{\theta} = (1.5, 2.5)$.

		$d = 5$					$d = 15$				
γ	Est.	$\delta = \mu/\sigma$									
		1	1.5	2	2.5	3	1	1.5	2	2.5	3
0.25	(ML)	.070	.069	.068	.066	.069	.062	.061	.063	.063	.065
	(H)	.400	.284	.206	.152	.108	.398	.283	.207	.149	.107
	(H_{opt})	.395	.281	.203	.150	.106	.388	.274	.200	.145	.103
	$(H^{(2)})$	.260	.194	.145	.110	.080	.209	.160	.127	.094	.075
	$(H^{(2)}_{opt})$	.258	.192	.143	.108	.079	.204	.155	.122	.091	.070
0.5	(ML)	.080	.079	.079	.078	.078	.077	.074	.077	.075	.072
	(H)	.303	.183	.099	.059	.065	.301	.178	.103	.057	.060
	(H_{opt})	.297	.179	.097	.058	.065	.289	.170	.095	.053	.060
	$(H^{(2)})$	.183	.128	.087	.074	.083	.155	.110	.099	.085	.091
	$(H^{(2)}_{opt})$	.182	.124	.085	.073	.083	.149	.103	.091	.083	.088
0.75	(ML)	.094	.094	.091	.091	.091	.091	.085	.090	.088	.085
	(H)	.236	.122	.085	.121	.164	.237	.116	.086	.120	.168
	(H_{opt})	.231	.118	.084	.121	.164	.225	.107	.084	.121	.170
	$(H^{(2)})$	.161	.123	.111	.133	.154	.154	.130	.130	.140	.162
	$(H^{(2)}_{opt})$	.157	.121	.109	.132	.154	.143	.123	.126	.137	.161

Viglione et al. (2007) only spatially independent samples are considered, and the marginal distributions and hypotheses are formulated in terms of L-moments. We will continue to use the (γ, δ) characterization of marginal distributions.
To give an idea of the other tests, we briefly comment on these procedures:

- The statistic of test HW_1 is similar to that of $W_{\mathbf{k},\boldsymbol{\tau},n}$. For HW_1, each marginal sample ratio of L-scale divided by L-location is compared with a regional version computed from the whole data set. $H_{0,IF}$ is rejected, if the difference between these L-moment ratios is too large.

- HW_2 is similar to HW_1, with an additional term incorporating the distance of L-skewness divided by L-scale. Both, HW_1 and HW_2, are presented by Hosking and Wallis (2005, Chapter 4.3).

- The AD test is based on an Anderson-Darling type distance between marginal empirical distributions and a regional version computed from all available observations. In order to account for possibly different scales under $H_{0,IF}$, all observations are first divided by their marginal sample median.

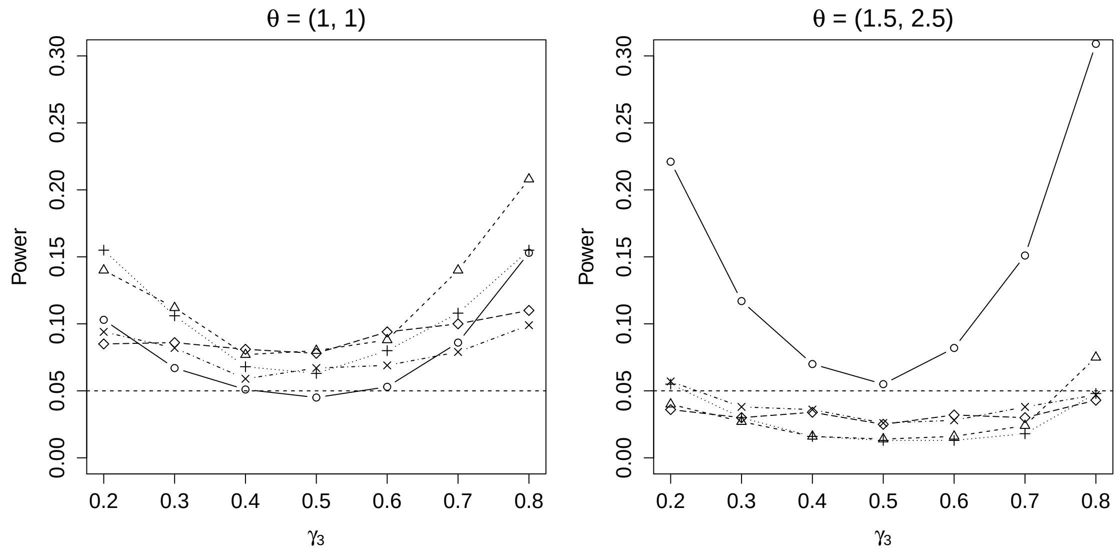

Figure 2: Rejection rates of $H_{1,evi}$ for tests W_{ev} ($\circ$), HW_1 (Δ), HW_2 ($+$), AD ($\times$) and DK ($\diamond$) computed from 4000 samples such that margins $j = 1, 2, 4, 5$ follow a $GEV(\mu = 2, \sigma = 1, \gamma = 0.5)$ and margin $j = 3$ follows a $GEV(\mu = 2, \sigma = 1, \gamma = \gamma_3)$. All 5 margins have sample length $n = 50$ and the spatial dependence corresponds to (21) with $\mathbf{a} = (1, 1)$ (left) and $\mathbf{a} = (1.5, 2.5)$ (right).

- DK is based on a goodness-of-fit statistic proposed by Durbin and Knott (1972). Just like for AD, all observations are first divided by their marginal sample median. The test is based on the fact that if F is the true distribution function of a continuous random variable X, then $U = F(X)$ has a uniform distribution.

We studied two versions of test $W_{\mathbf{k},\boldsymbol{\tau},n}$, with either the empirical estimator $\widehat{\Sigma}_{emp}$ or the CFG-based estimator $\widehat{\Sigma}_{ev}$ plugged in into the statistic. Recall from the discussion in Section 3.2 that the bias of the Hill estimator is less important when the test is applied, provided $H_{0,IF}$ holds. Therefore, we decided to set $k_j = \lfloor 2n_j^{2/3} \rfloor$ for all dimensions d. In order to slightly reduce the type 1 error, we multiplied the statistics with the asymptotically negligible factor $1 - d/(5 \min_j n_j)$ at the cost of a slight loss of power. The corresponding tests are denoted by W_{emp} and W_{ev}, respectively. We address the following questions:

1. How well do the tests keep their nominal level under $H_{0,IF}$?

2. Which test has the largest power against certain alternatives of $H_{0,IF}$? Specifically, against alternatives (a) $H_{1,evi} \cap H_{0,delta}$ or (b) $H_{0,evi} \cap H_{1,delta}$ such that $H_{0,IF}$ holds for the group of four margins $j = 1, 2, 4, 5$ and where margin 3 differs by either $\gamma_3 \neq \gamma$ or $\delta_3 \neq \delta$.

All test were carried out at a nominal level of $\alpha = 5\%$ and with data drawn from multivariate extreme value distributions discussed at the beginning of Section 4.

Table 2: Rejection rates of H_0^{IF} in % computed from 4000 samples under H_0^{IF}. The nominal level is 5%.

		$\boldsymbol{\tau} = (1, 0.9, 0.8, 0.7, 0.6)$									
		$\boldsymbol{\theta} = (1,1)$					$\boldsymbol{\theta} = (1.5, 2.5)$				
γ	Test	$\delta = \mu/\sigma$									
		1	1.5	2	2.5	3	1	1.5	2	2.5	3
		$n = 50$									
0.25	W_{ev}	8.3	4.0	3.8	4.1	4.0	17.3	10.2	7.8	7.8	6.5
	W_{emp}	14.3	9.8	9.2	9.7	9.2	28.5	21.5	20.8	19.6	18.0
	HW_1	3.6	4.6	5.0	4.9	5.1	0.7	1.0	1.1	1.4	1.0
	HW_2	4.2	4.6	5.0	4.6	4.6	1.4	1.5	1.7	1.8	1.6
	AD	4.7	4.3	4.4	4.6	6.2	2.3	2.7	2.5	2.7	3.5
	DK	6.6	3.8	4.1	5.0	6.2	4.0	2.1	2.3	2.5	2.1
0.5	W_{ev}	5.1	3.8	4.8	5.7	6.4	12.2	6.6	6.2	5.2	6.7
	W_{emp}	11.3	10.0	10.4	11.7	12.3	24.9	19.1	17.9	17.9	18.4
	HW_1	7.6	9.0	9.2	9.5	8.0	1.8	2.0	1.9	1.6	1.6
	HW_2	6.3	7.6	6.9	7.8	7.4	2.2	1.8	2.0	1.8	2.1
	AD	4.3	5.0	6.6	7.3	8.4	2.3	2.2	3.1	3.7	4.0
	DK	4.5	4.6	7.6	10.3	13.4	2.5	2.2	3.2	4.5	4.9
0.75	W_{ev}	4.4	5.1	7.0	8.6	11.5	7.0	6.6	5.9	6.5	7.3
	W_{emp}	10.1	11.2	12.9	14.6	17.8	19.7	18.1	17.1	17.8	18.4
	HW_1	16.6	18.6	17.7	15.9	16.4	4.0	5.4	4.3	4.9	4.0
	HW_2	10.2	12.5	11.5	10.8	12.5	3.4	3.8	3.0	4.0	4.2
	AD	4.9	7.3	10.0	12.2	13.4	2.7	4.0	5.5	5.0	7.0
	DK	4.4	9.2	16.4	22.3	28.0	1.9	4.1	7.4	10.5	12.8
		$n = 100$									
0.25	W_{ev}	3.6	2.9	2.6	3.2	3.5	9.8	7.3	6.0	4.8	4.7
	W_{emp}	6.8	5.8	6.2	6.3	6.5	17.5	14.1	12.3	10.9	11.6
	HW_1	3.3	3.8	4.0	3.9	4.0	0.5	0.9	0.5	0.6	0.7
	HW_2	3.5	3.6	4.3	3.7	4.2	1.2	1.1	1.0	1.0	1.1
	AD	5.1	4.5	4.6	5.1	5.2	2.2	2.7	3.0	3.5	3.1
	DK	7.0	5.2	4.9	5.4	7.1	4.4	2.9	2.2	2.2	2.9
0.5	W_{ev}	3.3	3.5	4.5	5.6	6.3	6.9	5.2	5.6	4.9	5.2
	W_{emp}	6.6	6.5	8.2	9.4	10.0	12.8	11.2	11.5	10.5	10.3
	HW_1	6.4	6.0	6.0	6.5	5.4	0.8	1.2	1.1	0.9	0.8
	HW_2	5.8	5.0	5.4	5.7	4.7	1.1	0.9	1.4	0.9	1.5
	AD	4.0	5.3	6.0	7.1	7.6	1.9	2.6	3.4	3.5	3.3
	DK	5.2	5.2	8.8	11.7	13.9	2.5	2.6	3.2	5.1	7.2
0.75	W_{ev}	3.3	3.9	7.2	8.0	10.3	5.5	5.0	5.1	6.6	6.3
	W_{emp}	7.0	7.9	11.1	12.0	15.0	11.2	11.2	10.0	12.3	11.2
	HW_1	9.6	10.7	11.6	10.6	8.8	2.5	2.3	2.3	1.2	1.4
	HW_2	5.9	6.7	8.0	8.6	7.7	2.3	2.5	2.8	1.7	1.4
	AD	4.8	6.0	9.1	9.9	11.8	2.2	3.0	4.7	4.9	4.9
	DK	4.5	11.0	19.3	25.9	32.2	2.0	4.5	9.2	12.0	15.7

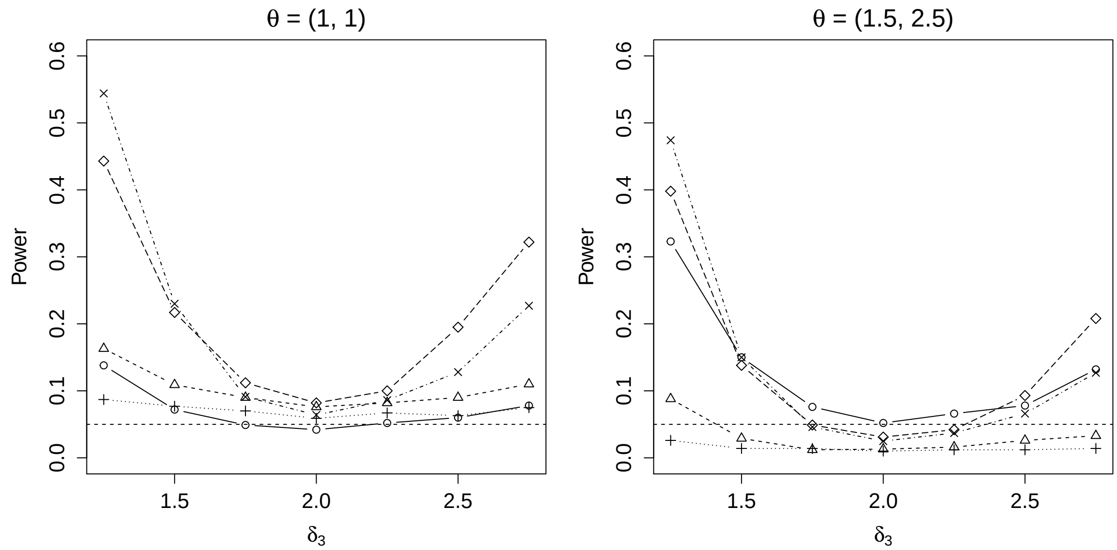

Figure 3: Rejection rates of $H_{1,delta}$ for tests W_{ev} ($\circ$), HW_1 (Δ), HW_2 ($+$), AD ($\times$) and DK ($\diamond$) computed from 4000 samples such that margins $j = 1, 2, 4, 5$ follow a $GEV(\mu = 2, \sigma = 1, \gamma = 0.5)$ and margin $j = 3$ follows a $GEV(\mu = \delta_3, \sigma = 1, \gamma = 0.5)$. All 5 margins have sample length $n = 50$ and the spatial dependence corresponds to (21) with $\mathbf{a} = (1, 1)$ (left) and $\mathbf{a} = (1.5, 2.5)$ (right).

Empirical levels under spatial independence: The left part of Table 2 reports rejection rates in percent of all considered tests estimated from 4000 samples under $H_{0,IF}$ and $\boldsymbol{\theta} = (1, 1)$. The level of W_{emp} is overall not acceptable, whereas test W_{ev} keeps its level reasonably well except for some cases with $\gamma = 0.75$. With increasing heaviness γ of the tails, all other tests fail to get close to the nominal level.

Empirical levels under spatial dependence: The right hand side of Table 2 reports rejection rates as before, but with $\boldsymbol{\theta}$ set to $(1.5, 2.5)$. In case of $\mathbf{a} = (0.9, 0.7, 0.5, 0.3, 0.1)$, this leads to an average Spearman's rho for the pairs of about $\rho = 0.5$. Such a strength of dependence is not uncommon in hydrological applications. Test W_{ev} keeps its level reasonably well, except for some settings with $\gamma = 0.25$. In contrast, all other methods are overall far from attaining the nominal level of 5%, because they do not take into account spatial dependence. We also studied the performance for $\boldsymbol{\tau} = (1, 2)$, which led to an average Spearman's rho of about $\rho = 0.25$. The results were very similar and are therefore not reported here.

Empirical power under $H_{1,evi} \cap H_{0,delta}$: Figure 2 presents rejection rates of tests W_{ev}, HW_1, HW_2, AD and DK under $H_{1,evi} \cap H_{0,delta}$ versus γ_3 estimated from 4000 samples of length $n = 50$ with $\tau = (1, 1, 1, 1, 1)$ such that all but the third component follow a GEV with $\gamma = 0.5$ and $\delta = 2$ and the third component follows a GEV with $\gamma_3 \in \{0.2, 0.3, \ldots, 0.8\}$ and $\delta = 2$. It is remarkable that all tests except W_{ev} have almost

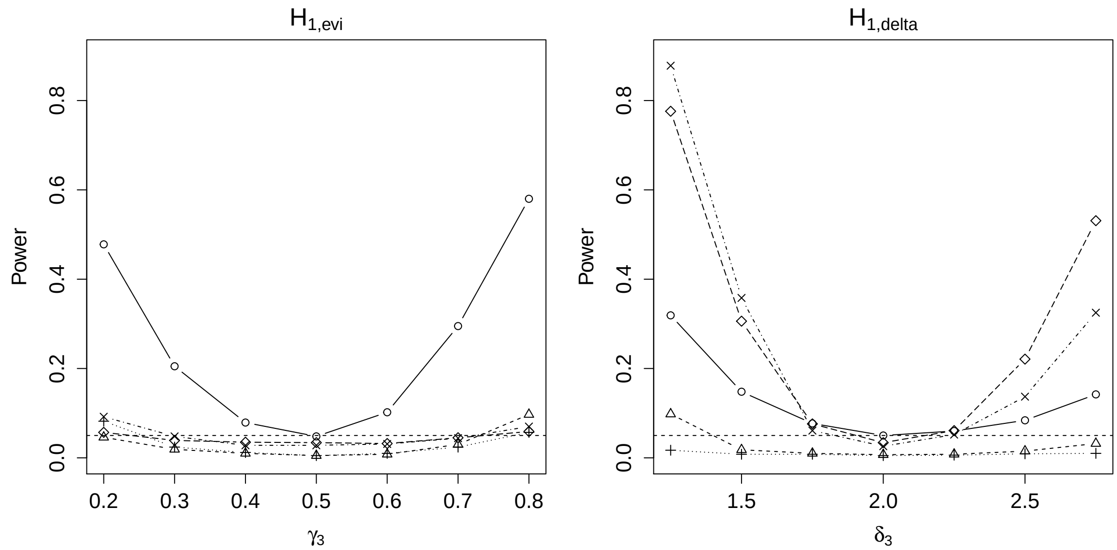

Figure 4: Rejection rates of $H_{1,evi}$ (left) and $H_{1,delta}$ (right) for tests W_{ev} ($\circ$), HW_1 (Δ), HW_2 ($+$), AD ($\times$) and DK ($\diamond$) computed from 4000 samples such that margins $j = 1, 2, 4, 5$ follow a $GEV(\mu = 2, \sigma = 1, \gamma = 0.5)$. Margin $j = 3$ follows a $GEV(\mu = 2, \sigma = 1, \gamma = \gamma_3)$ (left) and $GEV(\mu = \delta_3, \sigma = 1, \gamma = 0.5)$ (right). The spatial dependence corresponds to (21) with $\mathbf{a} = (1.5, 2.5)$. All 5 margins have sample length $n = 100$.

no power under positive dependence (right plot of Fig. 2), while the power of test W_{ev} is even higher than under independence (left plot of Fig. 2). The left plot in Figure 4, where we set $n = 100$, confirms these findings.

Empirical power under $H_{0,evi} \cap H_{1,delta}$**:** Figure 3 presents rejection rates of tests W_{ev}, HW_1, HW_2, AD and DK under $H_{0,evi} \cap H_{1,delta}$ versus δ_3 estimated from 4000 samples of length $n = 50$ with $\tau = (1, 1, 1, 1, 1)$ such that all but the third component follow a GEV with $\gamma = 0.5$ and $\delta = 2$, while the third component follows a GEV with $\gamma = 0.5$ and $\delta_3 \in \{1.25, 1.5, \ldots, 2.75\}$. Although test W_{ev} is designed to detect deviations from $H_{0,evi}$, these results indicate that W_{ev} is rather a test for $H_{0,IF}$. The right plot of Figure 4 depicts results for the same experiment, but with sample length set to $n = 100$. Although tests AD and DK do not take the spatial dependence into account, they are more powerful than W_{ev} in this scenario.

Altogether we conclude that the proposed test W_{ev} keeps its level well in reasonable settings from hydrology. Additionally, the new test is the only one that detects deviations from $H_{0,evi}$ under spatial dependence. On the other hand, test W_{ev} has little power against $H_{1,delta}$ compared to AD and DK. When hypothesis $H_{0,IF}$ is rejected by W_{ev}, tests AD and DK serve as auxiliary tools to indicate whether the deviation from hypothesis $H_{0,IF}$ is due to $H_{1,delta}$ or not.

5 Illustration

Many studies in regional flood frequency analysis focus on peak discharges Q (in m^3/sec) observed at several stations of some region of interest. In order to avoid non-stationarity due to seasonal effects, the block maxima method with block length covering one season is applied on each marginal series. Thanks to the asymptotic theory, these marginal series can be modeled by the parametric class of generalized extreme value distributions (GEV) and the spatial dependence by the nonparametric class of extreme value copulas (de Haan and Ferreira, 2006).

Our region of interest is the Mulde river basin in Saxony, Germany. We have monthly data from 116 stations with between 6 and 100 years of observations per station and an average of about 52 years. Here we focus on the analysis of hydrological summer maxima, namely the maximal peak Q measured between May and October for each station and year available. There are two reasons for restricting to summer maxima. First, most winter floods are produced from melting snow, whereas summer floods are due to short but heavy rainfalls. These very different meteorological causalities lead to different distributions. Second, very high peak flows, which are of particular interest, have been observed only during summer. For our data set of 116 stations, the difference between winter and summer peaks is illustrated in Figure 5. Each point represents a ML fit $(\hat{\gamma}_{ML}, \hat{\delta}_{ML})$ to the generalized extreme value distribution with $\delta = \mu/\sigma$, where a fit is based on either the series of summer ($\circ$) or winter maxima ($*$) of the stations. The size of each point is taken proportional to the corresponding sample length available for estimation. Note that winter and summer maxima are systematically different in distribution and that the range of the summer estimates is covered well by our simulation settings from Section 4.

Canonical correlation analysis (CCA) is popular in flood frequency analysis. It is used to identify homogeneous groups (Ouarda et al., 2001), i.e. groups of stations such that assumption (4) is met. For this, a relationship between some characteristics of a gauge (e.g. the hight and size of the catchment area, mean annual precipitation, slope of main channel, ...) and its peak flow distribution is imposed. CCA identifies dominant linear combinations of (transformed) variables, which are supposed to discriminate best between different stations in terms of their peak flow distributions. A disadvantage of CCA (and other grouping techniques) is that the outcome strongly depends on the choice of variables and other tuning parameters. Different hydrologists will usually derive different groupings. Therefore, it is important to test whether a selected group of stations satisfies the homogeneity assumption or not.

Suppose that the interest is in estimation of γ at some specific station, e.g. station # 16 in Table 3. Because the information available for the target station is unsatisfactory for adequate estimation, we want to incorporate observations from a whole group of stations that shares the same EVI γ.

Based on a CCA, we select a group of 18 stations as possibly homogeneous, which are summarized in Table 3 together with some statistics of interest. The last column of Table 3 consists of p-values of a goodness-of-fit procedure, which evaluates the assumption that a marginal distribution is of GEV type and which is of interest in order to apply a ML

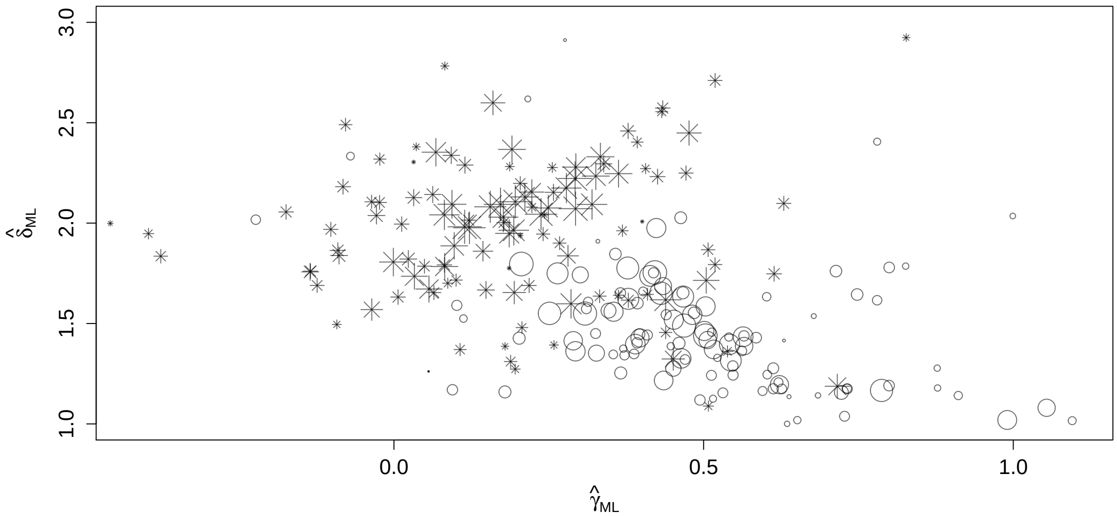

Figure 5: Each point represents a maximum likelihood fit $(\hat{\gamma}_{ML}, \hat{\delta}_{ML})$ of the $GEV(\mu, \sigma, \gamma)$ distribution, where $\delta = \mu/\sigma$. We fitted winter $(*)$ and summer maxima $(\circ)$ series of 116 stations that are located in Saxony, Germany. The size of each point was taken proportional to the available sample length at the corresponding station.

based approach for comparative reasons. More precisely, we applied the test statistic

$$S_n = n \int_{\mathbb{R}} \left[F_n(x) - F_{\hat{\theta}_n}(x) \right]^2 dF_n(x),$$

where F_θ and F_n are the GEV and empirical distribution function, respectively, and $\theta = (\mu, \sigma, \gamma)$ is estimated from the available observations by maximum likelihood. p-values are computed from 1000 parametric bootstrap replicates (Kojadinovic and Yan, 2012b). It should be noted that, however, such goodness-of-fit tests have only little power when the number of observations is small (i.e. $n \leqslant 100$). This, together with the fact that the GEV is an asymptotic model for block maxima distributions (with block size tending to ∞), motivates procedures that are built under less restrictive assumptions like the methods proposed here.

Recall from the discussion in Section 3.2 and from the simulation results in Section 4.2 that $k_j = k_j^{(1)} = \lfloor 2n_j^{2/3} \rfloor$ is appropriate for test W_{ev}, although this choice is not optimal for the joint estimation. With these k_j values we applied test W_{ev} on the selected group. The resulting p-value of $p = 0.02$ indicates that there is strong evidence against assumption (4).

In order to reduce heterogeneity, we examined a scatter plot of the 18 pairs $(\hat{\gamma}_{ML}, \hat{\delta}_{ML})$ from Table 3. The points corresponding to the station numbers 1, 4, 7 and 8 are quite isolated from the others. Moreover, taking into account the multiple testing, there is some weak evidence that the GEV assumption for station # 11 is violated. Overall we excluded stations 1, 4, 7, 8 and 11 and applied test W_{ev} again. The resulting p-value is $p = 0.22$,

making the assumption of homogeneity more plausible than for the larger group considered before.
Interestingly enough, none of the competing tests HW_1, HW_2, AD and DK rejects the homogeneity hypothesis for the larger group. A reason for this is the large spatial dependence, with an average pairwise Spearman's rho value of about 0.66. Recall that in such a case the competing methods are not able to detect deviations from $H_{0,evi}$ (e.g. right plot in Figure 2). In addition, the fact that tests AD and DK remain quite powerful against $H_{1,delta}$ even for dependent data (right plot of Figure 4) suggests that the heterogeneity detected by W_{ev} is indeed due to a violation of assumption $H_{0,evi}$.

The last part of this section deals with the estimation of γ under the assumption that $\gamma_j = \gamma$ holds for all $j \in G = \{1, \ldots, 18\} \backslash \{1, 4, 7, 8, 11\}$. Here the choice of appropriate integers k_j, $j \in G$, is of major importance. A recommended rule for the choice of marginally optimal k_j values is based on the examination of so-called Hill plots $(k, H_{k,n})_{1 \leqslant k < n}$ (Drees et al., 2000). An integer $1 \leqslant k < n$ is chosen such that the plot is approximately constant (stable) in a neighborhood of k. On the other hand, under the assumption that each margin is GEV distributed, we are able to calculate the asymptotically optimal rate of $k_j = \lfloor 2n_j^{2/3} \rfloor$. Interestingly, for our application, both methods yield very similar results, except for station # 18. For that we found that $k_{18} = 12$ is within a stable area in contrast to $\lfloor 2n_{18}^{2/3} \rfloor = 24$.
Recall from the discussion in Section 3.1 and the simulation results in Section 4.2 that the marginally optimal k_j values are not optimal for joint estimation. For the joint estimation we choose $k_j^{(d)} = \lfloor 2n_j^{2/3}/d^{1/3} \rfloor$, $j \neq 18$, and $k_{18}^{(13)} = \lfloor k_{18}/d^{1/3} \rfloor = 5$ with $d = 13$. Together with the asymptotically optimal weights $\hat{\boldsymbol{w}}_{opt}$ estimated under the extreme value dependence assumption we get an estimate of $\hat{\gamma} = 0.43$ with estimated 95% confidence interval $[0.27,\ 0.59]$ derived from (12).
In comparison, the same procedure with marginally optimal integers $k_j = k_j^{(1)} = \lfloor 2n_j^{2/3} \rfloor$ (under the GEV assumption) leads to an estimate of $\hat{\gamma} = 0.59$ with confidence interval $[0.45,\ 0.73]$. The ML based joint estimator $\hat{\gamma}_{ML} = \sum_{j \in G} w_j \hat{\gamma}_{ML,j}$ with weights $w_j = n_j / \sum_{k \in G} n_k$ proportional to the marginal sample lengths gives us $\hat{\gamma}_{ML} = 0.45$, which supports the first estimate rather than the second one, provided the GEV assumption is met for this data set.

6 Conclusion and outlook

The problem of predicting the risk of extreme realizations of heavy-tailed distributions is closely related to the extreme value index (EVI) estimation problem. Recently, Lekina et al. (2014) studied the Weissman estimator and related nonparametric methodology in an univariate hydrological framework. They argue that parametric models are not always appropriate for the estimation of high quantiles in flood frequency analysis. On the other hand, the estimation of nonparametric models is associated with relatively high uncertainty. Typically, these models are useful only in applications with many data points available.

Table 3: A group of 18 stations was selected based on a canonical correlation analysis (not reported here). The statistics were computed from the corresponding summer maxima series.

#	station	obs. years	τ_j	k_j	H_{k_j,n_j}	$\hat{\gamma}_{j,ML}$	$\hat{\delta}_{j,ML}$	GoF p-value
1)	560051	1961-2009	.49	26	.75	.71	1.76	.526
2)	562115	1910-2009	1	43	.54	.42	1.75	.946
3)	563790	1928-2009	.82	37	.63	.48	1.54	.732
4)	564410	1910-2009	1	43	.47	.21	1.80	.185
5)	566010	1936-2009	.74	35	.71	.57	1.39	.261
6)	566040	1926-2009	.84	38	.71	.56	1.43	.435
7)	566100	1961-2009	.49	26	.77	.75	1.64	.469
8)	567400	1960-2009	.50	27	.47	.46	2.03	.168
9)	567451	1910-2009	1	43	.65	.50	1.44	.518
10)	567470	1933-2009	.77	36	.63	.40	1.43	.624
11)	567700	1961-2009	.49	26	.42	.36	1.85	**.005**
12)	567850	1921-2009	.89	39	.50	.38	1.62	.275
13)	568140	1921-2009	.89	39	.58	.47	1.63	.608
14)	568160	1929-2009	.81	37	.72	.45	1.52	.089
15)	568350	1929-2009	.81	37	.59	.36	1.56	.883
16)	576410	1961-2009	.49	26	.65	.48	1.55	.064
17)	576421	1966-2009	.44	24	.66	.37	1.65	.338
18)	577100	1968-2009	.42	12	.42	.33	1.45	.385
mean					.62	.46	1.61	

In regional flood frequency analysis, where we observe the same variable at many gauges, pooling methods are used to overcome the problem of having only short marginal sequences available. The methods proposed here are based on a weighting of marginal Hill estimators initially proposed by Clémençon and Dematteo (2014). Although theory is developed for a broad class of heavy-tailed distributions, we are particularly interested in the applicability to data generated by a componentwise block maxima mechanism.

The main findings from our simulations are as follows: First, the asymptotically optimal weighting scheme for the joint estimation has only little practical benefit in small to moderately large samples. It is more important to incorporate the dimension d into the choice of integers k_j for the number of upper order statistics. Second, the proposed test W_{ev}, which is designed to detect deviations from $H_{0,evi}$ stated in (5), performs rather well as a test for the more restrictive assumption $H_{0,IF}$ stated in (4). While competing procedures considered in the simulations are not able to detect deviations from $H_{0,evi}$ under spatial dependence, the proposed test is powerful in such situations.

Coming back to the framework without extreme value distributional assumption, one might be interested in the estimation of high quantiles $F_j^{-1}(p)$ of some heavy-tailed marginal distribution F_j. This can be achieved by plugging in any consistent estimator $\hat{\gamma}$ of γ into the extrapolation formula of Weissman (1978) known as Weissman estimator. Specifically, if $\hat{\gamma} = \hat{\gamma}_{\mathbf{k},\boldsymbol{\tau},n}(\boldsymbol{w})$, confidence intervals for the Weissman estimator can be deduced from the asymptotic normality of $\widehat{\gamma}_{\mathbf{k},\boldsymbol{\tau},n}(\boldsymbol{w})$ stated in (12) together with de Haan and Ferreira (2006, Theorem 4.3.8).

Acknowledgements

We are grateful to Professor Andreas Schumann from the Department of Civil Engineering, Ruhr-University Bochum, Germany, for providing us hydrological data and for helpful discussions. The financial support of the Deutsche Forschungsgemeinschaft (SFB 823, "Statistical modelling of nonlinear dynamic processes") is gratefully acknowledged.

A Proofs

For sake of readability the proof of Proposition 1 is given only for $d = 2$. Bold symbols are used for (random) vectors, where ^{T} denotes transpose. $\mathbf{1} = (1, \ldots, 1)^T \in \mathbb{R}^d$, $x \wedge y = \max(x, y)$ and $x \vee y = \min(x, y)$. $f_n \sim g_n$ means $f_n/g_n \to 1$ for $n \to \infty$. For ease of presentation, we assume the same beginning and different end points, i.e. we observe $X_1, \ldots, X_{\lfloor n\tau_1 \rfloor} \sim F_X$ and $Y_1, \ldots, Y_{\lfloor n\tau_2 \rfloor} \sim F_Y$ from an i.i.d. process $(X_i, Y_i)_{i \geqslant 1}$. For $t > 1$ we define $a(t) = F_X^{-1}(1 - 1/t)$ and $b(t) = F_Y^{-1}(1 - 1/t)$.

Proof of Proposition 1. The proof for $\tau_1 = \tau_2 = 1$ is treated in Clémençon and Dematteo (2014). Since the proof for $\tau_1, \tau_2 \in (0, 1]$ is similar, some technical details are omitted and can be found in Clémençon and Dematteo (2014).

Recall that

$$\begin{aligned} H^{(1)}_{k_1,\tau_1,n} &= \frac{1}{k_1}\sum_{i=1}^{k_1}\log\frac{X_{\lfloor n\tau_1\rfloor-i+1:\lfloor n\tau_1\rfloor}}{X_{\lfloor n\tau_1\rfloor-k_1:\lfloor n\tau_1\rfloor}} \\ &= \int_1^{\infty}\frac{1}{k_1}\sum_{i=1}^{\lfloor n\tau_1\rfloor}\mathbb{1}\left(X_i > X_{\lfloor n\tau_1\rfloor-k_1:\lfloor n\tau_1\rfloor}\cdot x\right)\frac{dx}{x} \end{aligned}$$

and similarly $H^{(2)}_{k_2,\tau_2,n}$ for the 2nd component.
Let

$$Z_i^X(x) = \mathbb{1}\left(X_i > a\left(\frac{\lfloor n\tau_1\rfloor}{k_1}\right)x\right) - \bar{F}_X\left(a\left(\frac{\lfloor n\tau_1\rfloor}{k_1}\right)x\right)$$

and similarly define $Z_i^Y(y)$, where $\mathbb{1}(A)$ is the indicator function and F_X, F_Y are marginal distribution functions of X_1 and Y_1, respectively. First, we show weak convergence of

$$\left(\frac{1}{\sqrt{k_1}}\sum_{i=1}^{\lfloor n\tau_1\rfloor}Z_i^X(x), \frac{1}{\sqrt{k_2}}\sum_{j=1}^{\lfloor n\tau_2\rfloor}Z_j^Y(y)\right)^T = \left(S_n^X(x,\tau_1), S_n^Y(y,\tau_2)\right)^T = S_n(x,y)$$

in $D^2 = D(\mathbb{R}_+)\times D(\mathbb{R}_+)$. This follows from the proof in Clémençon and Dematteo (2014) and a Cramer-Wold device for D^2 (Davidson, 1994, Theorem 29.16): Let $\lambda = (\lambda_1,\lambda_2)^T \in \mathbb{R}^2$ and without loss of generality $\tau_1 \leqslant \tau_2$. Then

$$\begin{aligned} \lambda^T\cdot S_n(x,y) &= \frac{1}{\sqrt{k_1}}\sum_{i=1}^{\lfloor n\tau_1\rfloor}\left[\lambda_1 Z_i^X(x) + \lambda_2\sqrt{\frac{k_2}{k_1}}Z_i^Y(y)\right] \\ &\quad + \frac{1}{\sqrt{k_2}}\sum_{j=\lfloor n\tau_1\rfloor+1}^{\lfloor n\tau_2\rfloor}\lambda_2 Z_j^Y(y). \end{aligned} \tag{22}$$

For both summands on the right hand side of (22) weak convergence follows from the proof given in Clémençon and Dematteo (2014) (i.e. $\tau_1 = \tau_2 = 1$). Because these two summands are independent, we have weak convergence of $\lambda^T S_n(x,y)$ for each $\lambda\in\mathbb{R}^2$ and thus, by applying the Cramer-Wold device for D^2, we obtain weak convergence of $S_n(x,y)$ in D^2.

Note: From this point on we will write $n\tau$ instead of $\lfloor n\tau\rfloor$, e.g. $X_{n\tau-i:n\tau} = X_{\lfloor n\tau\rfloor-i:\lfloor n\tau\rfloor}$ and $\sum_{i=1}^{n\tau} = \sum_{i=1}^{\lfloor n\tau\rfloor}$.

Now replace the unknown quantities $a\left(\frac{n\tau_1}{k_1}\right)$ and $b\left(\frac{n\tau_2}{k_2}\right)$ by its empirical counterparts

$$\hat{a}\left(\frac{n\tau_1}{k_1}\right) = X_{n\tau_1-k_1:n\tau_1} \quad\text{and}\quad \hat{b}\left(\frac{n\tau_2}{k_2}\right) = Y_{n\tau_2-k_2:n\tau_2}.$$

Apply the map $(f,g) \mapsto (\int_1^\infty f(x)dx/x, \int_1^\infty g(y)dy/y)$. Just as for the proof where $\tau_1 = \tau_2 = 1$, the next step is to show

$$\begin{aligned} &\sqrt{k_1}\left(H^{(1)}_{k_1,\tau_1,n} - \int_{\hat{a}\left(\frac{n\tau_1}{k_1}\right)}^\infty \frac{n\tau_1}{k_1}\bar{F}_X(x)\frac{dx}{x}, H^{(2)}_{k_2,n,\tau_2} - \int_{\hat{b}\left(\frac{n\tau_2}{k_2}\right)}^\infty \frac{n\tau_2}{k_2}\bar{F}_Y(y)\frac{dy}{y}\right) \\ = &\sqrt{k_1}\left(\int_1^\infty \frac{1}{k_1}\sum_{i=1}^{n\tau_1} Z_i^X(x)\frac{dx}{x}, \int_1^\infty \frac{1}{k_2}\sum_{i=1}^{n\tau_2} Z_i^Y(y)\frac{dy}{y}\right) \xrightarrow{D} N(0,\Sigma^*) \end{aligned} \tag{23}$$

for some covariance matrix $\Sigma^* \in \mathbb{R}^{2\times 2}$.
Recall that $c_i = \lim_{n\to\infty}\frac{k_1}{k_i}$ and note that for each component of the vector on the left of (23), the asymptotic distribution does not depend on $\tau_1 \in (0,1]$ and $\tau_2 \in (0,1]$, respectively. Thus, the diagonal elements of Σ^* are the same as for $\tau_1 = \tau_2 = 1$ and given by $\Sigma^*_{i,i} = 2c_i\gamma^2$. The calculation of $\Sigma^*_{1,2}$ requires some more effort. For this, recall that $a, b \in RV_\gamma$ and thus

$$a\left(\frac{n\tau_1}{k_1}\right)\cdot a\left(\frac{n}{k_1}\right)^{-1} \longrightarrow \tau_1^\gamma = (c_1\tau_1)^\gamma,\ b\left(\frac{n\tau_2}{k_2}\right)\cdot b\left(\frac{n}{k_1}\right)^{-1} \longrightarrow (c_2\tau_2)^\gamma.$$

Consequently and because of assumption (10) we have

$$\begin{aligned} &\frac{n}{k_1}P\left(X_1 > a\left(\frac{n\tau_1}{k_1}\right)x, Y_1 > b\left(\frac{n\tau_2}{k_2}\right)y\right) \\ = &\frac{n}{k_1}P\left(X_1 > a\left(\frac{n}{k_1}\right)\frac{a\left(\frac{n\tau_1}{k_1}\right)}{a\left(\frac{n}{k_1}\right)}x, Y_1 > b\left(\frac{n}{k_1}\right)\frac{b\left(\frac{n\tau_2}{k_2}\right)}{b\left(\frac{n}{k_1}\right)}y\right) \\ &\longrightarrow \nu\left(\tau_1^\gamma x, (c_2\tau_2)^\gamma y\right) \end{aligned}$$

uniformly in x, y. Since in addition

$$\frac{n}{k_1}P\left(X_1 > a\left(\frac{n\tau_1}{k_1}\right)x\right)\cdot P\left(Y_1 > b\left(\frac{n\tau_2}{k_2}\right)y\right) \to 0,$$

we arrive at

$$\begin{aligned} &k_1\cdot E\left[\int_1^\infty \frac{1}{k_1}\sum_{i=1}^{n\tau_1} Z_i^X(x)\frac{dx}{x}\cdot\int_1^\infty \frac{1}{k_2}\sum_{i=1}^{n\tau_2} Z_i^Y(y)\frac{dy}{y}\right] \\ \sim\ &\int_1^\infty\int_1^\infty \frac{n(\tau_1\wedge\tau_2)}{k_2}P\left(X_1 > \hat{a}\left(\frac{n\tau_1}{k_1}\right)x, Y_1 > \hat{b}\left(\frac{n\tau_2}{k_2}\right)y\right)\frac{dxdy}{xy} \\ \to\ &c_2(\tau_1\wedge\tau_2)\int_1^\infty\int_1^\infty \nu\left(\tau_1^\gamma x, (c_2\tau_2)^\gamma y\right)\frac{dxdy}{xy} = \Sigma^*_{1,2}. \end{aligned}$$

In the next step we remove the random centering, i.e. in (23) we replace

$$\int_{\hat{a}\left(\frac{n\tau_1}{k_1}\right)}^\infty \frac{n\tau_1}{k_1}\bar{F}_X(x)\frac{dx}{x} \quad\text{and}\quad \int_{\hat{b}\left(\frac{n\tau_2}{k_2}\right)}^\infty \frac{n\tau_2}{k_2}\bar{F}_Y(y)\frac{dy}{y}$$

by γ. Note that

$$\gamma = \lim_{n\to\infty} \int_{a\left(\frac{n\tau_1}{k_1}\right)}^{\infty} \frac{n\tau_1}{k_1} \bar{F}_X(x) \frac{dx}{x} = \lim_{n\to\infty} \int_{b\left(\frac{n\tau_2}{k_2}\right)}^{\infty} \frac{n\tau_2}{k_2} \bar{F}_Y(y) \frac{dy}{y}$$

and

$$\begin{aligned}
& \lim_{n\to\infty} E\left[k_1 \left(H^{(1)}_{k_1,\tau_1,n} - \gamma\right)\left(H^{(2)}_{k_2,\tau_2,n} - \gamma\right)\right] \\
= \; & c_2(\tau_1 \wedge \tau_2) \int_1^{\infty}\int_1^{\infty} \nu\left(\tau_1^{\gamma} x, (\tau_2 c_2)^{\gamma} y\right) \frac{dxdy}{xy} && (24)\\
- \; & \lim_{n\to\infty} E\left[k_1 \int_{a\left(\frac{n\tau_1}{k_1}\right)}^{\hat{a}\left(\frac{n\tau_1}{k_1}\right)} \frac{n\tau_1}{k_1} \bar{F}_X(x)\frac{dx}{x} \int_{b\left(\frac{n\tau_2}{k_2}\right)}^{\hat{b}\left(\frac{n\tau_2}{k_2}\right)} \frac{n\tau_2}{k_2} \bar{F}_Y(y) \frac{dy}{y}\right] && (25)\\
- \; & \lim_{n\to\infty} E\left[k_1 \left(H^{(1)}_{k_1,\tau_1,n} - \gamma\right) \int_{b\left(\frac{n\tau_2}{k_2}\right)}^{\hat{b}\left(\frac{n\tau_2}{k_2}\right)} \frac{n\tau_2}{k_2} \bar{F}_Y(y) \frac{dy}{y}\right] && (26)\\
- \; & \lim_{n\to\infty} E\left[k_1 \left(H^{(2)}_{k_2,n,\tau_2} - \gamma\right) \int_{a\left(\frac{n\tau_1}{k_1}\right)}^{\hat{a}\left(\frac{n\tau_1}{k_1}\right)} \frac{n\tau_1}{k_1} \bar{F}_X(x) \frac{dx}{x}\right]. && (27)
\end{aligned}$$

Thus, it remains to show that

$$(24) - (25) - (26) - (27) \stackrel{!}{=} c_2(\tau_1 \wedge \tau_2)\gamma^2 \cdot \nu\left(\tau_1^{\gamma}, (\tau_2 c_2)^{\gamma}\right). \qquad (28)$$

We proceed in the same way as Clémençon and Dematteo (2014) do (i.e. case $\tau_1 = \tau_2 = 1$, see Lemmas 6.1-6.8 therein). First, we apply Lemmas 6.3 and 6.4 given in Clémençon and Dematteo (2014) to prove modified versions of Lemmas 6.5, 6.6 and 6.7. Next, we use all these auxiliary results to prove modified versions of Lemmas 6.1 and 6.2, which give analytical solutions for (25), (26) and (27). Finally, we prove a modified version of Lemma 6.8 which gives us (28).
For ease of notation, let $p_{i,n\tau} = \frac{n\tau - i + 1}{n\tau}$, $U_i = F_X(X_i)$ and $V_i = F_Y(Y_i)$.

Lemma A.1 (Modified version of Lemma 6.5)
For intermediate sequences k_1, k_2, i.e. $k_j \to \infty$ and $k_j/n \to 0$, we have

$$\begin{aligned}
& E\left[\log X_{n\tau_1 - i+1:n\tau_1} \int_{b\left(\frac{n\tau_2}{k_2}\right)}^{\hat{b}\left(\frac{n\tau_2}{k_2}\right)} \frac{n\tau_2}{k_2} \bar{F}_Y(y) \frac{dy}{y}\right] \\
= \; & M_{n,\tau_1,\tau_2}(i,k_2) + R_{n,\tau_1,\tau_2,1}(i,k_2) + R_{n,\tau_1,\tau_2,2}(i,k_2)
\end{aligned}$$

with $M_{n,\tau_1,\tau_2}(i,k_2)$, $R_{n,\tau_1,\tau_2,1}(i,k_2)$ and $R_{n,\tau_1,\tau_2,2}(i,k_2)$ given in the proof and

$$R_{n,\tau_1,\tau_2,1}(k_1,k_2) = O\left(\frac{n^{-3/2}(\log\log n)^{-1/2}(\log n)^{-1}}{a(n\tau_1/k_1)b(n\tau_2/k_2)}\right), \qquad (29)$$

$$R_{n,\tau_1,\tau_2,2}(k_1,k_2) = O\left(\frac{n^{-3/4}(\log\log n)^{-1/4}(\log n)^{-1/2}}{b(n\tau_2/k_2)}\right). \qquad (30)$$

Proof. With Lemma 6.3 in Clémençon and Dematteo (2014) we have

$$
\begin{aligned}
& E\left[\log X_{n\tau_1-i+1:n\tau_1}\int_{b\left(\frac{n\tau_2}{k_2}\right)}^{\hat{b}\left(\frac{n\tau_2}{k_2}\right)}\frac{n\tau_2}{k_2}\bar{F}_Y(y)\frac{dy}{y}\right] \\
\overset{6.3}{=} \;& E\left[\left(\log a\left(\frac{n\tau_1}{i}\right)+\frac{p_{i,n\tau_1}-\frac{1}{n\tau_1}\sum_{j=1}^{n\tau_1}1_{\{U_j\leqslant p_{i,n\tau_1}\}}}{a\left(\frac{n\tau_1}{i}\right)f_X\left(a\left(\frac{n\tau_1}{i}\right)\right)}\right)\right. \\
& \left.\cdot\left(\frac{p_{k,n\tau_2}-\frac{1}{n\tau_2}\sum_{j=1}^{n\tau_2}1_{\{V_j\leqslant p_{k,n\tau_2}\}}}{b\left(\frac{n\tau_2}{k_2}\right)f_Y\left(b\left(\frac{n\tau_2}{k_2}\right)\right)}\right)\right] \\
+ \;& E\left[O\left(\frac{T_{n\tau_1}(p_{i,n\tau_1})}{a\left(\frac{n\tau_1}{i}\right)}\int_{b\left(\frac{n\tau_2}{k_2}\right)}^{\hat{b}\left(\frac{n\tau_2}{k_2}\right)}\frac{n\tau_2}{k_2}\bar{F}_Y(y)\frac{dy}{y}\right)\right] \\
+ \;& E\left[\log X_{n\tau_1-i+1:n\tau_1}\frac{T_{n\tau_2}(p_{k_2,n\tau_2})}{b\left(\frac{n\tau_2}{k_2}\right)}\right] \\
= \;& M_{n,\tau_1,\tau_2}(i,k_2)+R_{n,\tau_1,\tau_2,1}(i,k_2)+R_{n,\tau_1,\tau_2,2}(i,k_2),
\end{aligned}
$$

where, since $P(V_j \leqslant p) - p = 0$ for $p \in (0,1)$, the first summand is equal to

$$
\begin{aligned}
& M_{n,\tau_1,\tau_2}(i,k_2) \\
= \;& E\left[\frac{\frac{1}{n\tau_1}\sum_{j=1}^{n\tau_1}(1_{\{U_j\leqslant p_{i,n\tau_1}\}}-p_{i,n\tau_1})\frac{1}{n\tau_2}\sum_{j=1}^{n\tau_2}(1_{\{V_j\leqslant p_{k_2,n\tau_2}\}}-p_{k_2,n\tau_2})}{a(n\tau_1/i)f_X(a(n\tau_1/i))b(n\tau_2/k_2)f_Y(b(n\tau_2/k_2))}\right].
\end{aligned}
$$

(29) and (30) follow from Clémençon and Dematteo (2014, Lemma 6.5). □

Lemma A.2 (Modified version of Lemma 6.6)
For $i = 1, \ldots, k$ we have

$$M_{n,\tau_1,\tau_2}(i,k_2) \sim (\tau_1 \wedge \tau_2)\gamma^2\frac{n}{ik_2}P\left(X_1 > a\left(\frac{n\tau_1}{i}\right), Y_1 > b\left(\frac{n\tau_2}{k_2}\right)\right)$$

and in particular

$$M_{n,\tau_1,\tau_2}(k_1,k_2) \sim (\tau_1 \wedge \tau_2)\gamma^2\frac{1}{k_2}\cdot\nu\left(\tau_1^{\gamma},(c_2\tau_2)^{\gamma}\right).$$

Proof. Note that $1 - F_X(a(n\tau_1/i)) = 1 - p_{i,n\tau_1} \sim \frac{i}{n\tau_1}$ and thus, by applying von Mises condition (8), we have

$$a\left(\frac{n\tau_1}{i}\right)f_X\left(a\left(\frac{n\tau_1}{i}\right)\right) \sim \gamma^{-1}\frac{i}{n\tau_1} \quad\text{and}\quad b\left(\frac{n\tau_2}{k_2}\right)f_Y\left(b\left(\frac{n\tau_2}{k_2}\right)\right) \sim \gamma^{-1}\frac{k_2}{n\tau_2}.$$

This leads to

$$
\begin{aligned}
& M_{n,\tau_1,\tau_2}(i,k_2) \\
= \; & E\left[\frac{\frac{1}{n\tau_1}\sum_{j=1}^{n\tau_1}(1_{\{U_j\leqslant p_{i,n\tau_1}\}}-p_{i,n\tau_1})\frac{1}{n\tau_2}\sum_{j=1}^{n\tau_2}(1_{\{V_j\leqslant p_{k_2,n\tau_2}\}}-p_{k_2,n\tau_2})}{a\left(\frac{n\tau_1}{i}\right)f_X(a\left(\frac{n\tau_1}{i}\right))b\left(\frac{n\tau_2}{k_2}\right)f_Y(b\left(\frac{n\tau_2}{k_2}\right))}\right] \\
= \; & \frac{\tau_1\wedge\tau_2}{\tau_1\tau_2}\frac{P\left(X_1>a\left(\frac{n\tau_1}{i}\right),Y_1>b\left(\frac{n\tau_2}{k_2}\right)\right)-(1-p_{i,n\tau_1})(1-p_{k_2,n\tau_2})}{n\cdot a\left(\frac{n\tau_1}{i}\right)f_X(a\left(\frac{n\tau_1}{i}\right))b\left(\frac{n\tau_2}{k_2}\right)f_Y(b\left(\frac{n\tau_2}{k_2}\right))}. \\
\sim \; & (\tau_1\wedge\tau_2)\gamma^2\frac{n}{ik_2}P\left(X_1>a\left(\frac{n\tau_1}{i}\right),Y_1>b\left(\frac{n\tau_2}{k_2}\right)\right).
\end{aligned}
$$

Consequently,

$$
\begin{aligned}
M_{n,\tau_1,\tau_2}(k_1,k_2) &\sim (\tau_1\wedge\tau_2)\gamma^2\frac{1}{k_2}\frac{n}{k_1}P\left(X_1>a\left(\frac{n\tau_1}{k_1}\right),Y_1>b\left(\frac{n\tau_2}{k_2}\right)\right) \\
&\sim (\tau_1\wedge\tau_2)\gamma^2\frac{1}{k_2}\cdot\nu\left(\tau_1^{\gamma},(c_2\tau_2)^{\gamma}\right).
\end{aligned}
$$

□

Lemma A.3 (Modified version of Lemma 6.7)
Write $\bar{F}(a,b)=P(X_1>a,Y_1>b)$. We have

$$
\lim_{n\to\infty}\sum_{i=1}^{k_1}\frac{n}{ik_2}\bar{F}\left(a\left(\frac{n\tau_1}{i}\right),b\left(\frac{n\tau_2}{k_2}\right)\right)=c_2\frac{1}{\gamma}\int_1^{\infty}\nu\left(\tau_1^{\gamma}x,(c_2\tau_2)^{\gamma}\right)\frac{dx}{x} \tag{31}
$$

The proof follows from exactly the same arguments as in Clémençon and Dematteo (2014, Lemma 6.7) and is thus omitted.

Now, by using Lemmas A.1, A.2 and A.3, we are able to calculate (25), (26) and (27). We start with (25):

Lemma A.4 (Modified version of Lemma 6.1)
We have

$$
(25)=c_2(\tau_1\wedge\tau_2)\gamma^2\cdot\nu\left(\tau_1^{\gamma},(c_2\tau_2)^{\gamma}\right). \tag{32}
$$

Proof. We use (Clémençon and Dematteo, 2014, Lemma 6.3) and Lemma A.2 to get

$$
\begin{aligned}
& E\left[k_1 \int_{a\left(\frac{n\tau_1}{k_1}\right)}^{\hat{a}\left(\frac{n\tau_1}{k_1}\right)} \frac{n\tau_1}{k_1} \bar{F}_X(x) \frac{dx}{x} \int_{b\left(\frac{n\tau_2}{k_2}\right)}^{\hat{b}\left(\frac{n\tau_2}{k_2}\right)} \frac{n\tau_2}{k_2} \bar{F}_Y(y) \frac{dy}{y}\right] \\
\overset{6.3}{=} \quad & k_1 E\left[\left(\frac{p_{k_1,n\tau_1} - F^X_{n\tau_1}\left(a\left(\frac{n\tau_1}{k_1}\right)\right)}{a\left(\frac{n\tau_1}{k_1}\right) \cdot f_X\left(a\left(\frac{n\tau_1}{k_1}\right)\right)} + \frac{T_{n\tau_1}(p_{k_1,n\tau_1})}{a\left(\frac{n\tau_1}{k_1}\right)}\right)\right. \\
& \left. \cdot \left(\frac{p_{k_2,n\tau_2} - F^Y_{n\tau_2}\left(b\left(\frac{n\tau_2}{k_2}\right)\right)}{b\left(\frac{n\tau_2}{k_2}\right) \cdot f_Y\left(b\left(\frac{n\tau_2}{k_2}\right)\right)} + \frac{T_{n\tau_2}(p_{k_2,n\tau_2})}{b\left(\frac{n\tau_2}{k_2}\right)}\right)\right] \\
= \quad & k_1 E\left[\frac{\frac{1}{n\tau_1\tau_2}\sum_{i=1}^{n(\tau_1\wedge\tau_2)}\left[P\left(X_i > a\left(\frac{n\tau_1}{k_1}\right), Y_i > b\left(\frac{n\tau_2}{k_2}\right)\right) - (1-p_{k_1,n\tau_1})(1-p_{k_2,n\tau_2})\right]}{a\left(\frac{n\tau_1}{k_1}\right) \cdot f_X\left(a\left(\frac{n\tau_1}{k_1}\right)\right) \cdot b\left(\frac{n\tau_2}{k_2}\right) \cdot f_Y\left(b\left(\frac{n\tau_2}{k_2}\right)\right)}\right] \\
& + o(1) \\
= \quad & \frac{k_1}{k_2} k_2 \cdot M_{n,\tau_1,\tau_2}(k_1,k_2) + o(1) \overset{A\ A.2}{\longrightarrow} c_2(\tau_1 \wedge \tau_2)\gamma^2 \cdot \nu\left(\tau_1^\gamma, (c_2\tau_2)^\gamma\right).
\end{aligned}
$$

□

Next we derive an analytical expression for (26) and similarly for (27):

Lemma A.5 (Modified version of Lemma 6.2)
We have

$$
(26) = c_2(\tau_1 \wedge \tau_2)\left[\gamma \int_1^\infty \nu\left(\tau_1^\gamma x, (c_2\tau_2)^\gamma\right) \frac{dx}{x} - \gamma^2 \nu\left(\tau_1^\gamma, (c_2\tau_2)^\gamma\right)\right]
$$

and

$$
(27) = c_2(\tau_1 \wedge \tau_2)\left[\gamma \int_1^\infty \nu\left(\tau_1^\gamma, (c_2\tau_2)^\gamma y\right) \frac{dy}{y} - \gamma^2 \nu\left(\tau_1^\gamma, (c_2\tau_2)^\gamma\right)\right].
$$

Proof. With the same arguments as in Clémençon and Dematteo (2014, Lemma 6.2) we arrive at

$$
\begin{aligned}
(26) \quad & = \quad \lim_{n\to\infty} \sum_{i=1}^{k_1} \left(M_{n,\tau_1,\tau_2}(i,k_2) - M_{n,\tau_1,\tau_2}(k_1,k_2)\right) \\
& \overset{A\ A.2}{=} \quad (\tau_1 \wedge \tau_2)\gamma^2 \lim_{n\to\infty} \sum_{i=1}^{k_1} \frac{n}{ik_2} P\left(X_1 > a\left(\frac{n\tau_1}{i}\right), Y_1 > b\left(\frac{n\tau_2}{k_2}\right)\right) \\
& \qquad - \lim_{n\to\infty} k \cdot M_{n,\tau_1,\tau_2}(k_1,k_2) \\
& \overset{A\ A.3}{=} \quad c_2(\tau_1 \wedge \tau_2)\left[\gamma \int_1^\infty \nu\left(\tau_1^\gamma x, (c_2\tau_2)^\gamma\right) \frac{dx}{x} - \gamma^2 \nu\left(\tau_1^\gamma, (c_2\tau_2)^\gamma\right)\right].
\end{aligned}
$$

□

Finally, we apply Lemmas A A.4 and A A.5 to show (28):

Lemma A.6 (Modified version of Lemma 6.8) *We have*

$$\int_1^\infty \int_1^\infty \nu\left(\tau_1^\gamma x, (c_2\tau_2)^\gamma y\right) \frac{dxdy}{xy} = \gamma \int_1^\infty \nu\left(\tau_1^\gamma x, (c_2\tau_2)^\gamma\right) \frac{dx}{x} + \gamma \int_1^\infty \nu\left(\tau_1^\gamma, (c_2\tau_2)^\gamma y\right) \frac{dy}{y}$$

and in particular, (28) follows.

Proof. Recall that $\gamma = \int_1^\infty y^{-1/\gamma} \frac{dy}{y}$ and $\nu(tx, ty) = t^{-1/\gamma}\nu(x, y)$ for all $t, x, y > 0$. So,

$$\gamma \int_1^\infty \nu\left(\tau_1^\gamma x, (c_2\tau_2)^\gamma\right) \frac{dx}{x} = \int_1^\infty \int_1^\infty y^{-1/\gamma} \nu\left(\tau_1^\gamma x, (c_2\tau_2)^\gamma\right) \frac{dxdy}{xy}$$
$$= \int_1^\infty \int_1^\infty \nu\left(\tau_1^\gamma yx, (c_2\tau_2)^\gamma y\right) \frac{dxdy}{xy} = \int_1^\infty \int_y^\infty \nu\left(\tau_1^\gamma x, (c_2\tau_2)^\gamma y\right) \frac{dxdy}{xy}$$

and similarly,

$$\gamma \int_1^\infty \nu\left(\tau_1^\gamma, (c_2\tau_2)^\gamma y\right) \frac{dy}{y} = \int_1^\infty \int_x^\infty \nu\left(\tau_1^\gamma x, (c_2\tau_2)^\gamma y\right) \frac{dydx}{yx}.$$

Finally, note that $[1, \infty)^2 = \{(x, y) : \ y \geqslant 1, x \geqslant y\} \dot\cup \{(x, y) : \ x \geqslant 1, y \geqslant x\}$. This completes the proof of Proposition 1. □

Proof of Proposition 2. Let $\mathbf{H}^c_{\mathbf{k},\boldsymbol{\tau},n} = \mathbf{H}_{\mathbf{k},\boldsymbol{\tau},n} - \gamma\mathbf{1}$ and note that we have

$$\mathbf{H}_{\mathbf{k},\boldsymbol{\tau},n} - \hat{\gamma}_{\mathbf{k}}(\boldsymbol{w})\mathbf{1} = A \cdot \mathbf{H}^c_{\mathbf{k},\boldsymbol{\tau},n}$$

with matrix $A = I - \mathbf{1} \cdot \boldsymbol{w}^T \in \mathbb{R}^{d\times d}$ and identity matrix $I \in \mathbb{R}^{d\times d}$.
Let $\mathbf{Z}$ denote a d-dimensional standard normal distributed random vector. By assumption, we have (11) and as a byproduct, $\hat{\gamma}_{\mathbf{k},\boldsymbol{\tau},n}(\boldsymbol{w}) \xrightarrow{P} \gamma$, which, from the continuous mapping theorem, implies that

$$\tilde{W}_{\mathbf{k},\boldsymbol{\tau},n}(\boldsymbol{w}) = \frac{k_1}{\hat{\gamma}_{\mathbf{k},\boldsymbol{\tau},n}(\boldsymbol{w})^2}(A\mathbf{H}^c_{\mathbf{k},\boldsymbol{\tau},n})^T\hat{\Sigma}^{-1}A\mathbf{H}^c_{\mathbf{k},\boldsymbol{\tau},n} \xrightarrow{D} \mathbf{Z}^T\left(A\Sigma^{1/2}\right)^T\Sigma^{-1}A\Sigma^{1/2}\mathbf{Z}.$$

Note that $B = \left(A\Sigma^{1/2}\right)^T\Sigma^{-1}A\Sigma^{1/2}$ is a symmetric matrix. The spectral theorem from linear algebra guarantees the existence of a matrix O with $O^T \cdot O = O \cdot O^T = I$ and a diagonal matrix D containing all eigenvalues of B, such that $B = ODO^T$ holds. Because $O \cdot Z \overset{D}{=} Z$, the latter asymptotic result can be rewritten in the more compact form

$$\tilde{W}_{\mathbf{k},\boldsymbol{\tau},n}(\boldsymbol{w}) \xrightarrow{D} \mathbf{Z}^T D\mathbf{Z}.$$

To complete the first part of the proof, it remains to calculate all d diagonal elements, i.e. the eigenvalues of B:
The first step is achieved by recognizing that $\Sigma^{-1/2}\mathbf{1}$ is an eigenvector with eigenvalue 0. Let $V = \left(span(\Sigma^{-1/2}\mathbf{1}, \Sigma^{1/2}\boldsymbol{w})\right)^\perp$, where $^\perp$ denotes the orthogonal complement, and note that $B \cdot v = v$ for all $v \in V$. Since $\dim(V) \in \{d-2, d-1\}$, at least $d-2$ elements of D are

equal to 1. For $\dim(V) = d-2$, it remains to present one last eigenvalue:
Let $C = (\Sigma^{-1/2}\mathbf{1}, \Sigma^{1/2}\boldsymbol{w}) \in \mathbb{R}^{d\times 2}$,

$$D = \begin{pmatrix} -\boldsymbol{w}^T\Sigma\boldsymbol{w} & 1 \\ 1 & 0 \end{pmatrix} \in \mathbb{R}^{2\times 2}$$

and note that $I - B = CDC^T$. From linear algebra we know that every eigenvalue $\lambda \neq 0$ of $C \cdot DC^T$ is necessarily also an eigenvalue of the matrix $DC^T \cdot C$, which is

$$DC^TC = \begin{pmatrix} 1 - \boldsymbol{w}^T\Sigma\boldsymbol{w}\mathbf{1}^T\Sigma^{-1}\mathbf{1} & 0 \\ \mathbf{1}^T\Sigma^{-1}\mathbf{1} & 1 \end{pmatrix}.$$

From the latter expression we conclude that $\lambda = 1 - \boldsymbol{w}^T\Sigma\boldsymbol{w}\mathbf{1}^T\Sigma^{-1}\mathbf{1}$ is an eigenvalue of $I - B$, which implies that $1-\lambda$ is an eigenvalue of B.
Next note that $\boldsymbol{w}_{opt}^T\Sigma\boldsymbol{w}_{opt}\mathbf{1}^T\Sigma^{-1}\mathbf{1} = 1$ for $\boldsymbol{w}_{opt} = (\mathbf{1}^T\Sigma^{-1}\mathbf{1})^{-1} \cdot \Sigma^{-1}\mathbf{1}$. Finally, from $\hat{\boldsymbol{w}}_{opt} \xrightarrow{P} \boldsymbol{w}_{opt}$ and the continuous mapping theorem, we then have that $W_{\mathbf{k},\boldsymbol{\tau},n} \xrightarrow{D} \chi^2_{d-1}$.
For the remaining part of the proof let F_j satisfy (9), $j = 1, \ldots, d$, but with $\gamma_i \neq \gamma_j$ for some $1 \leqslant i, j \leqslant d$. From Resnick (2007, Theorem 4.2) we have that $\mathbf{H}_{\mathbf{k},\boldsymbol{\tau},n} - \widehat{\gamma}_{\mathbf{k},\boldsymbol{\tau},n}(\boldsymbol{w})\mathbf{1} \xrightarrow{P} \mathbf{b} \in \mathbb{R}^d$, $\mathbf{b} \neq 0$. Together with the positive definiteness of Σ and the consistency of $\hat{\Sigma}$ we have $W_{\mathbf{k},\boldsymbol{\tau},n}/k_1 \xrightarrow{P} const. > 0$, which implies that $W_{\mathbf{k},\boldsymbol{\tau},n} \xrightarrow{P} \infty$. This completes the proof.
□

References

Clémençon, S. and A. Dematteo (2014). On Tail Index Estimation based on Multivariate Data. *ArXiv e-prints*.

Cooley, D., D. Nychka, and P. Naveau (2007). Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association 102*(479), 824–840.

Davidson, J. (1994). *Stochastic Limit Theory : An Introduction for Econometricians.* Oxford University Press.

de Haan, L. and A. Ferreira (2006). *Extreme Value Theory: An Introduction* (Auflage: 2006 ed.). Springer.

Dixon, M. J., J. A. Tawn, and J. M. Vassie (1998). Spatial modelling of extreme sea-levels. *Environmetrics 9*(3), 283–301.

Drees, H., L. de Haan, and S. Resnick (2000). How to make a hill plot. *Ann. Statist. 28*(1), 254–274.

Durante, F. and G. Salvadori (2010). On the construction of multivariate extreme value models via copulas. *Environmetrics 21*(2), 143–161.

Durbin, J. and M. Knott (1972). Components of cramer-von mises statistics. i. *Journal of the Royal Statistical Society. Series B (Methodological) 34*(2), pp. 290–307.

Fuentes, M., J. Henry, and B. Reich (2013). Nonparametric spatial models for extremes: application to extreme temperature data. *Extremes 16*(1), 75–101.

Genest, C. and J. Segers (2009). Rank-based inference for bivariate extreme-value copulas. *Ann. Statist. 37*(5B), 2990–3022.

Gomes, M. I. and D. Pestana (2007). A sturdy reduced-bias extreme quantile (var) estimator. *Journal of the American Statistical Association 102*(477), 280–292.

Hall, P. and A. H. Welsh (1985). Adaptive estimates of parameters of regular variation. *Ann. Statist. 13*(1), 331–341.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist. 3*(5), 1163–1174.

Hofert, M., I. Kojadinovic, M. Mächler, and J. Yan (2014). *copula: Multivariate Dependence with Copulas.* R package version 0.999-9.

Hosking, J. R. M. and J. R. Wallis (2005). *Regional Frequency Analysis: An Approach Based on L-Moments.* Cambridge University Press.

Jarušková, D. and M. Rencová (2008). Analysis of annual maximal and minimal temperatures for some european cities by change point methods. *Environmetrics 19*(3), 221–233.

Khoudraji, A. (1995). *Contributions à l'étude des copules et à la modélisation des valeurs extrêmes bivariées.* Ph. D. thesis, Université Laval, Québec, Canada.

Kojadinovic, I. and J. Yan (2012a). *fgof: Fast Goodness-of-fit Test.* R package version 0.2-1.

Kojadinovic, I. and J. Yan (2012b). Goodness-of-fit testing based on a weighted bootstrap: A fast large-sample alternative to the parametric bootstrap. *Canadian Journal of Statistics 40*(3), 480–500.

Lekina, A., F. Chebana, and T. Ouarda (2014). Weighted estimate of extreme quantile: an application to the estimation of high flood return periods. *Stochastic Environmental Research and Risk Assessment 28*(2), 147–165.

Lettenmaier, D. P., J. R. Wallis, and E. F. Wood (1987). Effect of regional heterogeneity on flood frequency estimation. *Water Resources Research 23*(2), 313–323.

Ouarda, T. B., C. Girard, G. S. Cavadias, and B. Bobée (2001). Regional flood frequency estimation with canonical correlation analysis. *Journal of Hydrology 254*(1–4), 157 – 173.

Pickands, III, J. (1981). Multivariate extreme value distributions. In *Proceedings of the 43rd session of the International Statistical Institute, Vol. 2 (Buenos Aires, 1981)*, Volume 49, pp. 859–878, 894–902. With a discussion.

R Core Team (2013). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing.

Resnick, S. I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling.* Springer.

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

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris 8*, 229–231.

Viglione, A. (2012). *homtest: Homogeneity tests for Regional Frequency Analysis.* R package version 1.0-5.

Viglione, A., F. Laio, and P. Claps (2007). A comparison of homogeneity tests for regional frequency analysis. *Water Resources Research 43*(3).

Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association 73*(364), 812–815.

Würtz, D., many others, and see the SOURCE file (2013). *fExtremes: Rmetrics - Extreme Financial Market Data.* R package version 3010.81.