EXTREMES ON RIVER NETWORKS[1]

BY PEIMAN ASADI[*], ANTHONY C. DAVISON[†] AND SEBASTIAN ENGELKE[*,†]

Université de Lausanne[*] and Ecole Polytechnique Fédérale de Lausanne[†]

Max-stable processes are the natural extension of the classical extreme-value distributions to the functional setting, and they are increasingly widely used to estimate probabilities of complex extreme events. In this paper we broaden them from the usual situation in which dependence varies according to functions of Euclidean distance to situations in which extreme river discharges at two locations on a river network may be dependent because the locations are flow-connected or because of common meteorological events. In the former case dependence depends on river distance, and in the second it depends on the hydrological distance between the locations, either of which may be very different from their Euclidean distance. Inference for the model parameters is performed using a multivariate threshold likelihood, which is shown by simulation to work well. The ideas are illustrated with data from the upper Danube basin.

1. Introduction. Modeling extreme events has recently become of great interest. The financial crisis, heat waves, storms and heavy precipitation underline the importance of assessing rare phenomena when few relevant data are available.

There is a vast literature on modeling the univariate upper tail of the distribution of environmental quantities such as precipitation or river discharges at a fixed location t. If $X_i(t)$ ($i = 1, \ldots, n$) are $n \in \mathbb{N}$ independent measurements of a random spatial process X at location t, then the probability law of the maximum of the n observations can be approximated by the generalized extreme value distribution (GEVD)

$$(1) \quad \mathbb{P}\left\{ \max_{i=1,\ldots,n} \frac{X_i(t) - b_n}{a_n} \le x \right\} \approx G(x) = \exp\{-(1 + \xi x)_+^{-1/\xi}\}, \qquad x \in \mathbb{R},$$

where $z_+ = \max(z, 0)$ and $b_n \in \mathbb{R}$, $a_n > 0$ and $\xi \in \mathbb{R}$ are the location, scale and shape parameters, respectively. For $\xi = 0$, $G(x)$ is read as the limit $\exp\{-\exp(-x)\}$. In fact, (1) represents the only possible nondegenerate limit for maxima of independent and identically distributed sequences of random variables [see, e.g., Coles (2001), Chapter 3]. This justifies the extrapolation to high quantiles using the parametric tail approximation (1) for u close to the upper endpoint

of the distribution of $X(t)$ by

$$(2) \qquad \mathbb{P}\{X(t) > u\} \approx \frac{1}{n}\left(1 + \xi \frac{u - b_n}{a_n}\right)_+^{-1/\xi}.$$

Often, however, univariate considerations are insufficient, because near-simultaneous extreme events may cause the most severe damage. In considering flooding of a river basin, for example, it is crucial to understand the *extremal dependence* between flows at different gauging stations. Many authors have analyzed this using multivariate copulas or multivariate extreme value distributions [e.g., Renard and Lang (2007), Salvadori and De Michele (2010)], but the explosion of the number of parameters in high dimensions limits the applicability of such models, and information on the geographical location of the stations cannot be readily incorporated. Meteorological considerations suggest that extremal dependence can be modeled as a function of the distance between two locations. Indeed, for precipitation, temperature or wind data, the use of *Euclidean distance* has become standard in spatial extremes [e.g., Davison and Gholamrezaee (2012), Engelke et al. (2015), Huser and Davison (2014)]. An important class of probability models for extreme spatial dependence on the Euclidean space $\mathbb{R}^2$ is the class of *max-stable processes*, giving several flexible models whose dependence is parameterized in terms of covariance functions [Opitz (2013), Schlather (2002)] or of negative definite kernels [Brown and Resnick (1977), Kabluchko (2011), Kabluchko, Schlather and de Haan (2009)]. Almost all such models have hitherto presupposed that extremal dependence depends only on the Euclidean distance between two locations, but this may be too restrictive when more is known about the physical processes underlying the data: locations on a river network may interact because of the flow of water downstream between them.

In this paper we focus on assessment of the risk of extreme discharges on river networks in order to understand and prevent flooding. There is longstanding interest in the application of extreme value statistics in hydrology [e.g., Katz, Parlange and Naveau (2002), Keef, Svensson and Tawn (2009), Keef, Tawn and Svensson (2009)]. In Europe, floods are major natural hazards that can end human lives and cause huge material damage. Figure 1 shows the upper Danube basin, which covers most of the German state of Bavaria and parts of Baden-Würtemberg, Austria and Switzerland, and is regularly affected by flooding. For this reason there is a well-developed system of gauging stations that measure the daily average river discharge on this river network; the locations of 31 stations are shown on the map. For each fixed location t_j ($j = 1, \ldots, 31$) on the network, the approximation (2) can be applied to daily measurements $X_i(t_j)$ ($i = 1, \ldots, n$) of river discharge (m³/s) in order to model univariate tail probabilities.

Dependence modeling is more challenging. The extremal coefficient $\theta(t_i, t_j) \in [1, 2]$ measures the degree of dependence of large values at two locations t_i and t_j on the river network; it ranges from $\theta(t_i, t_j) = 1$ for complete dependence to
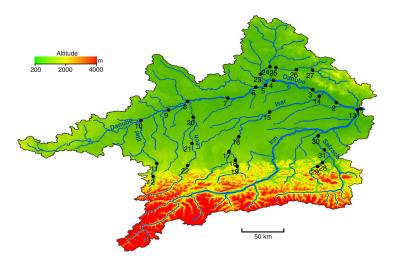
FIG. 1. *Topographic map of the upper Danube basin, showing sites of* 31 *gauging stations (black circles) along the Danube and its tributaries. Water flows toward gauging station* 1.

$\theta(t_i, t_j) = 2$ for independence. The left panel of Figure 2 shows its values for all pairs of stations in Figure 1, plotted against their Euclidean distances. Unlike similar plots for extreme precipitation, the non-Euclidean structure of the network means that this graph shows only a weak relationship.

In this paper we aim to exploit both the geographical structure of the river basin and the hydrological properties of the network in order to provide a parsimonious model for extremal dependence. The resulting dependence function has two parts:
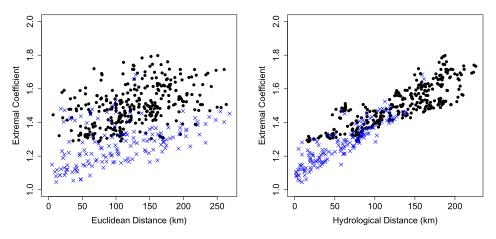

FIG. 2. *Extremal coefficients (estimated using the madogram) of all pairs of gauging stations plotted against Euclidean distance (left) and hydrological distance (right); those for flow-connected pairs are blue crosses, and those for flow-unconnected pairs are black circles.*

- since precipitation is the major source of extreme river discharges and it is spatially dependent, one also expects higher dependence of river discharges at stations which are close. The left panel of Figure 2 suggests that the Euclidean distance between stations has low explanatory power, so we shift each gauging station to a new position in the center of its sub-catchment, which we call its *hydrological position*. The extremal coefficients plotted against the *hydrological distance* between the hydrological positions exhibit a strong functional relationship, shown in the right panel of Figure 2, which is exploited in the dependence model described in Section 3.3;
- the crosses in Figure 2 represent the extremal coefficients of pairs of flow-connected stations, which have one station located upstream of the other. Such pairs are generally more dependent than flow-unconnected pairs, not only because the catchments are close but also owing to the flow of water along the river. In Section 3.2 we explain how knowledge about the network structure and river sizes can be included in the dependence model for flooding using ideas of Ver Hoef and Peterson (2010), who defined covariance functions on river networks.

As one application of such a model, we would like to be able to compute the multivariate counterpart of (2), that is, the probability of a rare event such as

$$\mathbb{P}\{X(s_1) > u_1, \ldots, X(s_k) > u_k\}$$

for large $u_1, \ldots, u_k > 0$, where $s_1, \ldots, s_k \in T$ can be any stations on the river network, even without measurements there. More complicated quantities, such as the sum of discharges at several stations, may also be of interest.

2. Preliminaries.

2.1. *Extreme value theory.* The only nontrivial limiting distribution for the normalized maxima of an independent and identically distributed sequence of scalar random variables is the max-stable GEVD, expression (1). In the multivariate case, we can transform each margin such that the max-limit has a standard Fréchet cumulative distribution function $\exp(-1/x)$ ($x > 0$). In this way, without loss of generality, we can concentrate on the multivariate dependence between the components [Resnick (1987), Proposition 5.8].

Let $\mathbf{X}_i = (X_{1,i}, \ldots, X_{m,i})$ ($i = 1, \ldots, n$) be independent copies of an m-variate random vector $\mathbf{X}$ and assume that for each $j = 1, \ldots, m$ the maximum $\max_i X_{j,i}$ converges to a GEVD G_j, as in (1), with norming constants $b_{j,n} \in \mathbb{R}$, $a_{j,n} > 0$ and shape parameter ξ_j. Define the transformations

$$(3) \qquad U_j(x) = -1/\log G_j(x) = (1 + \xi_j x)_+^{1/\xi_j},$$

and note that

$$\lim_{n \to \infty} \mathbb{P}\left\{ \max_{i=1,\ldots,n} U_j\left(\frac{X_{j,i} - b_{j,n}}{a_{j,n}} \right) \le x \right\} = \exp(-1/x), \qquad j = 1, \ldots, m.$$

We say that $\mathbf{X}$ is in the multivariate maximum domain of attraction (MDA) of a random vector $\mathbf{Z} = (Z_1, \ldots, Z_m)$, if for any $\mathbf{z} = (z_1, \ldots, z_m)$,

$$
\lim_{n \to \infty} \mathbb{P}\left\{ \max_{i=1,\ldots,n} U_1\left(\frac{X_{1,i} - b_{1,n}}{a_{1,n}}\right) \le z_1, \ldots, \max_{i=1,\ldots,n} U_m\left(\frac{X_{m,i} - b_{m,n}}{a_{m,n}}\right) \le z_m \right\}
$$
(4)
$$
= \mathbb{P}(\mathbf{Z} \le \mathbf{z});
$$

call this joint distribution $F_{\mathbf{Z}}(\mathbf{z})$. In this case, $\mathbf{Z}$ is max-stable with standard Fréchet marginal distributions; see before (9). Moreover, by Resnick (1987), Proposition 5.8, we may write

$$
F_{\mathbf{Z}}(\mathbf{z}) = \exp\{-V(\mathbf{z})\}, \qquad \mathbf{z} \in \mathbb{R}^m,
$$
(5)

where the exponent measure V is a measure defined on the cone $E = [0, \infty)^m \setminus \{\mathbf{0}\}$ and $V(\mathbf{z})$ is shorthand for $V([\mathbf{0}, \mathbf{z}]^C)$. The object V incorporates the extremal dependence structure of $\mathbf{Z}$, where $V(\mathbf{z}) = 1/\min(z_1, \ldots, z_m)$ and $V(\mathbf{z}) = 1/z_1 + \cdots + 1/z_m$ represent complete dependence and independence, respectively. The measure V is homogeneous of order -1, that is, $V(\lambda \mathbf{z}) = \lambda^{-1} V(\mathbf{z})$, for $\lambda > 0$, and it satisfies $V(z, \infty, \ldots, \infty) = 1/z$ for $z > 0$ and any permutation of its arguments. There are many parametric models for the exponent measure V and thus for multivariate extreme value distributions or copulas. The explosion of parameters in most such models makes fitting them feasible only in low dimensions.

By Proposition 5.17 of Resnick (1987) the convergence in (4) is equivalent to

$$
\lim_{n \to \infty} n \mathbb{P}\left[\left\{ U_1\left(\frac{X_1 - b_{1,n}}{a_{1,n}}\right), \ldots, U_m\left(\frac{X_m - b_{m,n}}{a_{m,n}}\right) \right\} \in A \right] = V(A)
$$
(6)

for any Borel subset $A \subset E$ which is bounded away from $\mathbf{0}$ and satisfies $V(\partial A) = 0$, where ∂A is the boundary of A. This important observation allows us to approximate the probability that $\mathbf{X}$ falls into a rare region. For instance, if $A = (u_1, \infty) \times \cdots \times (u_m, \infty)$ $(u_1, \ldots, u_m \in \mathbb{R})$, then for large n (6) implies that

$$
\mathbb{P}(X_1 > u_1, \ldots, X_m > u_m) \approx \frac{1}{n} V\left\{ \prod_{j=1}^m \left(U_j\left(\frac{u_j - b_{j,n}}{a_{j,n}}\right), \infty \right) \right\},
$$
(7)

where $\prod$ denotes the Cartesian product. More complicated events such as $A = \{\mathbf{x} \in \mathbb{R}^m : \sum_{i=1}^m x_i > u\}$ for some $u \in \mathbb{R}$ can also be considered. Equation (6) implies that as $n \to \infty$ the empirical point process

$$
\left\{ \left(U_1\left(\frac{X_{1,i} - b_{1,n}}{a_{1,n}}\right), \ldots, U_m\left(\frac{X_{m,i} - b_{m,n}}{a_{m,n}}\right) \right) : i = 1, \ldots, n \right\}
$$

converges vaguely to a Poisson point process on E with intensity measure V [Resnick (1987), Proposition 3.21]. In Section 4 this result will be used to derive the asymptotic distribution of exceedances and to fit parametric models for V.

In the bivariate case $m = 2$, a common summary statistic for the dependence among components of $F_{\mathbf{Z}}$ is the extremal coefficient $\theta \in [1, 2]$ [see, e.g., Schlather and Tawn (2003)], which is defined through the expression

$$(8) \qquad \mathbb{P}(Z_1 \leq u, Z_2 \leq u) = \mathbb{P}(Z_1 \leq u)^{\theta}, \qquad u > 0,$$

or, equivalently, $\theta = V(1, 1)$. Consequently, the cases $\theta = 1$ and $\theta = 2$ correspond to complete dependence and independence. Model-free estimation of the extremal coefficient is possible through the madogram [Cooley, Naveau and Poncet (2006)], and these estimates of θ can be used for model-checking.

2.2. *Max-stable processes.* Max-stable processes can be defined on any index set T, though this is usually taken to be a subset of an Euclidean space $\mathbb{R}^d$. A random process $\{Z(t) : t \in T\}$ is called *max-stable* if there exists a sequence $(X_i)_{i \in \mathbb{N}}$ of independent copies of a process $\{X(t) : t \in T\}$ and functions $a_n(t) > 0$, $b_n(t) \in \mathbb{R}$, such that the convergence

$$(9) \qquad Z(t) = \lim_{n \to \infty} \left\{ \max_{i=1,\dots,n} X_i(t) - b_n(t) \right\} / a_n(t), \qquad t \in T,$$

holds in the sense of finite dimensional distributions. In this case, the process X is said to lie in the max-domain of attraction of Z.

The class of max-stable processes is generally too large for statistical modeling, so one typically considers parametric subclasses of models. Examples include mixed moving maxima processes [Wang and Stoev (2010)], Schlather processes [Schlather (2002)] and Brown–Resnick processes [Brown and Resnick (1977), Kabluchko, Schlather and de Haan (2009)]. In this paper we rely on the construction principle for a large class of max-stable processes given in Kabluchko (2011); see also Kabluchko, Schlather and de Haan (2009). A *negative definite kernel* Γ on an arbitrary nonempty set T is a mapping $\Gamma : T \times T \to [0, \infty)$ such that for any $n \in \mathbb{N}$ and $a_1, \dots, a_n \in \mathbb{R}$ with $\sum_{i=1}^{n} a_i = 0$, we have

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \Gamma(t_i, t_j) \leq 0, \qquad t_1, \dots, t_n \in T.$$

The following result states that there corresponds a max-stable process to any negative definite kernel on T.

THEOREM 2.1 [Kabluchko (2011), Theorem 1]. *Suppose that W_i ($i \in \mathbb{N}$) are independent copies of the zero-mean Gaussian process $\{W(t) : t \in T\}$ whose incremental variance $\mathbb{E}\{W(s) - W(t)\}^2$ equals $\Gamma(s, t)$ for all $s, t \in T$. Let $\sigma^2(t) = \mathbb{E}\{W(t)^2\}$ denote the variance function of W and let $\{U_i : i \in \mathbb{N}\}$ denote a Poisson process on $(0, \infty)$ with intensity $u^{-2}\,du$. Then the process*

$$(10) \qquad \eta_{\Gamma}(t) = \max_{i \in \mathbb{N}} U_i \exp\{W_i(t) - \sigma^2(t)/2\}, \qquad t \in T,$$

is max-stable, has standard Fréchet margins, and its distribution depends only on Γ.

If $T = \mathbb{R}^d$ and W is an intrinsically stationary Gaussian process, then η_Γ is called a Brown–Resnick process [Brown and Resnick (1977), Kabluchko, Schlather and de Haan (2009)]. This is a popular model for complex extreme events. The generation of random samples from Brown–Resnick type processes is challenging [cf. Engelke, Kabluchko and Schlather (2011), Oesting, Kabluchko and Schlather (2012)], but recent advances provide exact and efficient algorithms [Dieker and Mikosch (2015), Dombry, Engelke and Oesting (2016)].

REMARK 2.2. (a) For any negative definite kernel Γ there are many different Gaussian processes with incremental variance Γ [Kabluchko (2011), Remark 1]. In particular, for $u \in T$, we can choose a unique Gaussian process $W^{(u)}$ with incremental variance Γ and $W^{(u)}(u) = 0$ almost surely. The covariance function of this process is

$$(11) \qquad \mathbb{E}\{W^{(u)}(t)W^{(u)}(s)\} = \{\Gamma(s, u) + \Gamma(t, u) - \Gamma(s, t)\}/2.$$

Thus, there is a one-to-one correspondence between negative definite kernels Γ and the class of max-stable processes η_Γ.

(b) If $\{X(t) : t \in T\}$ is a zero-mean Gaussian process with covariance function $C : T \times T \to \mathbb{R}$, then $\Gamma(s, t) = C(s, s) + C(t, t) - 2C(s, t)$ is a negative definite kernel on T.

The bivariate distribution function of $(\eta_\Gamma(s), \eta_\Gamma(t))$ $(s, t \in T)$ is

$$\mathbb{P}\{\eta_\Gamma(s) \le x, \eta_\Gamma(t) \le y\}$$

$$(12) \quad = \exp\left\{-\frac{1}{x}\Phi\left[\frac{\sqrt{\Gamma(s, t)}}{2} + \frac{\log(y/x)}{\sqrt{\Gamma(s, t)}}\right] - \frac{1}{y}\Phi\left[\frac{\sqrt{\Gamma(s, t)}}{2} + \frac{\log(x/y)}{\sqrt{\Gamma(s, t)}}\right]\right\},$$
$$x, y > 0,$$

where Φ is the standard normal distribution function. Analogously to the extremal coefficient in (8), one considers the *extremal coefficient function* $\theta(s, t)$ $(s, t \in T)$, defined as the extremal coefficient of the bivariate vector $(\eta_\Gamma(s), \eta_\Gamma(t))$, as a measure of the functional extremal dependence of the max-stable process η_Γ. By (12), we conclude that

$$(13) \qquad \theta(s, t) = 2\Phi\left\{\frac{\sqrt{\Gamma(s, t)}}{2}\right\},$$

so the negative definite kernel Γ parameterizes the extremal dependence between observations at positions s and t; small and large values of $\Gamma(s, t)$ correspond to strong and weak dependence, respectively. By Remark 2.2(a), any kernel Γ yields a max-stable process η_Γ, so in Section 3 we can and will focus on finding a parametric model for Γ suitable for our application.

The higher dimensional distributions of η_Γ are more complicated. For instance, for $\mathbf{t} = (t_1, \ldots, t_m) \in T^m$, the random vector $(\eta_\Gamma(t_1), \ldots, \eta_\Gamma(t_m))$ is max-stable
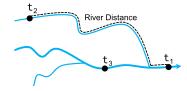
FIG. 3.   *River network with three locations $t_1, t_2, t_3 \in T$; t_1 is flow-connected with both t_2 and t_3, but t_2 and t_3 are flow-unconnected.*

and its exponent measure $V_{\Gamma, \mathbf{t}}$ defined in (5) is characterized by [Kabluchko (2011)]

$$(14) \qquad V_{\Gamma, \mathbf{t}}(x_1, \dots, x_m) = \mathbb{E}\left[\max_{i=1,\dots,m}\left\{\frac{W(t_i) - \sigma^2(t_i)/2}{x_i}\right\}\right].$$

This multivariate max-stable distribution is called the Hüsler–Reiss distribution [Hüsler and Reiss (1989)]. Computation of the expected value in (14) involves high-dimensional integrals and thus is awkward in general.

3. Model.

3.1. *River network.* In the previous section we showed how to define max-stable processes on an arbitrary index set T. From here on, T will represent a river network and we will construct a kernel Γ flexible enough to explain the extremal dependence observed in data.

Let us first fix some notation for river networks [Ver Hoef and Peterson (2010)]. We embed our network T in the Euclidean space $\mathbb{R}^2$ representing the geographical river basin. To this end, let $T \subset \mathbb{R}^2$ denote the collection of piecewise differentiable curves, called river segments, that are connected at the junctions of the river and whose union constitutes the river network. There is a finite number $M \in \mathbb{N}$ of such segments and we index them by $i \in \mathcal{S} = \{1, \dots, M\}$. The network is dendritic, in the sense that there is one most downstream segment, which splits up into other segments when going upstream; see Figure 3. For a location $t_i \in T$ on the ith segment, we let $D_i \subseteq \mathcal{S}$ denote the index set of river segments downstream of t_i, including the ith segment. Moreover, for another location $t_j \in T$ on the jth segment we say that t_i and t_j are *flow-connected*, written $t_i \leftrightarrow t_j$, if and only if $D_i \subseteq D_j$ or $D_j \subseteq D_i$. If t_i and t_j are not flow-connected, we say that they are *flow-unconnected* and write $t_i \not\leftrightarrow t_j$. If t_j is *upstream* of t_i, that is, $D_i \subset D_j$, then we denote the set of segments between t_j and t_i, inclusive of the jth but exclusive of the ith segment, by $B_{i,j} = D_j \setminus D_i$. If t_j is downstream of t_i, then $B_{i,j} = D_i \setminus D_j$. In the case that t_i and t_j are on the same segment, that is, $D_i = D_j$, we put $B_{i,j} = \varnothing$.

We define the *river distance* $d(t_1, t_2)$ between two arbitrary points t_1, t_2 on the network T as the shortest distance along T, that is, we sum the arc-lengths of the
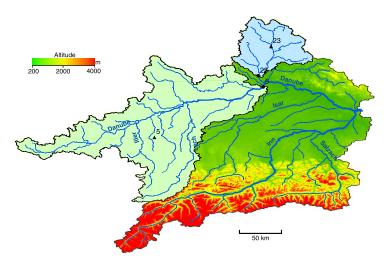
FIG. 4. *Gauging stations* 5 *and* 23 (*black circles*), *their sub-catchments in light green and blue,* *respectively, and their hydrological locations* (*black triangles*) *as defined in* (17).

segment curves lying between t_1 and t_2; see Figure 3. The embedding of the river network T in the Euclidean space $\mathbb{R}^2$ has the advantage that we can exploit the geographical structure of the river basin. To this end, associate to each location $t = (x, y) \in T \subset \mathbb{R}^2$ the set $S_t \subset \mathbb{R}^2$ of all points on the geographical map such that water from this point will eventually flow through point t on the river. The set S_t is called the sub-catchment of location t; see Figure 4.

As explained in Section 2.2, we need to construct a negative definite kernel Γ on the space $T \times T$ that captures the dependence structure of extreme values on the river network T. Figure 2 suggests that this should be based on two components: one, Γ_{Riv}, for the flow-connected dependence along the river, taking into account the hydrological properties of the river network; and another, Γ_{Euc}, for the dependence resulting from the geographical structure of the river basin and spatially distributed meteorological variables.

3.2. *Dependence measure* Γ_{Riv}. There are many models for Gaussian random fields where the covariance between two locations depends only on the Euclidean distance between two points. Such covariances are not valid with metrics such as the river distance d on our network because they may not be positive definite. Recent work [Cressie et al. (2006), Ver Hoef and Peterson (2010), Ver Hoef, Peterson and Theobald (2006)] has developed covariances that are positive definite as functions of river distance. A related approach, the top-kriging of Skøien, Merz and Blöschl (2006), uses variograms integrated over catchments, but does not provide closed-form formulae, so we focus on river distance methods.

Following the "upstream construction" in Ver Hoef, Peterson and Theobald (2006), we can define a covariance function based on river distance for $t_i, t_j \in T$ by

$$(15) \qquad C_{\text{Riv}}(t_i, t_j) = \begin{cases} \prod_{k \in B_{i,j}} \sqrt{\pi_k} C_1\{d(t_i, t_j)\}, & t_i \leftrightarrow t_j, \\ 0, & t_i \not\leftrightarrow t_j, \end{cases}$$

where the covariance function C_1 arises from a moving average construction on $\mathbb{R}$. If $B_{i,j} = \varnothing$ in (15), then $\prod_{k \in B_{i,j}} \sqrt{\pi_k}$ is set to 1. The corresponding weights π_k ($k \in B_{i,j}$) are chosen such that the variance is constant, that is, $C_{\text{Riv}}(t_i, t_i) = C_{\text{Riv}}(t_j, t_j) = C_1(0)$ for all $t_i, t_j \in T$. For a fuller treatment, see Ver Hoef, Peterson and Theobald (2006) and Ver Hoef and Peterson (2010), who also provide different parametric classes for the covariance function C_1, including the *linear with sill model*

$$C_1(h) = (1 - h/\tau)_+, \qquad \tau > 0,$$

which we use below. Intuitively, the covariance function (15) can be understood as follows: an event at a downstream location, for example, t_1 in Figure 3, can be caused by an event on one of the two branches of an upstream bifurcation. The weights π_k quantify the proportions of events coming from the branches. If several bifurcations lie between two flow-connected locations, then the weights along the connection must be multiplied. The choice of the weights in the covariance function C_{Riv} in (15) is crucial and depends on the application. As we consider extreme discharges on river networks, the weights at a bifurcation should reflect the proportion of large discharge values at the downstream river that are caused by a large discharge of one of the upstream rivers. In Figure 3, for example, a natural choice for the weights π_2, π_3 on the river segments of t_2, t_3 is to take the proportion of mean water volumes, that is, $\pi_i = E_{t_i}/(E_{t_2} + E_{t_3})$, where E_{t_i} is the average discharge at location t_i ($i = 2, 3$). This, however, requires measurements at all bifurcations. Since we would like to use our model for extrapolation to parts of the network without measurements, we must approximate E_{t_1} and E_{t_2}. A digital elevation model can be used to extract the geographical coordinates of the sub-catchment S_t corresponding to each location $t \in T$ on the river network, including the altitude $h(x, y)$ at all $(x, y) \in S_t$. Exploratory analysis shows that altitude is an excellent covariate for average precipitation, so we define E_t^* as the integrated altitude over S_t, that is,

$$E_t^* = \int_{S_t} h(x, y) \, dx \, dy,$$

which is thus approximately proportional to the average runoff accumulated in the sub-catchment S_t. We then define the weights in the above example to be

$$(16) \qquad \pi_i = E_{t_i}^*/(E_{t_2}^* + E_{t_3}^*), \qquad i = 2, 3.$$

By the second part of Remark 2.2 and the construction of the positive definite covariance function in (15), we obtain a negative definite kernel Γ_{Riv} on the river network T by setting

$$\Gamma_{\text{Riv}}(t_i, t_j) = \begin{cases} 1 - \prod_{k \in B_{i,j}} \sqrt{\pi_k}(1 - d(t_i, t_j)/\tau)_+, & t_i \leftrightarrow t_j, \\ 1, & t_i \nleftrightarrow t_j. \end{cases}$$

3.3. *Dependence measure* Γ_{Euc}. Two flow-unconnected locations on the river network can have dependent extreme discharges, since precipitation is spatially dependent. As shown in Figure 2, the usual Euclidean distance between two points cannot fully explain this dependence, because the total amount of water at location $t \in T$ on the river network comes not only from precipitation there, but also from the accumulated runoff from its sub-catchment S_t. Thus, instead of the Euclidean distance between two points $s, t \in T$, we should consider a *hydrological distance* that appropriately describes the distance between runoff in sub-catchments S_s and S_t due to precipitation. For this purpose we first shift each location $t \in T$ to a *hydrological location* by a function $H : T \to \mathbb{R}^2$. In our case, the center of mass of mean annual precipitation on the sub-catchment S_t gives a good choice [Merz and Blöschl (2005)]. As noted in Section 3.2, precipitation data on a dense grid is often difficult to obtain, so we use the altitude $h(x, y)$ at location $(x, y) \in S_t$ instead.

The hydrological location $H(t)$, or "altitude weighted centroid," of a point on the river network is

$$(17) \quad H(t) = \left(\frac{1}{E_t^*} \int_{S_t} x h(x, y) \, dx \, dy, \frac{1}{E_t^*} \int_{S_t} y h(x, y) \, dx \, dy \right)^{\text{T}}, \qquad t \in T,$$

and the hydrological distance between $s, t \in T$ is $\|H(s) - H(t)\|$, where $\|\cdot\|$ denotes Euclidean distance. Figure 4 shows two stations on the river network that are close in terms of Euclidean distance but whose hydrological locations are far apart. The right-hand panel of Figure 2 reveals strong functional dependence of the extremal coefficients on hydrological distance.

A variogram that is valid on the Euclidean space $\mathbb{R}^2$ can be applied to the hydrological positions $H(t)$ ($t \in T$). The fractal variogram family $\Gamma_\alpha(x, y) = \|x - y\|^\alpha$ ($x, y \in \mathbb{R}^2$), where $\alpha \in (0, 2]$ is called the shape parameter, is commonly used, but it is isotropic: the dependence decreases at the same rate in each direction. Extremal meteorological data often exhibit anisotropies that can be captured by including a rotation and dilation matrix [Blanchet and Davison (2011), Engelke et al. (2015)]

$$(18) \quad R \equiv R(\beta, c) = \begin{pmatrix} \cos\beta & -\sin\beta \\ c\sin\beta & c\cos\beta \end{pmatrix}, \qquad \beta \in [\pi/4, 3\pi/4], c > 0,$$

where the restriction of β to one quadrant ensures the identifiability of the parameters (β, c). Applying the kernel Γ_α and transformation R to the positions $H(t)$,

we obtain a negative definite kernel on the river network T, that is,

$$\Gamma_{\text{Euc}}(t_i, t_j) = \| R \cdot H(t_i) - R \cdot H(t_j) \|^{\alpha}, \qquad t_i, t_j \in T,$$

where $R \cdot v$ denotes matrix multiplication of R and the vector $v \in \mathbb{R}^2$.

3.4. *Max-stable process on T*. In Sections 3.2 and 3.3 we defined two negative definite kernels on the river network T: Γ_{Riv} models the extremal dependence of flow-connected stations due to the specific hydrological properties of the river network, and Γ_{Euc} describes additional dependence between all stations due to the geographical structure of the river basin and spatially distributed precipitation. We combine these to obtain our final dependence model: given weights $\lambda_{\text{Riv}}, \lambda_{\text{Euc}} \geq 0$, we put

$$
\begin{aligned}
(19) \quad \Gamma(t_i, t_j) &= \lambda_{\text{Riv}} \Gamma_{\text{Riv}}(t_i, t_j) + \lambda_{\text{Euc}} \Gamma_{\text{Euc}}(t_i, t_j) \\
&= \begin{cases} \lambda_{\text{Riv}} \left\{ 1 - \prod_{k \in B_{i,j}} \sqrt{\pi_k} (1 + d(t_i, t_j)/\tau)_+ \right\} \\ \quad + \lambda_{\text{Euc}} \| R \cdot H(t_i) - R \cdot H(t_j) \|^{\alpha}, & t_i \leftrightarrow t_j, \\ \lambda_{\text{Riv}} + \lambda_{\text{Euc}} \| R \cdot H(t_i) - R \cdot H(t_j) \|^{\alpha}, & t_i \not\leftrightarrow t_j, \end{cases}
\end{aligned}
$$

for any $t_i, t_j \in T$. By Remark 2.2 we can define a Gaussian random field W on T with variogram Γ, and by Theorem 2.1 we obtain a max-stable process η_Γ on T, defined in (10), with dependence function Γ. The process η_Γ is nonstationary: indeed, since it is not defined on a Euclidean space, even the notion of stationarity is unclear.

The process η_Γ has standard Fréchet margins. However, even after normalization of the data with scale and location parameters at each location $t \in T$ as in (1), the univariate tail distributions will have different shapes. We must therefore transform the standard Fréchet margins in (10) to GEVD. We set

$$(20) \qquad \tilde{\eta}_\Gamma(t) = \frac{\eta_\Gamma(t)^{\xi(t)} - 1}{\xi(t)}, \qquad t \in T,$$

where $\xi(t) \in \mathbb{R}$ is the shape parameter at point $t \in T$. It is then easily verified that the margins of $\tilde{\eta}_\Gamma$ follow a GEVD, that is,

$$\mathbb{P}\{\tilde{\eta}_\Gamma(t) \leq x\} = \exp[-\{1 + \xi(t)x\}_+^{-1/\xi(t)}], \qquad x \in \mathbb{R}.$$

4. Inference.

4.1. *General*. Inference for the extremes of univariate data is well developed [Coles (2001), de Haan and Ferreira (2006), Embrechts, Klüppelberg and Mikosch (1997)], so we merely sketch it in Section 4.2. Statistical inference for

multivariate or spatial models is more difficult, as their distributions are rarely known in closed form or involve high-dimensional integration. Composite likelihood methods based on bivariate densities have therefore been widely applied [Davison and Gholamrezaee (2012), Huser and Davison (2014), Padoan, Ribatet and Sisson (2010)]. Recent research has focused on methods that exploit full likelihoods of multivariate extreme observations through peaks-over-threshold approaches [Wadsworth and Tawn (2014), Engelke et al. (2015), Thibaud and Opitz (2015), Bienvenüe and Robert (2014)] and on M-estimators for spatial extremes [Einmahl et al. (2015)]. However, different definitions of an extreme event yield different inferences. One might call a multivariate observation extreme if at least one component is large, leading to multivariate generalized Pareto distributions [Rootzén and Tajvidi (2006)], whereas choosing data where a single fixed component exceeds a high threshold gives a conditional extreme value model [Heffernan and Tawn (2004)], and spectral estimation is based on observations where a suitable norm of the components is large [cf. Coles and Tawn (1991)]. For finite samples each choice has advantages and disadvantages [Huser, Davison and Genton (2014)].

We consider two estimation procedures tailor-made for a max-stable process η_Γ whose finite-dimensional margins follow the Hüsler–Reiss distribution (14). Engelke et al. (2015) compute the spectral density of the exponent measure (14) and introduce an estimator for the parameters of a Brown–Resnick process [Kabluchko, Schlather and de Haan (2009)]. Wadsworth and Tawn (2014) use events for which at least one component exceeds a high threshold, and censor any components that stay below it.

In Section 4.3 we review these two methods, show how they can be adapted to our framework, and derive a new representation of the conditional densities, simpler than that in Wadsworth and Tawn (2014). Asadi, Davison and Engelke (2015) describe a small simulation study that aids in the choice of estimator for our application.

4.2. *Univariate margins.* We must estimate the univariate extreme value parameters, that is, the norming constants $a_{j,n}$, $b_{j,n}$, and the shape parameter ξ_j ($j = 1, \ldots, m$) in (1). This allows the calculation of univariate return levels at each location and is needed for the transformations $U_{j,n}$ in (3) that appear in the multivariate exceedance probabilities (7). We use the Poisson point process approach [Coles (2001), Section 7.3] to fit these models for the univariate exceedances.

Recall that $\mathbf{X}_i = (X_{1,i}, \ldots, X_{m,i})$ ($i = 1, \ldots, n$) are independent copies of an m-variate random vector $\mathbf{X}$ as in Section 2.1. For each location $j = 1, \ldots, m$, let $q_{j,p}$ be the empirical p-quantile, with $p \approx 1$, of the data $X_{j,1}, \ldots, X_{j,n}$, and write $\mathcal{I}_j = \{i \in \{1, \ldots, n\} : X_{j,i} > q_{j,p}\}$. Then the Poisson point process likelihood for the exceedances at station t_j, assumed independent, can be written as

[Coles (2001), (7.9)]

$$
L(\xi_j, a_{j,n}, b_{j,n}) \propto \exp\left\{-n_j\left[1 + \xi_j\left(\frac{q_{j,p} - b_{j,n}}{a_{j,n}}\right)\right]^{-1/\xi_j}\right\}
$$

(21)

$$
\times \prod_{i \in \mathcal{I}_j} a_{j,n}^{-1}\left[1 + \xi_j\left(\frac{X_{j,i} - b_{j,n}}{a_{j,n}}\right)\right]^{-1/\xi_j - 1},
$$

where n_j is the number of years of observations at location t_j. Owing to the inclusion of n_j, the parameters $a_{j,n}, b_{j,n}$ and ξ_j equal those in the GEVD (1) for yearly maxima. A joint model for the parameters at different locations, such as a linear model with environmental covariates, can be fitted by maximizing a so-called independence likelihood [Chandler and Bate (2007)] based on the product of (21) over all stations.

4.3. *Estimation of η_Γ.* In order to fit the max-stable process η_Γ introduced in Section 3 with dependence kernel (19), we must estimate the six parameters

$$
\lambda_{\text{Riv}} \geq 0, \qquad \lambda_{\text{Euc}} \geq 0, \qquad \tau > 0,
$$

(22)

$$
\alpha \in (0, 2], \qquad \beta \in [\pi/4, 3\pi/4], \qquad c > 0,
$$

that characterize the river and Euclidean dependence functions Γ_{Riv} and Γ_{Euc} and their weights. Below we write $\vartheta = (\lambda_{\text{Riv}}, \lambda_{\text{Euc}}, \tau, \alpha, \beta, c)$, and denote the corresponding parameter space by Θ. When stressing that Γ depends on the parameter ϑ, we write $\Gamma = \Gamma_\vartheta$.

We do not observe data from the asymptotic limit model η_Γ itself, so let us specify the assumptions for our observations. As in Section 3, let T denote the river network and assume that we have n observations $\mathbf{X}_1, \ldots, \mathbf{X}_n \in \mathbb{R}^m$ at m locations $\mathbf{t} = (t_1, \ldots, t_m) \in T^m$. Further, suppose that the data are normalized to standard Pareto margins with cumulative distribution function $1 - 1/x$ ($x \geq 1$) and that the vectors $\mathbf{X}_k$ ($k = 1, \ldots, n$) are independent copies of a random vector $\mathbf{X}$ in the max-domain of attraction of the max-stable process $\eta_\Gamma(\mathbf{t}) = (\eta_\Gamma(t_1), \ldots, \eta_\Gamma(t_m))$. This means that

(23)
$$
\lim_{n \to \infty} n\mathbb{P}(\mathbf{X}/n \in A) = V_{\Gamma, \mathbf{t}}(A),
$$

for any Borel subset $A \subset E$ which is bounded away from $\mathbf{0}$ and which has zero $V_{\Gamma, \mathbf{t}}$ measure on its boundary; recall the definition of the exponent measure in Section 2.1.

4.3.1. *Spectral estimation of Γ_ϑ.* The random vector $\eta_\Gamma(\mathbf{t})$ follows a multivariate Hüsler–Reiss distribution. Even though its multivariate densities are not available, the densities of its exponent measure $V_{\Gamma, \mathbf{t}}$ have closed forms for any dimensions and we can apply the spectral estimator proposed by Engelke et al.

(2015). Indeed, for large thresholds $u > 0$ the convergence in (23) justifies the approximation

$$(24) \qquad \mathbb{P}(\mathbf{X} \in d\mathbf{x}, \|\mathbf{X}\|_1 > u) \approx -\frac{\partial^m}{\partial x_1 \cdots \partial x_m} V_{\Gamma, \mathbf{t}}(x_1, \ldots, x_m) \, d\mathbf{x},$$

where $\|\mathbf{x}\|_1 = \sum_{j=1}^m x_j$ ($\mathbf{x} \in E$) denotes the L_1-norm, and $V_{\Gamma, \mathbf{t}}(\{\mathbf{x} \in E : \|\mathbf{x}\|_1 > 1\}) = m$. Owing to the homogeneity of the exponent measure $V_{\Gamma, \mathbf{t}}$ in Section 2.1, it suffices to specify the angular part of (24), namely, its *spectral density* on the positive L_1-sphere $S_{m-1} = \{\mathbf{x} \geq \mathbf{0} : \|\mathbf{x}\|_1 = 1\} \subset \mathbb{R}^m$ [Coles and Tawn (1991)]. Engelke et al. (2015) showed that the spectral density of the Hüsler–Reiss exponent measure is

$$g_\vartheta(\omega_1, \ldots, \omega_m) = \frac{1}{\omega_1^2 \omega_2 \cdots \omega_m (2\pi)^{(m-1)/2} |\det \Sigma_\vartheta|^{1/2}} \exp\left(-\frac{1}{2} \tilde{\boldsymbol{\omega}}^{\mathrm{T}} \Sigma_\vartheta^{-1} \tilde{\boldsymbol{\omega}}\right),$$

$$\boldsymbol{\omega} \in S_{m-1},$$

where $\tilde{\boldsymbol{\omega}} = (\log(\omega_j/\omega_1) + \Gamma_\vartheta(t_j, t_1)/2 : j = 2, \ldots, m)^{\mathrm{T}}$ and $\Sigma_\vartheta \subset \mathbb{R}^{(m-1) \times (m-1)}$ is the covariance matrix from Remark 2.2(a) for $u = t_1$, that is,

$$(25) \qquad \Sigma_\vartheta = \frac{1}{2} \{\Gamma_\vartheta(t_i, t_1) + \Gamma_\vartheta(t_j, t_1) - \Gamma_\vartheta(t_i, t_j)\}_{2 \leq i, j \leq m}.$$

Thus, denoting the index set of extremal observations by $\mathcal{I} = \{k = 1, \ldots, n : \|\mathbf{X}_k\|_1 > u\}$, the spectral estimator $\hat{\vartheta}_{\mathrm{SPEC}}$ of ϑ is defined by

$$(26) \qquad \hat{\vartheta}_{\mathrm{SPEC}} = \arg\max_{\vartheta \in \Theta} \sum_{k \in \mathcal{I}} \log g_\vartheta(\mathbf{X}_k / \|\mathbf{X}_k\|_1).$$

The advantage of this estimator over composite likelihood counterparts is that it uses a full likelihood and thus is fully efficient, thus giving improved estimation of Brown–Resnick processes; see the simulation study in Engelke et al. (2015). Owing to the explicit form of the spectral densities, this approach is feasible even for a large number m of locations.

4.3.2. *Censored estimation of Γ_ϑ.* Conditioning on the norm of observations being large, as in (24), might introduce bias, since the limit distribution may provide a poor density approximation to any of the $\mathbf{X}_k$ that have small individual components. To overcome this, Wadsworth and Tawn (2014) apply *censoring* to those components that do not exceed a fixed high threshold. We adopt their approach, giving a new, simpler expression for the censored likelihood, valid for any process with Hüsler–Reiss margins, not just for stationary Brown–Resnick processes.

Similarly to the spectral estimation based on (24), for large thresholds $u > 0$ we have the approximation

$$(27) \qquad \mathbb{P}\left(\mathbf{X} \in d\mathbf{x}, \max_{j=1,\ldots,m} X_j > u\right) \approx -\frac{\partial^m}{\partial x_1 \cdots \partial x_m} V_{\Gamma, \mathbf{t}}(x_1, \ldots, x_m) \, d\mathbf{x}.$$

Here, a multivariate observation is said to be extreme if at least one component exceeds the threshold. For the likelihood contribution from an observation $\mathbf{X} = (X_1, \ldots, X_m)$ we distinguish two cases:

- if at least one component exceeds the threshold, that is, $X_j > u$ for all $j \in \mathcal{K}$ and $X_j \leq u$ for all $j \in \mathcal{K}^C = \{1, \ldots, m\} \setminus \mathcal{K}$ for a nonempty subset $\mathcal{K} \subset \{1, \ldots, m\}$, we compute the likelihood $f_{\vartheta, \mathcal{K}}(\mathbf{X})$ by censoring all $\mathcal{K}^C$-components of the full likelihood $f_{\vartheta, 1:m}(\mathbf{X})$. We thus only use the information that those components are below the threshold u, but not their exact values. Without loss of generality, let $\mathcal{K} = \{1, \ldots, b\}$, for some $b \in \{1, \ldots, m\}$. Then the censored likelihood is

$$
\begin{aligned}
(28) \quad f_{\vartheta, \mathcal{K}}(\mathbf{x}) &= -\frac{\partial^b}{\partial x_1 \cdots \partial x_b} V_{\Gamma, \mathbf{t}}(x_1, \ldots, x_b, u, \ldots, u) \\
&= \frac{1}{x_1^2 x_2 \cdots x_b} \phi_{b-1}(\tilde{\mathbf{x}}_{2:b}; \Sigma_{2:b, 2:b}) \Phi_{m-b}(\mu_C; \Sigma_C),
\end{aligned}
$$

where $\Sigma = \Sigma_\vartheta$ is the covariance matrix in (25), $\tilde{\mathbf{x}} = (\log x_j - \log x_1 + \Gamma_\vartheta(t_j, t_1)/2 : j = 1, \ldots, m)^\mathrm{T} \in \mathbb{R}^m$, and $\phi_p(\cdot, \Psi)$ and $\Phi_p(\cdot, \Psi)$ denote the density and the cumulative distribution function of a p-dimensional, zero-mean normal distribution with covariance matrix Ψ. We set ϕ_0 to 1 if $b = 1$, and Φ_0 to 1 if $b = m$. The conditional mean μ_C and covariance matrix Σ_C are

$$
(29) \quad \mu_C = \left(\log u - \log x_1 + \Gamma_\vartheta(t_j, t_1)/2\right)_{j=b+1, \ldots, m} - \Sigma_{(b+1):m, 2:b} \Sigma_{2:b, 2:b}^{-1} \tilde{\mathbf{x}}_{2:b},
$$

$$
(30) \quad \Sigma_C = \Sigma_{(b+1):m, (b+1):m} - \Sigma_{(b+1):m, 2:b} \Sigma_{2:b, 2:b}^{-1} \Sigma_{2:b, (b+1):m}.
$$

  In the case $b = 1$, μ_C and Σ_C are unconditional, that is, the last summands in the formulas above vanish. The derivation of this new representation of $f_{\vartheta, \mathcal{K}}$ can be found in Asadi, Davison and Engelke (2015).
- if none of the components exceeds u, that is, $\mathcal{K} = \varnothing$, then the likelihood contribution is just the probability $f_{\vartheta, \mathcal{K}}(\mathbf{x}) = 1 - V_{\Gamma, \mathbf{t}}(\mathbf{u})$ that $\mathbf{X}$ lies entirely below the threshold.

Let $\mathcal{J} = \{i = 1, \ldots, n : \max_{k=1, \ldots, m} X_{i,k} > u\}$ denote the index set of observations extreme in the sense of (27) and, for each $i \in \mathcal{J}$, let $\mathcal{K}_i$ be the index set of those components of $\mathbf{X}_i$ that exceed u. Then, the censored estimator $\hat{\vartheta}_{\mathrm{CENS}}$ is obtained by maximizing the log-likelihood [Thibaud and Opitz (2015), Section 3]

$$
(31) \quad \hat{\vartheta}_{\mathrm{CENS}} = \arg\max_{\vartheta \in \Theta} \left[(n - |\mathcal{J}|) \log\{1 - V_{\Gamma, \mathbf{t}}(\mathbf{u})\} + \sum_{i \in \mathcal{J}} \log f_{\vartheta, \mathcal{K}_i}(\mathbf{X}_i) \right].
$$

This estimator has the advantage of using full likelihoods and reducing potential bias by censoring components that might not yet have converged, but the disadvantage of being slow when m is large, since the censored likelihood $f_{\vartheta, \mathcal{K}}$ then involves the burdensome evaluation of high-dimensional normal distribution functions.

4.3.3. *Simulation study.* The two estimators $\hat{\vartheta}_{\mathrm{SPEC}}$ and $\hat{\vartheta}_{\mathrm{CENS}}$ use different data and will have different behavior for finite sample sizes. We conducted a small simulation study to assess their performance in a setting similar to our application. Details can be found in Asadi, Davison and Engelke (2015). Both estimation procedures work for the simulated data, even with a low number of observations; only the extreme events contribute to the likelihoods. In simulated data, the advantage of censoring cannot be seen, but it will reduce any bias for real data. As also noted by Engelke et al. (2015) and Einmahl et al. (2015), the estimates of λ_{Euc} have larger variation than the others. In fact, owing to a near-functional relationship between the scale λ_{Euc} and the shape α of the fractal variogram, these two parameters are strongly related in the range considered here, and this near lack of identifiability gives highly variable estimators of λ_{Euc}.

5. Extreme river discharges in the upper Danube basin.

5.1. *Data.* We used data for average daily discharges recorded at $m = 31$ German gauging stations on 20 rivers in the upper Danube basin, made available by the Bavarian Environmental Agency (http://www.gkd.bayern.de). The average discharges at these stations range from around 20 m^3/s at high altitudes to around 1400m^3/s at the most downstream station. The major part of the runoff in the basin arises from the Alps, situated south of the Danube; see Figure 1. The series at individual stations have lengths from 50 to 130 years, with 50 years of data for all stations from 1960–2009. Originally, data were provided for 47 stations, but we excluded 16 stations which have very small discharges or whose largest discharges are affected by hydroelectric installations or dampened by big lakes; it might be possible to include these data by applying special preprocessing techniques, but we have not explored this.

Exploratory analysis shows that around one-half of the annual maxima in the basin occur in June, July and August. This agrees with the study of floods in the Danube tributaries Lech and Isar by Böhm and Wetzel (2006), which shows that nearly all major floods in recent decades have occurred in these three months; floods in this area are typically caused by heavy summer rain. In order to eliminate temporal nonstationarities and the effect of snow melt, we restrict our analysis to these months. For $k = 1, \ldots, N$, we let $\mathbf{Y}_k = (Y_{1,k}, \ldots, Y_{m,k})$ denote the daily mean discharge at the m stations on day k. The number of common measurements at all stations is thus $N = 50 \times 92 = 4600$, that is, 50 years of 92 daily observations in the summer months.

Seasonality and overall trend are the main sources of nonstationarity in river flow data, but as we use only the summer month discharges, the seasonality becomes negligible. National studies have concluded that there are no significant trends in the extremes of stream flows in our area of interest [Katz, Parlange and Naveau (2002), Kundzewicz et al. (2005)], in agreement with our exploratory analysis, so henceforth we treat our data as temporally stationary.

In addition to the time series of daily average discharges, we use a digital elevation model to obtain the following geographical covariates at each station: the latitude and longitude of both the station itself and the weighted centroid of its sub-catchment, and catchment attributes including its size, mean altitude and mean slope.

5.2. *Declustering.* Extreme discharges at a given station occur in clusters due to temporal dependence, which must be removed for spatial modeling. Moreover, a large value at an upstream station may cause a peak further downstream a day or two later. These slightly shifted maximum values on different rivers stem from the same event and should be treated as dependent. In the framework of meteorology, multivariate declustering is used by Tawn (1988), Coles and Tawn (1991) and Palutikof et al. (1999) to extract independent "storm events." We apply a similar technique to obtain a set of independent flood events $\tilde{\mathbf{X}}_1, \ldots, \tilde{\mathbf{X}}_n \in \mathbb{R}^m$ on the river network from the full time series $\mathbf{Y}_k$ ($k = 1, \ldots, N$).

In order to extract the flood events, we first identify nonoverlapping windows of length p days in each of the 50 summer periods. We replace each observation by its rank within its series, and then consider the day with the highest rank across all series, choosing this day randomly if it is not unique. We then take a window of p days centered upon the chosen day, and form an event by taking the largest observation for each series within this window. We delete the data in this window and then repeat the process of forming events, stopping when no windows of p consecutive days remain. Figure 5 illustrates this declustering procedure. In agreement with Kallache et al. (2010), our data suggest that flood events last no longer than 9 days, so we put $p = 9$; a sensitivity analysis showed that our results are robust to this choice. For the ith time window, the corresponding flood event $\tilde{\mathbf{X}}_i$ is the m-dimensional vector whose jth entry is the maximum discharge value at location t_j within this window. This procedure yields a declustered time series of $n = 428$ supposedly independent events $\tilde{\mathbf{X}}_i$ from the $N = 4600$ summer measurements common to the 31 series.

5.3. *Marginal fitting.* Before using the techniques from Section 4.3 to fit the multivariate dependence model, we assess the univariate tail behavior at individual gauging stations, obtaining the constants $a_{j,n}, b_{j,n}$ and shape parameters ξ_j that allow us to normalize the margins to lie in the standard Fréchet max-domain of attraction, using (3). The model $\tilde{\eta}_\Gamma$ in (20) is a max-stable stochastic process on the whole river network T, so in order to make predictions throughout T, we must allow the norming constants and shape parameters to vary with covariates that are easily obtainable even at locations without gauging stations or find some other way to extend the model to the entire network, such as kriging.

We fitted a generalized extreme value distribution (2) to the tail of the declustered daily discharges at each gauging station location t_j, estimating the extreme
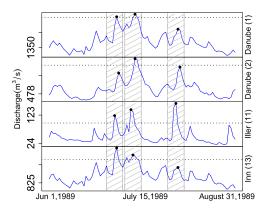
FIG. 5. *Declustered flood events at four gauging stations. The grey hatched areas are the p-day time windows around flood events. Only events for which at least one river exceeds its 90% quantile (dotted horizontal lines) are shown. The black circles show maxima for each river in each window.*

value parameters $a_{j,n}$, $b_{j,n}$ and ξ_j. At each location we tested whether the extremal behavior from any available earlier data changed relative to the 50 common years. In almost all cases there was no such change, and we could use the longer series of independent events, declustered using the procedure of Section 5.2, for each station. For the marginal fitting we use the independent events at gauging stations and estimate the GEV parameters by maximizing the joint Poisson process likelihood given in (21) in an independence likelihood [Chandler and Bate (2007)].

We fitted and compared a variety of different models using this technique, finally settling on a version of regional analysis, as widely used in hydrological applications. The idea is similar to the regionalization method of Merz and Blöschl (2005), who predict high quantiles of river flows using the catchment attributes of stations that are "hydrologically" close. Exploratory analysis suggests that for our purposes the upper Danube basin can be split into four disjoint regions: R1 contains eight stations in the southwest of the upper Danube basin and has mid-altitude sub-catchments; R2 comprises five stations in the Inn basin that are fed by precipitation in high-altitude alpine regions; R3 contains 13 stations in the center of the Danube basin that are fed by precipitation from regions with both high and low altitudes; and R4 contains five stations with sources north of the Danube. With $J_1, \ldots, J_4$ denoting the index sets of stations in regions $R_1, \ldots, R_4$, we let for $j \in J_i$ $(i = 1, \ldots, 4)$

(32)
$$\log(a_{j,n}) = \sum_{k=1}^{4} \alpha_k^{(i)} \log(P_{j,k}),$$

$$\log(b_{j,n}) = \sum_{k=1}^{4} \beta_k^{(i)} \log(P_{j,k}), \qquad \xi_j = \xi^{(i)},$$
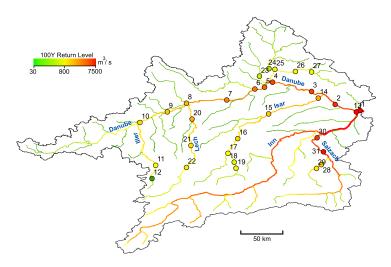
FIG. 6.   100-*year return levels for river flow* (m^3/s), *extrapolated to the entire network T* ; *the colors of the points indicate the return levels at the* 31 *numbered gauging stations.*

where $P_{j,1}, \ldots, P_{j,4}$ are the latitude of the centroid, the size, the mean altitude and the mean slope of the sub-catchment of gauging station j. Likelihood ratio statistics were used to further simplify the model, finally yielding a model with 28 parameters, compared to $93 = 3 \times 31$ parameters in the full model. Diagnostic plots indicate a very satisfactory fit of the simpler model, which is also strongly favored by the AIC. The estimated shape parameters and their standard errors for the four regions are 0.030 (0.025), 0.145 (0.034), 0.028 (0.022) and 0.294 (0.045), suggesting that catchments influenced by mountain regions tend to have heavier-tailed responses.

This model allows the extrapolation of the marginal fit to ungauged locations on the network T, thereby enabling computation of return levels throughout T; see Figure 6. More details are given in Asadi, Davison and Engelke (2015).

5.4. *Joint fitting.*   The generalized extreme value distributions constitute all possible limits for univariate maxima, but the dependence structure of multivariate extremes is infinite-dimensional, so we must first check that the extreme discharges at different stations on the river network are asymptotically dependent; if not, max-stable processes would not be suitable models. Keef, Svensson and Tawn (2009) note that the spatial dependence of extreme river flows is much stronger than that of precipitation data, since the former averages the latter and thus is less vulnerable to small-scale variation, and standard diagnostics [Coles, Heffernan and Tawn (1999)] show strong extremal dependence between all 31 stations in our data. Moreover, Figure 7 shows bivariate scatter plots of two flow-connected and two flow-unconnected stations. In both cases, the assumption of asymptotic dependence seems appropriate and, moreover, a symmetric model for the tail dependence can be justified.
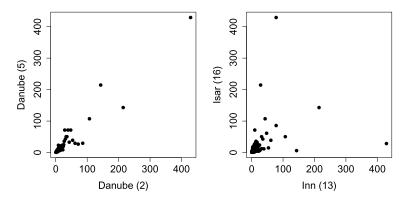
FIG. 7. *Scatter plots of declustered discharges* (*normalized to the unit Fréchet scale*) *of two flow-connected stations* (*left*) *and two flow-unconnected stations* (*right*).

The choice of a parametric subclass within the asymptotic dependence models must be a good approximation to the infinite-dimensional structure of multivariate max-stable distributions. Theorem 17 in Kabluchko, Schlather and de Haan (2009) gives some justification for the fitting of Hüsler–Reiss distributions and Brown–Resnick type processes, which are essentially the only possible limits of pointwise maxima of suitably rescaled and normalized, independent, stationary Gaussian processes.

In order to assess whether the Hüsler–Reiss distribution approximates the extremal dependence of our data well, we estimate the extremal coefficient $\hat{\theta}$ as in (8) for each pair of locations using the madogram [Cooley, Naveau and Poncet (2006)] based on summer maxima. We then fit the bivariate Hüsler–Reiss distribution (12) to these data by a censored peaks-over-threshold approach and use (13) to compute a model-based extremal coefficient estimate $\hat{\theta}_{\mathrm{HR}}$. The left panel of Figure 8 suggests that the Hüsler–Reiss model provides an excellent overall approximation to the bivariate extremal dependence structure of the discharge data, albeit with slight overestimation of dependence at longer distances for flow-unconnected pairs.

We compare four overall models for the dependence kernel Γ:

- the stationary variogram based on Euclidean distances with anisotropy matrix R as in (18),

$$\Gamma_1(s, t) = \lambda \| R \cdot (s - t) \|^{\alpha}, \qquad \lambda > 0, \alpha \in (0, 2], \beta \in [\pi/4, 3\pi/4], c > 0;$$

- a variogram using the transformation H to hydrological locations,

$$\Gamma_2(s, t) = \lambda \| R \cdot \{ H(s) - H(t) \} \|^{\alpha},$$

$$\lambda > 0, \alpha \in (0, 2], \beta \in [\pi/4, 3\pi/4], c > 0;$$

- a variogram that includes the hydrological properties of the river network for flow-connected locations, corresponding to (19),

$$\Gamma_3(s, t) = \lambda_{\mathrm{Riv}} \Gamma_{\mathrm{Riv}}(s, t) + \lambda_{\mathrm{Euc}} \| R \cdot \{ H(t) - H(s) \} \|^{\alpha},$$
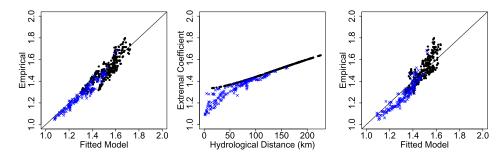
FIG. 8. *Comparison of empirical estimates of extremal coefficients found nonparametrically using the madogram and those implied by different models, for all pairs of gauging stations. Left: madogram-based estimates and extremal coefficients $\hat{\theta}_{HR}$ of the Hüsler–Reiss model, estimated by fitting to independent events. Center: estimates using $\hat{\Gamma}_3$ plotted against hydrological distance. Right: madogram-based estimates and those from fitted joint model $\hat{\Gamma}_3$. Those for flow-connected pairs are blue crosses, and those for flow-unconnected pairs are black circles.*

whose six parameters are given in (22); finally,

- we also consider the previous model without anisotropy,

$$\Gamma_4(s, t) = \lambda_{\mathrm{Riv}} \Gamma_{\mathrm{Riv}}(s, t) + \lambda_{\mathrm{Euc}} \| H(t) - H(s) \|^{\alpha},$$

$$\lambda_{\mathrm{Riv}}, \lambda_{\mathrm{Euc}} > 0, \tau > 0, \alpha \in (0, 2].$$

The weights in Γ_{Riv} are computed according to (16) using a digital elevation model.

In Section 5.2 we extracted $n = 428$ independent multivariate flood events $\tilde{\mathbf{X}}_1, \ldots, \tilde{\mathbf{X}}_n$, whose univariate extremal behavior was analyzed in Section 5.3. In order to fit the multivariate dependence structure, we use the marginal empirical distribution functions to transform the distribution at each gauging station to standard Pareto, and denote the resulting data by $\mathbf{X}_1, \ldots, \mathbf{X}_n$. We fit the functions $\Gamma_1, \ldots, \Gamma_4$ for the negative definite kernel in η_Γ to these data using the inference procedures described in Section 4.3, first obtaining the spectral estimate $\hat{\vartheta}_{\mathrm{SPEC}}$ in (26) by grid search on the parameter space Θ, and then using this as an initial value for the more demanding computation of the censored estimate $\hat{\vartheta}_{\mathrm{CENS}}$ in (31). It would be preferable to fit the univariate margins and the dependence structure simultaneously, but here this is infeasible since the optimization for the dependence structure is very time intensive.

The maximized log-likelihoods corresponding to $\Gamma_1, \ldots, \Gamma_4$ are -6629.17, -6161.86, -5907.49 and -5915.97; Γ_3 has six parameters, and the others all have four parameters. The use of hydrological distances for $\Gamma_2, \Gamma_3, \Gamma_4$ gives a huge improvement over the use of Euclidean distances in Γ_1, and adding the component Γ_{Riv} for flow-connected dependence means that Γ_3 is much better than Γ_2. The drop from Γ_3 to Γ_4 shows that the anisotropy matrix R also contributes to the good fit of the model based on Γ_3.

The center and right panels of Figure 8 (recall also the right panel of Figure 2) compare the extremal coefficients obtained with the madogram and those implied by the fitted model Γ_3. The center panel shows that the latter do not lie on a smooth curve; flow-connected pairs at the same distance can have different extremal coefficients, depending on where the two stations lie on the network, because the river dependence kernel Γ_{Riv} is nonstationary, unlike those based on simple meteorology. Overall there is a fairly good fit, though the model tends to slightly understate dependence at short hydrological distances and to overstate it at long ones.

The parameter estimates $\hat{\vartheta}_{\text{CENS}}$ are $\hat{\lambda}_{\text{Riv}} = 0.73$ (0.07), $\hat{\lambda}_{\text{Euc}} = 1.93 \times 10^{-4}$ (0.75×10^{-4}), $\hat{\tau} = 839$ (280) km, $\hat{\alpha} = 1.75$ (0.08), $\hat{\beta} = 1.10$ (0.11) and $\hat{c} = 0.64$ (0.08), with standard errors in parentheses obtained from 100 nonparametric bootstrap simulations. The high uncertainty for $\hat{\lambda}_{\text{Euc}}$ was mentioned when discussing the simulation study; it does not translate into high variation of the fitted model.

The fitted weights $\hat{\lambda}_{\text{Riv}}$ and $\hat{\lambda}_{\text{Euc}}$ cannot be compared directly, because the variogram Γ_{Euc} is unbounded and thus does not have a natural normalization. The influences of the river and the Euclidean dependence kernel on the overall extremal dependence between two flow-connected points $s, t \in T$ can be measured by $\hat{\Gamma}_{\text{Riv}}(s, t)/\hat{\Gamma}_3(s, t)$ and $\hat{\Gamma}_{\text{Euc}}(s, t)/\hat{\Gamma}_3(s, t)$, respectively. In fact, for certain pairs of stations the river dependence kernel is dominant, whereas for others the Euclidean kernel has a stronger influence on the extremal dependence. The parameter $\hat{\tau}$ is the scale for dependence along the river; as expected, this dependence is very strong, decreasing to zero only after $\hat{\tau} = 839$ km. The shape parameter $\hat{\alpha}$ describes how local the influence of spatial meteorological events on river flows is; note that $\hat{\alpha} = 1.75$ is much larger than in applications on extreme precipitation, confirming the observation of Keef, Svensson and Tawn (2009) that extreme river flows exhibit stronger spatial dependence due to an averaging effect. The parameters $\hat{\beta}$ and $\hat{c}$ describe the anisotropy of meteorological dependence, since the transformation $R(\hat{\beta}, \hat{c})$ dilates the space in direction $(\sin\hat{\beta}, \cos\hat{\beta})$ by $\hat{c}$. As $\hat{c} < 1$, extremal dependence is increased in this direction, which corresponds approximately to the planar vector $(2, 1)$. Thus, in terms of hydrological distance, two stations that are 64 km apart in a direction roughly parallel to the Alps have the same dependence as two stations that are 100 km apart perpendicular to the Alps. In view of the orientations of the catchments and the blocking effect that the Alps have on weather systems, this seems quite plausible.

5.5. *Higher-order properties.* Figure 8 shows how the max-stable model η_{Γ_3} fits the bivariate extremal features of the data. In practice, higher-order properties such as multivariate exceedance probabilities are also of interest, and to check these we randomly choose groups of 3, 10, 15 and 31 stations and compute the quantiles of their observed group maxima, suitably rescaled [cf. Davison and Gholamrezaee (2012)]. Figure 9, which compares these quantiles with the theoretical
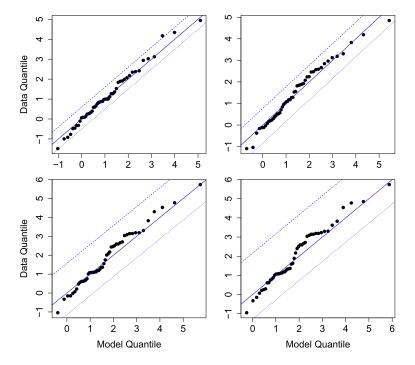
FIG. 9.   *QQ-plots* (*Gumbel scale*) *of observed groupwise yearly maxima and theoretical values from the fitted model, for groups of* 3 (*top left*), 5 (*top right*), 15 (*bottom left*) *and all* 31 (*bottom right*) *stations. Dashed lines and dotted lines correspond to values for complete independence and complete dependence, respectively, and the solid line corresponds to the fitted model.*

values derived from the fitted model, shows that the model captures even high order structures of the data very well. Moreover, the comparison of observed quantiles to those corresponding to complete independence and complete dependence underlines the importance of proper dependence modeling.

A joint extremal model allows the estimation of the risk of simultaneous exceedances of high thresholds at multiple locations. More precisely, we can use equation (7) to approximate these probabilities as a function of the univariate extreme value parameters and the exponent measure V of the dependence model. For three stations $\mathbf{t} = (t_1, t_2, t_3) \in T^3$, the exponent measure for our model is $V_{\Gamma, \mathbf{t}}$ as in (14). Let $q_{j,p}$ be the p-quantile of the distribution of daily discharges at station t_j. The probability of a flood that exceeds the respective p-quantiles at all three stations in the same summer can be approximated by

$$
(33) \quad
\begin{aligned}
&K\mathbb{P}\big(X(t_j) > q_{j,p}; j = 1, 2, 3\big) \\
&\approx V_{\hat{\Gamma}_3, \mathbf{t}}\left\{\prod_{j=1}^{3}\left(\left(1 + \hat{\xi}_j \frac{q_{j,p} - \hat{b}_{j,n}}{\hat{a}_{j,n}}\right)_{+}^{1/\hat{\xi}_j}, \infty\right)\right\},
\end{aligned}
$$

where K is the mean number of multivariate events per year. The estimates for the shape and scale parameters are taken from the fitted covariate model in (32), so this multivariate exceedance probability, and others for more complex events, can be computed for any locations, even ungauged, on the river network. To compare the model with empirical data, we randomly choose 500 out of the $\binom{31}{3}$ possible triplets of gauging stations and evaluate (33) for different values of p close to 1. The mean relative absolute differences of these model probabilities and their empirical counterparts are 15% for $p = 0.95$, 14% for $p = 0.97$, 19% for $p = 0.99$, and 31% for $p = 0.995$; the empirical counterparts are highly variable, since they are based on very few events.

6. Discussion. The approach described above was used to fit other max-stable processes, such as the extremal-t or Schlather models, but we found that the Brown–Resnick model was the best of those fitted; perhaps this is not surprising, since this model is flexible and allows independent extremes at long distances, unlike the Schlather model, for example.

Keef, Svensson and Tawn (2009), Keef, Tawn and Lamb (2013), Keef, Tawn and Svensson (2009) describe an alternative approach to modeling joint flooding that allows the possibility of asymptotically independent extremes through the fitting of the Heffernan and Tawn (2004) model. This can handle large-scale problems, but has the drawback of not treating the variables symmetrically, and it is not clear whether it corresponds to a well-defined joint model. In those papers, it is important to allow for asymptotic independence because the data arise from rivers that may be quite unrelated, whereas stronger dependence might be anticipated in a single river network, as in the present paper. Moreover, our approach uses the known structure of the river networks, which should provide better dependence modeling.

Finally, the ideas suggested here might be extended to similar problems for which Euclidean geometry does not seem natural, such as the transmission of earthquake shocks along fault lines, or communication networks, though it would then be important to allow for flows in different directions. In some applications it might be useful to include the relative timings of extremes at different nodes of the network.

SUPPLEMENTARY MATERIAL

Supplement to "Extremes on river networks" (DOI: 10.1214/ 15-AOAS863SUPP; .zip). The supplementary material contains the following: a PDF document containing the derivation of the new likelihood representation mentioned in Section 4.3.2, results of the simulation study mentioned in Section 4.3.3, and additional details germane to Section 5.3; and R code and data files to reproduce the data analysis and figures.

REFERENCES

ASADI, P., DAVISON, A. C. and ENGELKE, S. (2015). Supplement to "Extremes on river networks."
    DOI:10.1214/15-AOAS863SUPP.

BIENVENÜE, A. and ROBERT, C. (2014). Likelihood based inference for high-dimensional extreme
    value distributions. Available at http://arxiv.org/abs/1403.0065.

BLANCHET, J. and DAVISON, A. C. (2011). Spatial modeling of extreme snow depth. *Ann. Appl.
    Stat.* **5** 1699–1725. MR2884920

BÖHM, O. and WETZEL, K.-F. (2006). Flood history of the Danube tributaries Lech and Isar in the
    Alpine foreland of Germany. *Hydrological Sciences Journal* **51** 784–798.

BROWN, B. M. and RESNICK, S. I. (1977). Extreme values of independent stochastic processes.
    J. Appl. Probab. **14** 732–739. MR0517438

CHANDLER, R. E. and BATE, S. (2007). Inference for clustered data using the independence log-
    likelihood. *Biometrika* **94** 167–183. MR2367830

COLES, S. (2001). *An Introduction to Statistical Modeling of Extreme Values.* Springer, London.
    MR1932132

COLES, S., HEFFERNAN, J. and TAWN, J. (1999). Dependence measures for extreme value analyses.
    Extremes **2** 339–365.

COLES, S. G. and TAWN, J. A. (1991). Modelling extreme multivariate events. *J. R. Stat. Soc. Ser.
    B. Stat. Methodol.* **53** 377–392. MR1108334

COOLEY, D., NAVEAU, P. and PONCET, P. (2006). Variograms for spatial max-stable random fields.
    In *Dependence in Probability and Statistics* (P. Bertail, P. Soulier and P. Doukhan, eds.). *Lecture
    Notes in Statist.* **187** 373–390. Springer, New York. MR2283264

CRESSIE, N., FREY, J., HARCH, B. and SMITH, M. (2006). Spatial prediction on a river network.
    J. Agric. Biol. Environ. Stat. **11** 127–150.

DAVISON, A. C. and GHOLAMREZAEE, M. M. (2012). Geostatistics of extremes. *Proc. R. Soc.
    Lond. Ser. A Math. Phys. Eng. Sci.* **468** 581–608. MR2874052

DE HAAN, L. and FERREIRA, A. (2006). *Extreme Value Theory: An Introduction.* Springer, New
    York. MR2234156

DIEKER, A. B. and MIKOSCH, T. (2015). Exact simulation of Brown–Resnick random fields at a
    finite number of locations. *Extremes* **18** 301–314. MR3351818

DOMBRY, C., ENGELKE, S. and OESTING, M. (2016). Exact simulation of max-stable processes.
    Biometrika **103**. To appear.

EINMAHL, J., KIRILIOUK, A., KRAJINA, A. and SEGERS, J. (2015). An M-estimator of spatial tail
    dependence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77**. To appear.

EMBRECHTS, P., KLÜPPELBERG, C. and MIKOSCH, T. (1997). *Modelling Extremal Events: For In-
    surance and Finance. Applications of Mathematics* (*New York*) **33**. Springer, Berlin. MR1458613

ENGELKE, S., KABLUCHKO, Z. and SCHLATHER, M. (2011). An equivalent representation of the
    Brown–Resnick process. *Statist. Probab. Lett.* **81** 1150–1154. MR2803757

ENGELKE, S., MALINOWSKI, A., KABLUCHKO, Z. and SCHLATHER, M. (2015). Estimation of
    Hüsler–Reiss distributions and Brown–Resnick processes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*
    77 239–265. MR3299407

HEFFERNAN, J. E. and TAWN, J. A. (2004). A conditional approach for multivariate extreme values.
    J. R. Stat. Soc. Ser. B. Stat. Methodol. **66** 497–546. MR2088289

HUSER, R. and DAVISON, A. C. (2014). Space–time modelling of extreme events. *J. R. Stat. Soc.
    Ser. B. Stat. Methodol.* **76** 439–461. MR3164873

HUSER, R., DAVISON, A. C. and GENTON, M. G. (2014). A comparative study of parametric
    estimators for multivariate extremes. *Extremes.* Under review.

HÜSLER, J. and REISS, R.-D. (1989). Maxima of normal random vectors: Between independence
    and complete dependence. *Statist. Probab. Lett.* **7** 283–286. MR0980699

KABLUCHKO, Z. (2011). Extremes of independent Gaussian processes. *Extremes* **14** 285–310. MR2824498

KABLUCHKO, Z., SCHLATHER, M. and DE HAAN, L. (2009). Stationary max-stable fields associated to negative definite functions. *Ann. Probab.* **37** 2042–2065. MR2561440

KALLACHE, M., RUST, H. W., LANGE, H. and KROPP, J. P. (2010). Extreme value analysis considering trends: Application to discharge data of the Danube river basin. In *Extremis*: *Disruptive Events and Trends in Climate and Hydrology* (J. Kropp and H. Schellnhuber, eds.) 167–184. Springer, Berlin.

KATZ, R. W., PARLANGE, M. B. and NAVEAU, P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources* **25** 1287–1304.

KEEF, C., SVENSSON, C. and TAWN, J. A. (2009). Spatial dependence in extreme river flows and precipitation for Great Britain. *Journal of Hydrology* **378** 240–252.

KEEF, C., TAWN, J. A. and LAMB, R. (2013). Estimating the probability of widespread flood events. *Environmetrics* **24** 13–21. MR3042270

KEEF, C., TAWN, J. and SVENSSON, C. (2009). Spatial risk assessment for extreme river flows. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **58** 601–618. MR2750258

KUNDZEWICZ, Z. W., ULBRICH, U., BRÜCHER, T., GRACZYK, D., KRÜGER, A., LECKEBUSCH, G. C., MENZEL, L., PIŃSKWAR, I., RADZIEJEWSKI, M. and SZWED, M. (2005). Summer floods in central Europe–Climate change track? *Natural Hazards* **36** 165–189.

MERZ, R. and BLÖSCHL, G. (2005). Flood frequency regionalisation—Spatial proximity vs. catchment attributes. *Journal of Hydrology* **302** 283–306.

OESTING, M., KABLUCHKO, Z. and SCHLATHER, M. (2012). Simulation of Brown–Resnick processes. *Extremes* **15** 89–107. MR2891311

OPITZ, T. (2013). Extremal *t* processes: Elliptical domain of attraction and a spectral representation. *J. Multivariate Anal.* **122** 409–413. MR3189331

PADOAN, S. A., RIBATET, M. and SISSON, S. A. (2010). Likelihood-based inference for max-stable processes. *J. Amer. Statist. Assoc.* **105** 263–277. MR2757202

PALUTIKOF, J. P., BRABSON, B. B., LISTER, D. H. and ADCOCK, S. T. (1999). A review of methods to calculate extreme wind speeds. *Meteorol. Appl.* **6** 119–132.

RENARD, B. and LANG, M. (2007). Use of a Gaussian copula for multivariate extreme value analysis: Some case studies in hydrology. *Advances in Water Resources* **30** 897–912.

RESNICK, S. I. (1987). *Extreme Values*, *Regular Variation*, *and Point Processes*. *Applied Probability. A Series of the Applied Probability Trust* **4**. Springer, New York. MR0900810

ROOTZÉN, H. and TAJVIDI, N. (2006). Multivariate generalized Pareto distributions. *Bernoulli* **12** 917–930. MR2265668

SALVADORI, G. and DE MICHELE, C. (2010). Multivariate multiparameter extreme value models and return periods: A copula approach. *Water Resources Research* **46** W10501.

SCHLATHER, M. (2002). Models for stationary max-stable random fields. *Extremes* **5** 33–44. MR1947786

SCHLATHER, M. and TAWN, J. A. (2003). A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika* **90** 139–156. MR1966556

SKØIEN, J., MERZ, R. and BLÖSCHL, G. (2006). Top-kriging-geostatistics on stream networks. *Hydrol. Earth Syst. Sci.* **10** 277–287.

TAWN, J. A. (1988). An extreme-value theory model for dependent observations. *Journal of Hydrology* **101** 227–250.

THIBAUD, E. and OPITZ, T. (2015). Efficient inference and simulation for elliptical Pareto processes. *Biometrika* **102** 855–870.

VER HOEF, J. M. and PETERSON, E. E. (2010). A moving average approach for spatial statistical models of stream networks. *J. Amer. Statist. Assoc.* **105** 6–18. MR2757185

VER HOEF, J. M., PETERSON, E. and THEOBALD, D. (2006). Spatial statistical models that use flow and stream distance. *Environ. Ecol. Stat.* **13** 449–464. MR2297373

WADSWORTH, J. L. and TAWN, J. A. (2014). Efficient inference for spatial extreme value processes associated to log-Gaussian random functions. *Biometrika* **101** 1–15. MR3180654

WANG, Y. and STOEV, S. A. (2010). On the structure and representations of max-stable processes. *Adv. in Appl. Probab.* **42** 855–877. MR2779562

P. ASADI
FACULTÉ DES HAUTES ETUDES COMMERCIALES
UNIVERSITÉ DE LAUSANNE
EXTRANEF, UNIL-DORIGNY
1015 LAUSANNE
SWITZERLAND
E-MAIL: peiman.asadi@unil.ch

A. C. DAVISON
S. ENGELKE
ECOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
EPFL-FSB-MATHAA-STAT
STATION 8
1015 LAUSANNE
SWITZERLAND
E-MAIL: anthony.davison@epfl.ch
        sebastian.engelke@epfl.ch