Causal mechanism of extremes on a river network

Valérie Chavez-Demoulin[1]
Joint work with Linda Mhalla and Debbie Dupuis

[1]Faculty of Business and Economics,
University of Lausanne, Switzerland

RESIM, May 2021

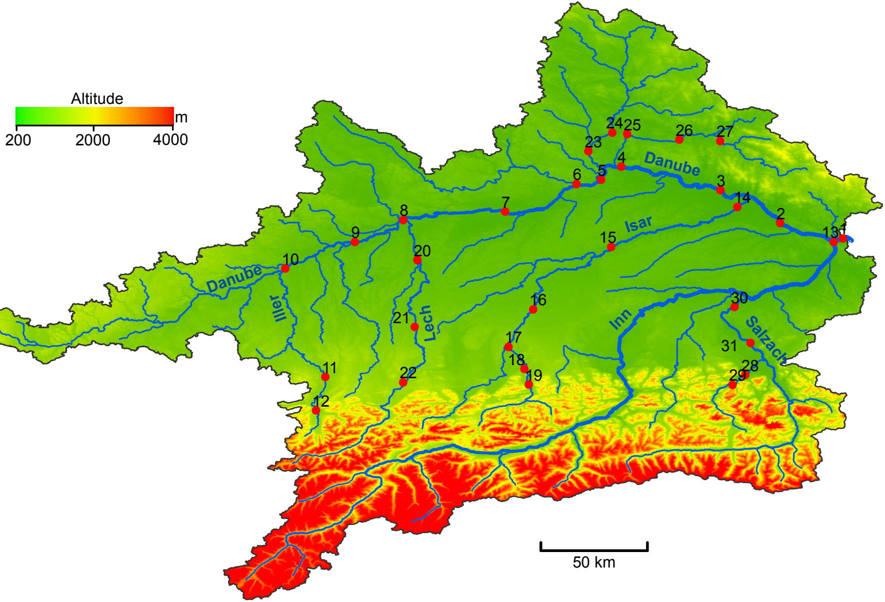
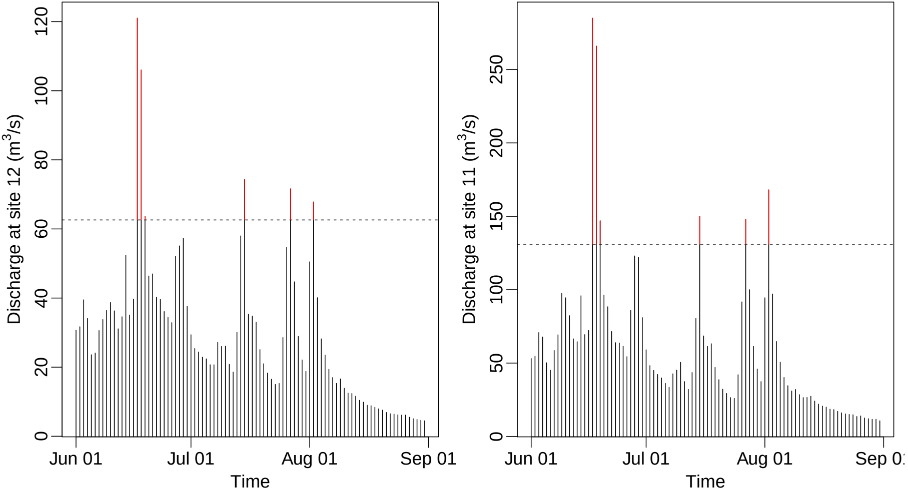
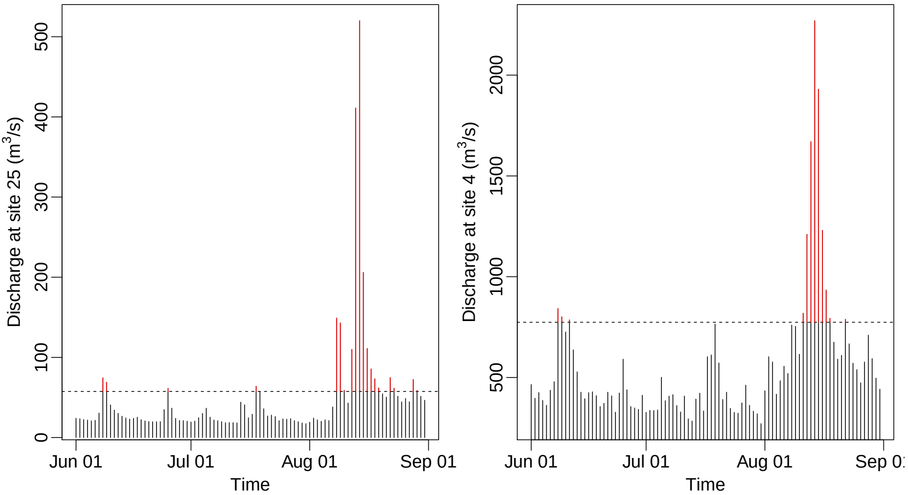


Figure from Asadi, P., Davison, A.C. and Engelke, S. (2015). Extremes on river networks. *The Annals of Applied Statistics*,

Extremes of upper Danube basin: sites 12 and 11



Extremes of upper Danube basin: sites 25 and 4



Aim

- We develop a new method to understand intrinsic causal mechanisms between extreme values of sites of a network
- Our causality approach is unconditional on time and on the real network flow structure
- We focus on the assessment of the causal relations between the extreme discharges in the upper Danube basin

Univariate EVT I

Let $(Y_i)_{i\geq 1}$ be a sequence of independent and identically distributed (iid) random variables with common distribution F. Let M_n be the maximum of a sequence of n such random variables, i.e. $M_n = \max\{Y_1, \ldots, Y_n\}$. The Fisher–Tippett theorem states that if there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that the normalized random variable M_n converges in distribution to a random variable with a non–degenerate distribution function G, i.e.,

$$\Pr\{(M_n - b_n)/a_n \leq y\} \to G(y), \quad n \to \infty, \tag{1}$$

then G belongs to the Generalized Extreme Value (GEV) family of distributions

$$\mathrm{GEV}_{(\mu,\sigma,\xi)}(y) = \begin{cases} \exp\left[-\{1+\xi(y-\mu)/\sigma\}_+^{-1/\xi}\right], & \xi \neq 0, \\ \exp\left[-\exp\{-(y-\mu)/\sigma\}\right], & \xi = 0, \end{cases}$$

defined on $\{y : 1 + \xi(y-\mu)/\sigma > 0\}$, with $-\infty < \mu, \xi < \infty$, $\sigma > 0$, and $x_+ = \max(x, 0)$.

Univariate EVT II

The following result allows the approximation of the conditional distribution of the exceedances above a high threshold. If there exist normalizing sequences $\{a_n > 0\}$ and $\{b_n\}$ such that (1) holds, i.e., F is in the max-domain of attraction of a $\text{GEV}_{(\mu,\sigma,\xi)}$, then, for a sufficiently high threshold u we can model the limiting distribution of the exceedances $Y - u | Y > u$ with a Generalized Pareto Distribution (GPD) $G_{(\tilde{\sigma}_u,\xi)}$,

$$\Pr\left\{\left(Y_n - b_n\right)/a_n > u + y \middle| \left(Y_n - b_n\right)/a_n > u\right\} \xrightarrow[n\to\infty]{} \begin{cases} (1 + \xi y/\tilde{\sigma}_u)_+^{-1/\xi}, \\ \exp(-y/\tilde{\sigma}_u), \end{cases}$$

where $\tilde{\sigma}_u = \sigma + \xi(u - \mu)$ and the shape parameter ξ equals that of the corresponding GEV distribution.

Multivariate EVT

Let $(\mathrm{Y}_i)_{i \geq 1}$ be iid copies of a d-dimensional random vector $\mathrm{Y} = (Y_1, \ldots, Y_d)$. When the marginal distributions F_j are unit Fréchet the law of the standardized componentwise maxima $n^{-1}\mathrm{M}_n$ converges in distribution to a multivariate extreme value distribution (MEVD), $G(\mathrm{z}) = \exp\{-V(\mathrm{z})\}$, with

$$V(\mathrm{z}) = \int_{S_d} \max\left(\frac{w_1}{z_1}, \ldots, \frac{w_d}{z_d}\right) dH(\mathrm{w}), \quad \mathrm{z} \in (0, \infty)^d, \int_{S_d} w_j \, dH(\mathrm{w}) = 1, \quad j$$

From the max-stability of the limiting distribution G, its associated copula

$$\begin{aligned} C^{EV}(\mathrm{v}) &= \exp\left[-V\{-1/\log(\mathrm{v})\}\right], \quad \mathrm{v} = (v_1, \ldots, v_d) \in [0,1]^d, \\ &= \exp\left\{\left(\sum_{j=1}^d \log v_j\right) A\left(\frac{\log v_1}{\sum_{j=1}^d \log v_j}, \ldots, \frac{\log v_d}{\sum_{j=1}^d \log v_j}\right)\right\} \end{aligned}$$

belongs to the large class of extreme value copulas $\mathcal{C}^{EV}$

The function A is the Pickands dependence function and is a continuous convex function defined on the unit simplex

From inference for extremes to causal discovery I

- Assume that $\mathrm{Y} = (Y_1, Y_2)$ is in the MDA of an MEVD
- Approximate for $\mathrm{Y} \geq \mathrm{u}$ (with inequality holding componentwise), the dependence structure between the exceedances of the high bivariate threshold u by an extreme value copula
- We make no assumption on the form of the extremal dependence in the random vector $\mathrm{Y} = (Y_1, Y_2)$

From inference for extremes to causal discovery II

The derivatives of the extreme value copula and the Pickands function are related:

$$\begin{aligned}
\partial_{v_1} C^{EV}(v_1, v_2) &= C(v_1, v_2)\left\{\frac{1}{v_1} A\left(\frac{\log v_1}{\log v_1 + \log v_2}\right) + \frac{1}{v_1}\frac{\log v_2}{\log v_1 + \log v_2} A'\left(\frac{\log v_1}{\log v_1 + \log v_2}\right)\right\}, \\
\partial_{v_2} C^{EV}(v_1, v_2) &= C(v_1, v_2)\left\{\frac{1}{v_2} A\left(\frac{\log v_1}{\log v_1 + \log v_2}\right) - \frac{1}{v_2}\frac{\log v_1}{\log v_1 + \log v_2} A'\left(\frac{\log v_1}{\log v_1 + \log v_2}\right)\right\},
\end{aligned}$$

where $A(\omega) \equiv A(\omega, 1-\omega)$ and $A'(\omega) = dA(\omega)/d\omega$

- We take a non-parametric inference approach using the min-projection approach of Mhalla et al. (2019)
- The resulting estimates of the extreme value copula as well as its derivatives are valid

Pairwise causal discovery of extremes

- $\{(X_i, Y_i)\}_{i=1}^{n}$ dataset
- $\{(X_i^{ext}, Y_i^{ext})\}_{i=1}^{n_u}$ extremes exceeding sufficiently high threshold in both margins
- Aim: study the causal relationships between X^{ext} and Y^{ext}

Postulate

The mechanism generating the random variable describing the cause, denoted by X, and the mechanism generating the random variable describing the effect given the cause, denoted by $Y|X$, are two independent mechanisms, i.e., they contain no information about each other

Kolmogorov complexity

Postulate

If the random variable X is the cause of a random variable Y, then the distribution of X, F_X, and the distribution of $Y|X$, $F_{Y|X}$, are algorithmically independent, that is

$$K(F_{X,Y}) = K(F_X) + K(F_{Y|X})$$

where $K(F)$ stands for the Kolmogorov complexity of a distribution F

Note that in a setting where X and Y are causally related, i.e., either X causes Y or Y causes X, the postulate implies that one should infer that X causally influences Y whenever

$$K(F_X) + K(F_{Y|X}) \leq K(F_Y) + K(F_{X|Y})$$

Minimum description length

- Kolmogorov complexity is not computable
- We replace it by the MDL which translates the Occam's razor and focuses on the description length of probability distributions
- Based on the MDL principle, the Kolmogorov complexity of a random variable can be practically approximated by its code length leading to the following inequality whenever X is the cause of a random variable Y:

$$CL_{\mathcal{M}_X}(X) + CL_{\mathcal{M}_{Y|X}}(Y|X) \leq CL_{\mathcal{M}_Y}(Y) + CL_{\mathcal{M}_{X|Y}}(X|Y)$$

where $CL_{\mathcal{M}_Z}(Z)$ denotes the code length of a random variable Z under the model class $\mathcal{M}_Z$

Causal discovery using quantile scoring

We derive the code lengths of the marginal and conditional random variables in the dataset $\{(X_i^{ext}, Y_i^{ext})\}_{i=1}^{n_u}$ using a quantile-based MDL

- Marginal model for the τ-th quantiles of X^{ext} and Y^{ext}

$$\begin{aligned} Q_{X^{ext}}(\tau) &= F^{-1}(\tau), \quad \text{where } F = F_{(\sigma_X,\xi_X)} \sim \text{GPD}(u_X, \sigma_X, \xi_X), \\ Q_{Y^{ext}}(\tau) &= G^{-1}(\tau), \quad \text{where } G = G_{(\sigma_Y,\xi_Y)} \sim \text{GPD}(u_Y, \sigma_Y, \xi_Y), \end{aligned}$$

- Conditional τ-th quantiles of $X^{ext}|Y^{ext}$ and $Y^{ext}|X^{ext}$

$$\begin{aligned} Q_{X^{ext}|Y^{ext}=y>u_Y}(\tau) &= F^{-1}\{(\partial_v C^{EV})^{-1}(\tau, G(y))\}, \\ Q_{Y^{ext}|X^{ext}=x>u_X}(\tau) &= G^{-1}\{(\partial_u C^{EV})^{-1}(F(x), \tau)\}, \end{aligned}$$

The level τ of the considered quantile ranges from a probability of 0 to a probability of 1 allowing for a coverage of the entire joint tail distribution

Code length decomposition

- Classes of models: $\mathcal{M}_{X^{ext}}$, $\mathcal{M}_{Y^{ext}}$, $\mathcal{M}_{X^{ext}|Y^{ext}}$, and $\mathcal{M}_{Y^{ext}|X^{ext}}$
- Models from these classes $\mathcal{F}_{X^{ext}} \in \mathcal{M}_{X^{ext}}$, $\mathcal{F}_{Y^{ext}} \in \mathcal{M}_{Y^{ext}}$, $\mathcal{F}_{X^{ext}|Y^{ext}} \in \mathcal{M}_{X^{ext}|Y^{ext}}$, and $\mathcal{F}_{Y^{ext}|X^{ext}} \in \mathcal{M}_{Y^{ext}|X^{ext}}$

If $CL^{\tau}_{\mathcal{F}_X}(X^{ext})$ denotes the code length of the random variable X^{ext} under the model $\mathcal{F}_X \in \mathcal{M}_X$ at a given τ-th quantile, then we have the following decomposition

$$CL^{\tau}_{\mathcal{F}_X}(X^{ext}) = CL^{\tau}_{\mathcal{F}_X}(\hat{\mathcal{F}}_X) + CL^{\tau}_{\mathcal{F}_X}(\hat{\epsilon}_X|\hat{\mathcal{F}}_X),$$

where $CL^{\tau}_{\mathcal{F}_X}(\hat{\mathcal{F}}_X)$ is the code length of the fitted model $\hat{\mathcal{F}}_X$ at the τ-th quantile and $CL^{\tau}_{\mathcal{F}_X}(\hat{\epsilon}_X|\hat{\mathcal{F}}_X)$ is the code length of the residuals conditional on the fitted model $\hat{\mathcal{F}}_X$ at the τ-th quantile

Code length for the fitted model

- From Rissanen (1989), we have that

$$CL^{\tau}_{\mathcal{F}_X}(\hat{\mathcal{F}}_X) = CL^{\tau}_{\mathcal{F}_Y}(\hat{\mathcal{F}}_Y) = \frac{p}{2}\log_2(n_u),$$

  where $p = 2$ is the number of parameters of the GPD

- We also have

$$CL^{\tau}_{\mathcal{F}_{X|Y}}(\hat{\mathcal{F}}_{X|Y}) = CL^{\tau}_{\mathcal{F}_{Y|X}}(\hat{\mathcal{F}}_{Y|X}),$$

  as the conditional models rely on the same marginal distributions and non-parametric extreme value copula. The analytical expression of their complexity is not needed in our enquiry about causal relationships between the joint extremes X^{ext} and Y^{ext}

Code length for the model residuals

- From Rissanen (1989) the code length of the unexplained portion of the model is given by the negative log-likelihood of the fitted model
- From Geraci and Bottai (2007) and Aue et al. (2014), the code lengths of the innovations in our τ-th quantile models are obtained through an application of the asymmetric Laplace (AL) likelihood, so that

$$
\begin{aligned}
L(\sigma_X, \xi_X) &= \tau^{n_u}(1-\tau)^{n_u} \exp\{-\hat{S}_{X^{ext}}(\tau)\} \\
L(\sigma_Y, \xi_Y) &= \tau^{n_u}(1-\tau)^{n_u} \exp\{-\hat{S}_{Y^{ext}}(\tau)\} \\
L(\sigma_X, \sigma_Y, \xi_X, \xi_Y, C^{EV}) &= \tau^{n_u}(1-\tau)^{n_u} \exp\{-\hat{S}_{X^{ext}|Y^{ext}}(\tau)\} \\
L(\sigma_X, \sigma_Y, \xi_X, \xi_Y, C^{EV}) &= \tau^{n_u}(1-\tau)^{n_u} \exp\{-\hat{S}_{Y^{ext}|X^{ext}}(\tau)\}
\end{aligned}
$$

$\hat{S}_{X^{ext}}(\tau)$, $\hat{S}_{Y^{ext}}(\tau)$, $\hat{S}_{X^{ext}|Y^{ext}}(\tau)$, and $\hat{S}_{Y^{ext}|X^{ext}}(\tau)$ are the estimates of the expected quantile scores (multiplied by n_u) of the τ-th quantile forecasts

Code length for the random variables

Summing both parts of the MDL, the code lengths of our random variables are:

$$
\begin{aligned}
CL^{\tau}_{\mathcal{F}_X}(X^{ext}) &= \log_2(n_u) + \hat{S}_{X^{ext}}(\tau) - n_u \log\{\tau(1-\tau)\} \\
CL^{\tau}_{\mathcal{F}_Y}(Y^{ext}) &= \log_2(n_u) + \hat{S}_{Y^{ext}}(\tau) - n_u \log\{\tau(1-\tau)\} \\
CL^{\tau}_{\mathcal{F}_{X|Y}}(X^{ext}|Y^{ext}) &= CL^{\tau}_{\mathcal{F}_{X|Y}}(\hat{\mathcal{F}}_{X|Y}) + \hat{S}_{X^{ext}|Y^{ext}}(\tau) - n_u \log\{\tau(1-\tau)\} \\
CL^{\tau}_{\mathcal{F}_{Y|X}}(Y^{ext}|X^{ext}) &= CL^{\tau}_{\mathcal{F}_{Y|X}}(\hat{\mathcal{F}}_{Y|X}) + \hat{S}_{Y^{ext}|X^{ext}}(\tau) - n_u \log\{\tau(1-\tau)\}
\end{aligned}
$$

Causal discovery using quantile scoring II

We can formulate the causality Postulate in terms of the τ-th quantile scores through the equivalence between the inequalities

$$\begin{aligned} CL(X^{ext}) + CL(Y^{ext}|X^{ext}) &\leq CL(Y^{ext}) + CL(X^{ext}|Y^{ext}) \\ &\Leftrightarrow \\ \hat{S}_{X^{ext}}(\tau) + \hat{S}_{Y^{ext}|X^{ext}}(\tau) &\leq \hat{S}_{Y^{ext}}(\tau) + \hat{S}_{X^{ext}|Y^{ext}}(\tau) \end{aligned}$$

Defining $\hat{S}_{X^{ext}} = \int_0^1 \hat{S}_{X^{ext}}(\tau)d\tau$ we modify our decision rule to

$$\hat{S}_{X^{ext}} + \hat{S}_{Y^{ext}|X^{ext}} \leq \hat{S}_{Y^{ext}} + \hat{S}_{X^{ext}|Y^{ext}}$$

Further, we conclude that X causes Y at extreme levels whenever

$$S^{ext}_{X \to Y} = \frac{\hat{S}_{Y^{ext}} + \hat{S}_{X^{ext}|Y^{ext}}}{\hat{S}_{X^{ext}} + \hat{S}_{Y^{ext}|X^{ext}} + \hat{S}_{Y^{ext}} + \hat{S}_{X^{ext}|Y^{ext}}} \geq 0.5$$

Simulation study

$$Y = h(X) + \epsilon, \quad X, \epsilon : \text{independent}$$

Scenario 1. $X \sim \text{GPD}(2, 0.3, 0.1)$

(a) $h(x) = \log(x + 10) + x^6$ and $\epsilon \sim t(\nu_\epsilon)$, with $\nu_\epsilon \in [2.1, 4]$

(b) $h(x) = x^3 + x$ and $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$, with $\sigma_\epsilon \in [0.1, 20]$

Scenario 2. $X \sim \mathcal{N}(1, 0.4^2)$

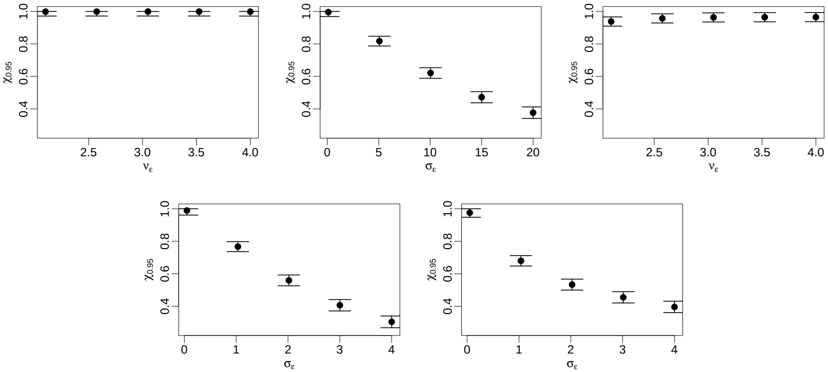
(c) $h(x) = \log(x + 10) + x^6$ and $\epsilon \sim t(\nu_\epsilon)$, with $\nu_\epsilon \in [2.1, 4]$

(d) $h(x) = x^3 + x$ and $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$, with $\sigma_\epsilon \in [0.05, 4]$

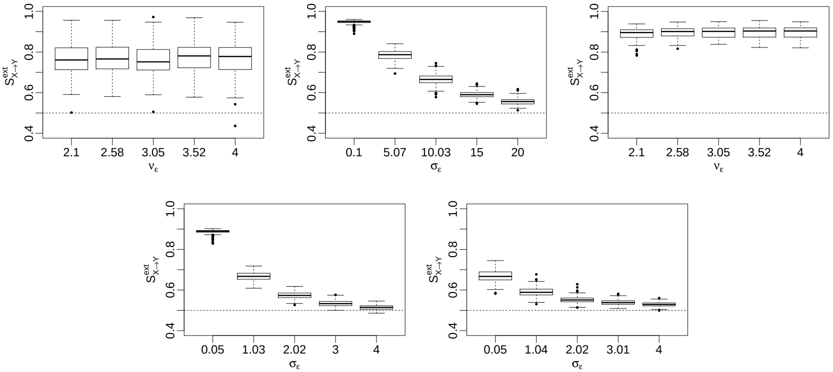
Scenario 3. $X \sim \text{GEV}(-2.8, 1, -0.1)$

(e) $h(x) = x^3 + x$ and $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$, with $\sigma_\epsilon \in [0.05, 4]$

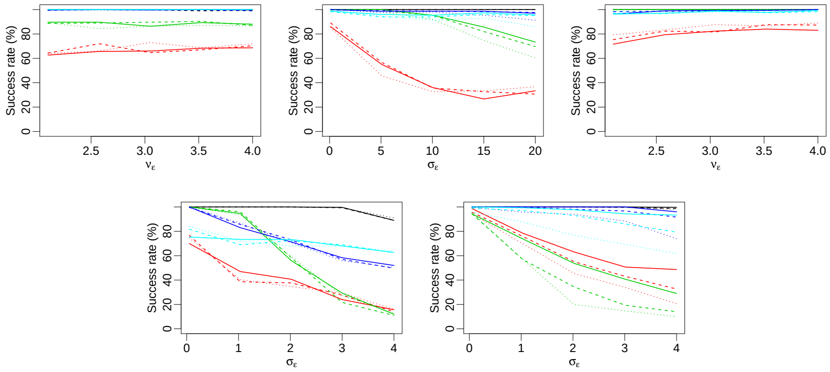
Empirical estimates of $\chi_{0.95}$



Boxplots of the estimated score $S^{ext}_{X\to Y}$

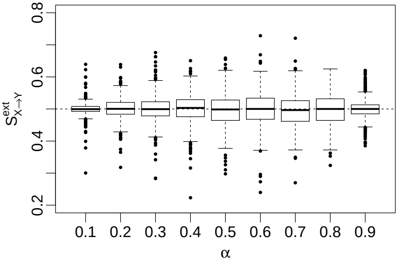


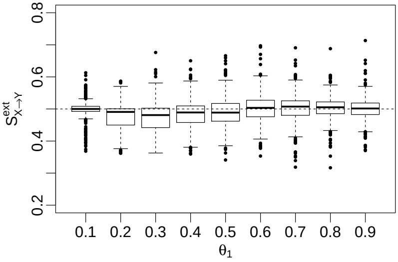
Comparison based on success rates

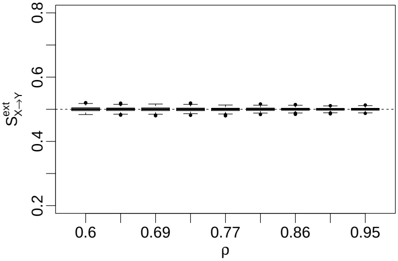


CausEV (black), LINGAM (green), IGCI (light blue), CAM (blue), and RESIT (red)

Robustness to the assumption of causality







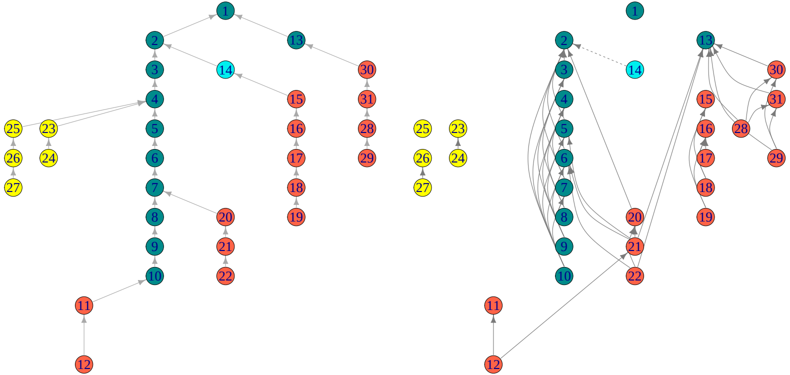
Extreme causal mechanism of the upper Danube basin I

Temporal declustering (Asadi *et al.* 2016)

- less temporally dependent extremes
- aim: unveil the inherent mechanism of causality
- n = 428 independent events at all 31 stations
- work with pairs of stations (465)

We estimate the extreme value copula using the min-projection of Mhalla et al. (2019)
CausEV for joint exceedances of 90% marginal quantiles
An edge is significant if $0.5 \notin 95\%$ boostrap bounds of the causal score $S^{ext}_{X \to Y}$

Extreme causal mechanism of the upper Danube basin II



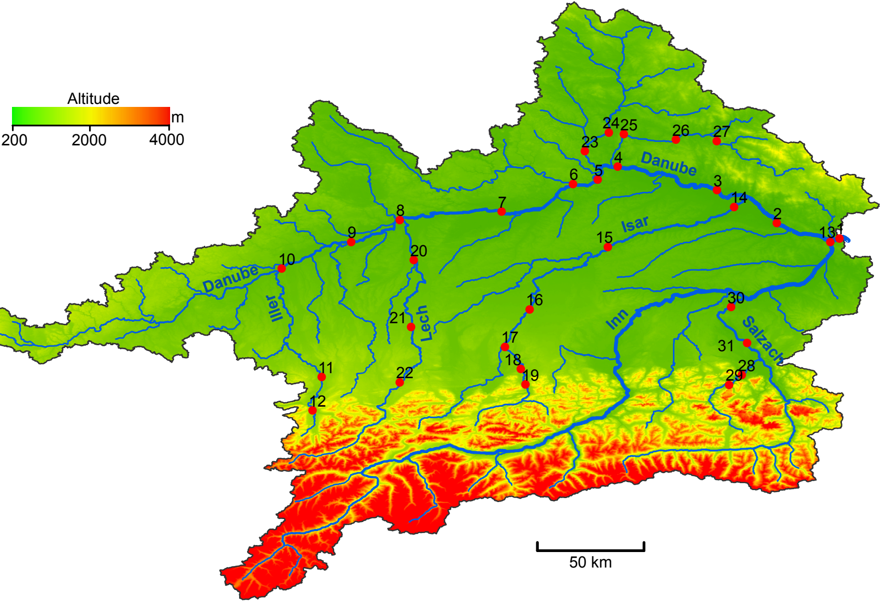


Figure from Asadi, P., Davison, A.C. and Engelke, S. (2015). Extremes on river networks. *The Annals of Applied Statistics*,

References

Thank you!

Aue, A., Cheung, R. C. Y., Lee, T. C. M. and Zhong, M. (2014) Segmented Model Selection in Quantile Regression Using the Minimum Description Length Principle. *Journal of the American Statistical Association*, **109**, 1241–1256. URL: `http://www.tandfonline.com/doi/abs/10.1080/01621459.2014.889022`.

Geraci, M. and Bottai, M. (2007) Quantile regression for longitudinal data using the asymmetric Laplace distribution. *Biostatistics*, **8**, 140–154.

Mhalla, L., Chavez-Demoulin, V. and Dupuis, D. (2020) Causal mechanism of extreme river discharges in the upper danube basin network. *JRSS, C*, **69**, 741–764.

Mhalla, L., Opitz, T. and Chavez-Demoulin, V. (2019) Exceedance-based nonlinear regression of tail dependence. *Extremes*. URL: `https://doi.org/10.1007/s10687-019-00342-6`.

Rissanen, J. (1989) *Stochastic Complexity in Statistical Inquiry Theory*. River Edge, NJ, USA: World Scientific Publishing Co., Inc.