Causal Discovery in Multivariate Extremes: A Study of Swiss Hydrological Catchments

L. Mhalla, V. Chavez-demoulin, Philippe Naveau

Environmetrics

WILEY

RESEARCH ARTICLE OPEN ACCESS

Causal Discovery in Multivariate Extremes: A Study of Swiss Hydrological Catchments

L. Mhalla[1] | V. Chavez-Demoulin[2] | P. Naveau[3]

[1]Institute of Mathematics, EPFL, Lausanne, Switzerland | [2]HEC Lausanne, University of Lausanne and Expertise Center for Climate Extremes (ECCE), Lausanne, Switzerland | [3]Laboratoire des Sciences du Climat et de l'Environnement–CNRS, Paris, France

Correspondence: L. Mhalla (linda.mhalla@epfl.ch)

Received: 9 July 2024 | **Revised:** 15 May 2025 | **Accepted:** 1 August 2025

Funding: Part of Naveau's research work was supported by European H2020 XAIDA (Grant agreement ID: 101003469) and the French Agence Nationale de la Recherche: EXSTA, the PEPR TRACCS (PC4 EXTENDING, ANR-22-EXTR-0005), the PEPR IRIMONT (France 2030 ANR-22-EXIR-0003) SHARE PEPR Maths-Vives (ANR-24-EXMA-0008) and from the Geolearning research chair, a joint initiative of Mines Paris and the French National Institute for Agricultural Research (INRAE).

ABSTRACT

Causally-induced asymmetry reflects the principle that an event qualifies as a cause only if its absence would prevent the occurrence of the effect. Thus, uncovering causal effects becomes a matter of comparing a well-defined score in both directions. Motivated by studying causal effects at extreme levels of a multivariate random vector, we propose to construct a model-agnostic causal score relying solely on the assumption of the existence of a max-domain of attraction. Based on a representation of a generalised Pareto random vector, we construct the causal score as the Wasserstein distance between the margins and a well-specified random variable. The proposed methodology is illustrated on a simulated dataset of different characteristics of catchments in Switzerland: discharge, precipitation, snowmelt, temperature, and evapotranspiration.

1 | Introduction

The primary emphasis of causality has been on causal effects pertaining to averaged outcomes. Moving from averages towards extremal quantiles can be a necessary shift for various risk analysis. For example, studying possible causes of exceptional floods, record heatwaves and their associated high societal costs can be paramount for hydrologists, climatologists, and re-insurance companies. Retrieving causal information from observational data, but at extreme levels poses a fundamental challenge across various scientific disciplines (see, e.g., Chavez-Demoulin and Mhalla 2024, for a recent review). For instance, in climatology, the discipline of extreme events attribution investigates causal connections from climate forcings such as greenhouse gases increases to observed responses on extreme phenomena like heatwaves or heavy rainfall (see, e.g., Naveau et al. 2020, for a review). In this area, causal generating mechanisms often exhibit distinct behaviours in the distribution bulk compared to its upper and lower tails. For example, moderate rainfall are influenced by factors such as prevailing wind patterns and orography. In the tails of the distribution, other atmospheric phenomena like atmospheric rivers (Dettinger 2011) can be added to the potential drivers of extreme heavy precipitation. Hence, there is a clear need, at least within the climate and hydrological communities, to develop simple and efficient causal tools for extremes. Besides a few recent theoretical advances (Engelke and Hitz 2020; Gissibl and Klüppelberg 2018), very few studies deal, in an accessible manner to practitioners, with the multivariate aspect of extremal causality. The goal of this study is to fill this gap. To motivate our particular setup, the hydrological cycle of Switzerland will be

the pedagogical thread. The reason for such a choice is that this country has been well studied in terms of its hydrological cycle (Fatichi et al. 2015), while the complex orography of Switzerland represents a challenging test bed to discover how causal extreme links may vary spatially. In particular, snow plays a crucial role in Switzerland's water cycle. Over 30% of its total precipitation falls as snow, and around 40% of Swiss river runoff comes from snowmelt. Seasonal snow acts as a temporary reservoir for precipitation, releasing it in a condensed manner over a relatively short time, leading to variable seasonal runoff (Schirmer and Jonas 2023). The complex atmospheric links between the hydrological variables are dynamic both spatially and temporally (Brunner et al. 2019) and are tightly connected to meteorological variables such as temperature and hydrometeorological variables such as evapotranspiration. Therefore, temperature and evapotranspiration will be incorporated into the analysis to investigate potential causal relationships among precipitation, discharge, snowmelt, temperature, and evapotranspiration, particularly under conditions where at least one variable is subject to extreme events.

To address this issue, we first need to recall two statistical models that have been particularly highlighted in the recent causal literature dealing with extremes: heavy-tailed linear structural causal models (LSCM) studied by Gnecco et al. (2021) and recursive max-linear models (RMLM) introduced by Gissibl and Klüppelberg (2018). Under the LSCM setting, a causal framework among multiple variables, say $\mathbf{Y} = (Y_1, \ldots, Y_d)^\top$, is delineated by a directed acyclic graph (DAG), denoted $\mathcal{G}$, where the node set $V := \{1, \ldots, d\}$ denotes random variable indices and the set of directed edges represents direct causal effects. The following relation

$$Y_j := \sum_{k \in \mathrm{pa}(j)} \beta_{kj} Y_k + \varepsilon_j, \quad j \in V \tag{1}$$

where $\mathrm{pa}(j) \subseteq V$ is the set of parents of j, $\beta_{kj} \in \mathbb{R} \setminus \{0\}$ is the causal weight of node k on node j, and $\varepsilon_1, \ldots, \varepsilon_d$ are jointly independent heavy-tailed noise variables, defines a heavy-tailed linear structural causal model (LSCM) with associated directed acyclic graph $\mathcal{G}$, in which the directed edge $(i, j) \in V \times V$ belongs to the set of vertices if and only if $i \in \mathrm{pa}(j)$. An extreme node observation Y_j is the result of an extreme noise ε_j, or of a sum of weighted observations from the parents of j in $\mathcal{G}$. In contrast to the additive structure of a LSCM, a RMLM is based on the max-operator, that is, maxima over weighted parent nodes are considered. The main idea is to propagate extremes throughout the max-linear equation defined as

$$Y_j := \max_{k \in \mathrm{pa}(j)} \max(c_{kj} Y_k, c_{jj} \varepsilon_j), \quad j \in V \tag{2}$$

with strictly positive weights c_{kj} for all $j \in V$ and $k \in \mathrm{pa}(j) \cup j$. The independent non-negative random variables $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_d)^\top \in \mathbb{R}_+^d$ represent the vector of innovations. An extreme node observation Y_j results from either an extreme innovation ε_j or from a large maximum of some weighted observations from Y_j's parents in $\mathcal{G}$. A particular feature of the model (2) is that the resulting joint distribution is discrete which may be inconvenient for some applications. Both models (1) and (2) can be expressed with a DAG. In addition, causal inference methods based on the LSCM (1) and RMLM (2) rely on the assumption of exogeneity, where predictor variables are assumed to be independent of the error term. This assumption is challenged by the presence of cycles that complexify the task of identifying causal directions and that are frequently encountered in the highly coupled Earth system. To handle possible presence of feedback loops within a causal study, we propose a novel definition of causality at extreme levels. The proposed definition moves away from the concept of structural equations models (SEM) and relies on a model-agnostic causal metric that is solely based on the assumption of the presence of a maximum domain of attraction (Ferreira and de Haan 2014). More precisely, leveraging a representation of a multivariate generalised Pareto random vector (see, e.g., Rootzén et al. 2018), we formulate the causal metric as the Wasserstein distance between marginal distributions and a well-defined random variable.

The paper is structured as follows. In Section 2, we review the basics of multivariate generalised Pareto models and recall some parametric examples that will be used in the simulation study. In Section 3, we propose a new definition of extremal causality in the bivariate setting. We assess our new causal methodology in Section 4 and apply our method to the Swiss hydrological system in Section 5. The paper ends by a conclusion in Section 6. In terms of notations, the multivariate sample $(\mathbf{Y}_1, \ldots, \mathbf{Y}_n)^\top$ of size n corresponds to independent and identically distributed real-valued copies of the d-dimensional random vector $\mathbf{Y} = (Y_1, \ldots, Y_d)^\top$ with associated realisation $\mathbf{y} = (y_1, \ldots, y_d)^\top$.

2 | Multivariate Generalised Pareto Model

A fundamental aspect of univariate extreme-value analysis involves fitting a generalised Pareto (GP) survival distribution to a set of exceedances beyond a high threshold. This survival GP distribution can be characterized by a positive scale parameter σ and a shape parameter ξ and is equal to

$$\overline{H}(z; \sigma, \xi) = \left(1 + \frac{\xi z}{\sigma}\right)_+^{-1/\xi}, \quad z > 0$$

where $\sigma > 0$, $\xi \in \mathbb{R}$, and $a_+ = \max(a, 0)$. This two-parameter expression allows for relatively straightforward statistical inference (Davison and Smith 1990). In the multivariate extremes context, the situation becomes complex as there is not a unique way to define a multivariate extreme event. In addition, the family of distributions suggested by asymptotic theory no longer remains parametric (Ferreira and de Haan 2014). Following Rootzén and Tajvidi (2006), we define multivariate extremes whenever, at least, one component of $\mathbf{Y}$ exceeds a large value. Such an extreme event is denoted[1] by $[\mathbf{Y} - \mathbf{u} | \mathbf{Y} \not\leq \mathbf{u}]$, with $\mathbf{u} = (u_1, \ldots, u_d)^\top$ a vector of large thresholds, and has support on the L-shaped region $\{\mathbf{x} \in \mathbb{R}^d : ||\mathbf{x}||_\infty > 0\}$. We also assume that $\mathbf{Y}$ belongs to the domain of attraction of a max-stable distribution, that is, if $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ are independent copies of $\mathbf{Y}$, there exist sequences $\mathbf{a}_n \in (\mathbf{0}, \infty)^d$ and $\mathbf{b}_n \in \mathbb{R}^d$ such that the distribution of the correctly re-normalized componentwise maxima defined by

$$\max_{i=1,\ldots,n} (\mathbf{Y}_i - \mathbf{b}_n)/\mathbf{a}_n$$

has a non-degenerate distribution as n gets large. It has been shown by Rootzén et al. (2018) and Ferreira and de Haan (2014) that the conditional vector $\mathbf{Y}-\mathbf{u}|\mathbf{Y}\not\leq\mathbf{u}$ can then be approximated, as $\mathbf{u}$ gets large, by a random vector $\mathbf{Z}$ with a multivariate generalised Pareto (MGP) distribution (see Naveau and Segers 2024, for a recent review on multivariate exteme value theory). Any conditional margin of the multivariate GP distributed $\mathbf{Z}$ has a univariate GP distribution in the sense that, for any $j=1,\ldots,d$,

$$\Pr(Z_j > z|Z_j > 0) = \overline{H}(z;\sigma_j,\xi_j)$$

where $\sigma_j > 0$ and $\xi_j \in \mathbb{R}$ correspond to the marginal scale and shape parameters, respectively. Any MGP vector $\mathbf{Z}$ with marginal parameters $\boldsymbol{\xi}=(\xi_1,\ldots,\xi_d)^\top$ and $\boldsymbol{\sigma}=(\sigma_1,\ldots,\sigma_d)^\top$ can be rewritten in standardized version in the following way

$$\mathbf{Z} \overset{d}{=} \boldsymbol{\sigma}\frac{e^{\boldsymbol{\xi}\mathbf{X}}-1}{\boldsymbol{\xi}} \tag{3}$$

with

$$\mathbf{X} = E + \mathbf{U} - \max(\mathbf{U}) \tag{4}$$

where E represents a univariate unit exponential random variable and $\mathbf{U}$ any multivariate random vector, independent of E and $\max(\mathbf{U})=\max_{1\leq j\leq d} U_j$. The multivariate vector $\mathbf{X}$ is called a standard Pareto vector with unit scale $\boldsymbol{\sigma}=\mathbf{1}$ and zero shape parameters $\boldsymbol{\xi}=\mathbf{0}$ and, by construction, its support is $\{\mathbf{x}\in\mathbb{R}^d : \mathbf{x}\not\leq\mathbf{0}\}$. As there is no constraint on the choice of $\mathbf{U}$, the dependence structure in the vector X is basically free and non-parametric by nature. In particular, if the random vector $\mathbf{U}$ has density $f_{\mathbf{U}}$ defined on $(-\infty,\infty)^d$, then the density of $\mathbf{X}$ can be expressed as

$$h_{\mathbf{X}}(\mathbf{x};\mathbf{1};\mathbf{0}) = \frac{\mathbb{1}_{\{\max(\mathbf{x})>\mathbf{0}\}}}{e^{\max(\mathbf{x})}}\int_0^\infty f_{\mathbf{U}}(\mathbf{x}+\log t)t^{-1}dt$$

Another construction of standard MGP pdfs is due to Rootzén et al. (2018). Suppose a d-dimensional random vector $\mathbf{T}$ with density $f_{\mathbf{T}}$ that satisfies $\mathbb{E}\left[e^{T_j}\right]<\infty$, for all $j=1,\ldots,d$, then a density of a multivariate GP distribution can be extracted as

$$h_{\mathbf{T}}(\mathbf{z};\mathbf{1};\mathbf{0}) = \frac{\mathbb{1}_{\{\max(\mathbf{z})>\mathbf{0}\}}}{\mathbb{E}\left[e^{\max(\mathbf{T})}\right]}\int_0^\infty f_{\mathbf{T}}(\mathbf{z}+\log t)dt \tag{5}$$

This representation has the convenient property that any subvector $\mathbf{T}'$ with at least one component above 0 of a GP random vector $\mathbf{Z}$ with density $h_{\mathbf{T}}$ is GP with same density $h_{\mathbf{T}'}$ adjusted to the dimension of $\mathbf{T}'$. Kiriliouk et al. (2018) provides a review of constructions of GP vectors. From this work, we list below three parametric models for the random vector $\mathbf{T}$ that lead to well-known parametric multivariate extreme value distributions.

Example 2.1 (Logistic max-stable distribution). Let $\mathbf{W}\in\mathbb{R}^d$ be a random vector with independent Gumbel components with equal positive scale α and defined by

$$P(W_j\leq w) = \exp\{\exp(-\alpha w)\}$$

If $\mathbf{T}\overset{D}{=}\mathbf{W}$ but with the restriction $\mathbb{E}[e^{T_j}]<\infty$, then (5) leads to the multivariare GP density with support $\{\mathbf{z}\in\mathbb{R}^d : \mathbf{z}\not\leq\mathbf{0}\}$ and associated logistic max-stable distribution. Note that in a similar way, the reverse Gumbel independent components lead to the multivariate GP distribution associated to the negative logistic max-stable distribution.

Example 2.2 (Dirichlet max-stable distribution). Same construction as in Example 2.1 but with W_j following the pdf

$$f_j(w) = \exp(\alpha_j w)\exp\{-\exp(w)\}/\Gamma(\alpha_j)$$

for $\alpha_j>0$ and $w\in(-\infty,\infty)$.

Example 2.3 (Hüsler–Reiss max-stable model). Same construction as in Example 2.1 but with $\mathbf{W}$ following the pdf

$$f_{\mathbf{W}}(\mathbf{w}) = (2\pi)^{-d/2}|\Sigma|^{-1/2}\exp\{-(\mathbf{w}-\boldsymbol{\beta})^T\Sigma^{-1}(\mathbf{w}-\boldsymbol{\beta})/2\}$$

where $\boldsymbol{\beta}\in\mathbb{R}^d$ is the mean vector and $\Sigma\in\mathbb{R}^{d\times d}$ is the positive-definite covariance matrix.

In the following section, we connect representation (4) to our new definition of causality for extremes.

3 | Defining Extremal Causality Through MGP Models

We consider the bivariate setting of a random vector $(Y_1,Y_2)^\top$ which represents two nodes of a graph. We suppose that the limiting tail behaviour of $(Y_1,Y_2)^\top$ is described by the MGP vector $\mathbf{Z}$ defined by (3) and its related standard Pareto vector $\mathbf{X}$. To understand the dependence strength between each component of $\mathbf{X}=(X_1,X_2)^\top$, we introduce the difference $V=X_1-X_2=U_1-U_2$ and we remark that Equation (4) can be written as

$$\begin{cases} X_1 = E + V - \max(0,V), \\ X_2 = E - \max(0,V) \end{cases} \tag{6}$$

where V and E are independent. From this system, we deduce that a absolute value of V close to zero corresponds to a strong dependence between X_1 and X_2, while a large value of $|V|$ reflects almost independence. In other terms, the strength of dependence in $\mathbf{X}=(X_1,X_2)^\top$ is fully described by the random scalar V and a strong (weak) departure from zero reflects a weak (strong) dependence within $\mathbf{X}$. Through the key role of $\max(0,V)$ in (6), the (a)symmetry of V will play a central role in our definition of extremal causality. If the distribution of V is strongly asymmetrical, say predominantly taking negative values, then $X_1=E+V$ and $X_2=E$ will occur more frequently than the alternative $X_1=E$ and $X_2=E-V$. On the other hand if the probability of V being positive is higher than that of being negative, than X_1 would be more often equal to the unit exponential E. Here, we argue that this asymmetrical feature in the extremal dependence structure can be exploited to detect extremal causality. Intuitively, if the extremal causal structure in $(Y_1,Y_2)^\top$ is monotonic, then we would expect that an extreme event in Y_1 would always cause an extreme event in Y_2 (assuming Y_1 to be the parent of Y_2 in the L-shaped region), while the opposite does not necessarily hold. Thus, we expect V to be more often negative than positive and X_2 to be closer to a unit exponential than X_1. Therefore, we propose to define the causally-induced asymmetry in the tails of $(Y_1,Y_2)^\top$ by comparing a distance with respect to the unit-exponential E. Different distances exist to measure the proximity to a target pdf.

In our case, the Wasserstein distance leads to direct computations with a simple interpretation in terms of the means of X_1 and X_2. The Wasserstein distance (Monge 1781; Villani 2008) is a metric between probability measures where one is interested in the "minimal effort" of moving one (probability) measure to another. It is a particularly interesting proper metric when one of the probability measures is derived from the other with a small random perturbation. The Wasserstein distance between two univariate random variables A and B is defined as

$$W_p(A,B) = \left\{ \int_0^1 \left(|F^{-1}(q) - G^{-1}(q)| \right)^p dq \right\}^{1/p}$$

with $F(x) = P(A \leq x)$ and $G(x) = P(B \leq x)$. The special case $p = 1$ is equivalent to

$$W_1(A,B) = \int_{\mathbb{R}} |F(t) - G(t)| dt \tag{7}$$

The following proposition (see proof in the Appendix A) is the stepping stone for our proposed definition of extremal causality.

Proposition 1. *Let* $\mathbf{X}$ *be a standard bivariate generalised Pareto vector expressed as in* (6)*. Then, the following equivalence holds*

$$W_1(X_1, E) \geq W_1(X_2, E) \Leftrightarrow \mathbb{E}(X_1) \leq \mathbb{E}(X_2)$$

Therefore, a comparison of the Wasserstein distances between the scaled tail margins and the unit exponential distribution is informative about the asymmetry in the tail dependence through the sign of $\mathbb{E}(X_1 - X_2)$. This is convenient as we have previously highlighted that the asymmetry in the difference $X_1 - X_2$ was key in the structure of (6). Proposition 1 tells us that a simple difference in means is enough to capture asymmetries in the Wasserstein distances to the unit-exponential. We can now introduce a novel notion of extremal causality in the d-dimensional setting.

Definition 3.1. Let $\mathbf{Y} = (Y_1, \ldots, Y_d)^\top$ be a d-dimensional vector with a joint distribution that is in the max-domain of attraction of a multivariate max-stable distribution. Denote by $\mathbf{X}$ its limiting standard generalised Pareto vector, and let

$$s_{i \to j} = \frac{W_1(X_i, E) - W_1(X_j, E)}{\max\limits_{k=1,\ldots,d} W_1(X_k, E)} \tag{8}$$

We refer to $s_{i\to j}$ as the causal score from the component i to the component j. If $s_{i\to j}$ is finite and strictly positive, then we say that the component i is the extremal cause of the component j.

The score (8) captures causally-induced asymmetries in the tail of $\mathbf{Y}$ regardless of the strength of extremal dependence. Without the standardisation term in (8), the score would be identical wherever both X_1 and X_2 are equally close to E or far from E. In general, low scores indicate that causality is weak. In the case of strong tail dependence, for example, where all distances are comparable, the causal score can be low even when tail causality is present. The latter scenario aligns with a non-identifiable case according to our definition. This can be analogized to non-extreme causal discovery methods that utilize restricted additive noise models, wherein identifiability hinges on the presence of either a non-linear causal effect or non-Gaussian noise. This effect is illustrated on SEM relations in Figure 3, when β is large. We should note that although the definition relies on a bivariate score, extremal causal discovery in the multivariate setting is equivalent to finding the topological order of the graph. For instance, the classical notion of conditional independence implies unconditional independence in max-stable vectors (Papastathopoulos and Strokorb 2016). Hence, the causal graph associated to the standard Pareto vector $\mathbf{X}$ in (4) is fully connected. Relying on our proposed definition of extremal causality, a topological order is directly retrieved by ordering the Wasserstein distances $W_1(X_i, E)$.

Figure 1 displays 10^4 bivariate samples of $(X_1, X_2)^\top$ derived from various distributions of $V = X_1 - X_2$ highlighting the impact of its asymmetry around zero and the irrelevance of its range, and hence of the strength of tail dependence, on the causal score. The color of each point indicates the value of V. For example, the two upper panels showcase symmetrical V, either following a zero-mean Gaussian (upper left) or defined as the difference between two independent Gumbel distributed random variables (upper right). The distributional symmetry is aligned with the absence of causal connection regardless of whether the tail dependence is strong (small values of V) or weak (large values of V). In contrast, the two lower panels correspond to asymmetrical V, especially in the lower left corner.

Definition 3.1 of extremal causality is broad. For instance, the structural equation models (SEM) defined by the LSCM (1) and the RMLM (2) with Pareto noise fall within this definition. That is, if two vertices are causally related in the heavy-tailed SEM, then the SEM-associated parent is the *extremal cause* of the SEM-associated child. Following our definition of extremal cause, then V is asymmetric with more negative values on the child axis. Figure 2 illustrates this situation where Y_1 is the parent of Y_2 in the LSCM (1) (top panels) and RMLM (2) (bottom panels) with $\beta = 1.2$ (left) and $\beta = 0.2$ (right) and Pareto noise ε_j, $j = 1, 2$, with shape parameter $\xi = 0.1$ in all these heavy-tailed cases. The asymmetry of V is more pronounced for large values of β as all extremes generated by Y_1 (or equivalently by ε_1) are also extremes of Y_2 but Y_2 also generates its extremes through ε_2 as β is close to 1. Consequently, the score (8) is higher for larger values of β. Causal links encoded in a heavy-tailed LSCM or RMLM remain valid at extremal levels and according to our definition, higher (absolute) values of (8) reflect stronger causal links, though the relation between the two is mediated by the tail index of the noise, as assessed by a simulation study in the next section.

To summarize, our definition of extremal causality stems from the strength and asymmetry of the extremal dependence structure. Specifically, when extremal dependence exhibits both strength and asymmetry, it indicates a potent extremal causal connection, even in the absence of structural causal relationships between variables. On the other hand, when extremal dependence is asymmetric but weak, it suggests a less influential extremal causal link. In cases where extremal dependence is either strong or weak but symmetrical, it signifies the absence of an extremal causal connection altogether.

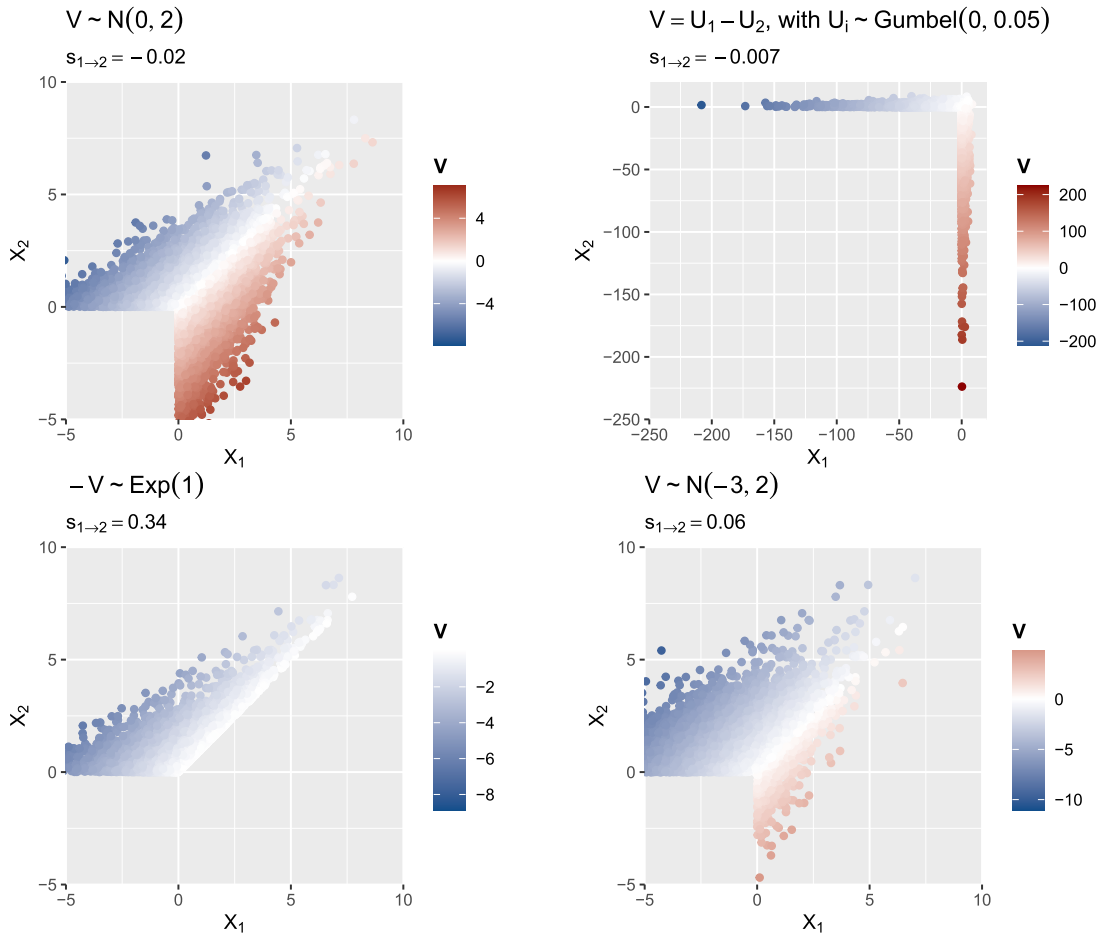

FIGURE 1 | MGPD samples ($n = 10^4$) from (6) with different distributions for the random variable V reflecting different tail behaviours: symmetric with strong dependence (top left panel), symmetric with weak dependence (top right panel), asymmetric with strong dependence (bottom left), and asymmetric with moderate dependence (bottom right panel).

4 | Simulation Study

We run different simulation studies to assess our extreme causality method in various contexts. In the sequel, marginal transformation of the data to the standard Pareto scale is performed empirically, that is, using a rank-based transformation. The empirical transformation led to nearly identical results compared to a semi-parametric modelling of the margins where marginal distributions are inferred by the empirical distribution below a high threshold and the limiting Generalised Pareto distribution above that threshold.

4.1 | Bivariate Setups

The first simulation study is based on the SEM relations LSCM (1) and RMLM (2). The heavy-tailed LSCM and RMLM from which we simulate $n = 10^4$ data and assume a MGPD above a marginal threshold at the 95% quantile are respectively

$$\text{LSCM:} \begin{cases} Y_1 &= \varepsilon_1, \\ Y_2 &= \beta Y_1 + \varepsilon_2 \end{cases}$$

and

$$\text{RMLM:} \begin{cases} Y_1 &= \varepsilon_1, \\ Y_2 &= \max(\beta Y_1, \varepsilon_2) \end{cases}$$

both with $\varepsilon_j \sim$ Pareto with shape parameter $\xi = 0.1$ and $\xi = 0.3$ and with β varying from 0.1 to 3. The boxplots of the resulting causal scores (8), represented in Figure 3, lay all on the strictly positive side for all values of β. For low values of β, causality exhibits reduced strength in both SEM relations, consequently yielding smaller scores. At the uppermost values of β, for a small shape parameter ξ, the SEM relations lean towards perfect dependence, rendering causal discovery more challenging, especially for RMLM, but still achievable. We finally run a sensitivity analysis of the causal score to a setup where the cause and the effect have different tail indices, namely ξ_1 for ε_1 and ξ_2 for ε_2. The boxplots of the resulting causal scores (8) are represented in Figure 4 and confirm the findings of Gnecco et al. (2021) that causal discovery becomes challenging when the ancestor has heavier tails than the descendant. For instance, when the cause has a heavier tail than the effect, causal discovery is achievable only for small values of β, after which the causal score exponentially decreases to zero without crossing it. However, when the effect is

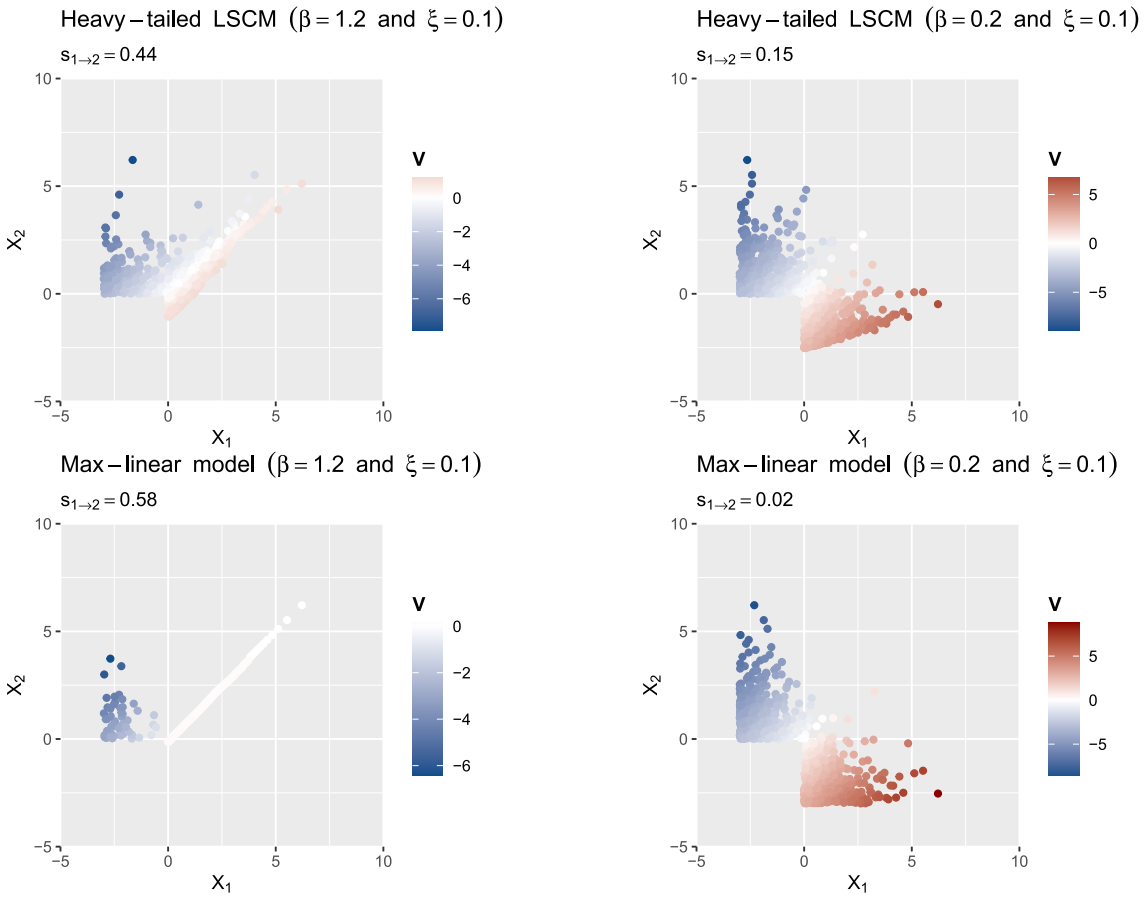

FIGURE 2 | Bivariate samples ($n = 10^4$) from LSCM (1) and RMLM (2).

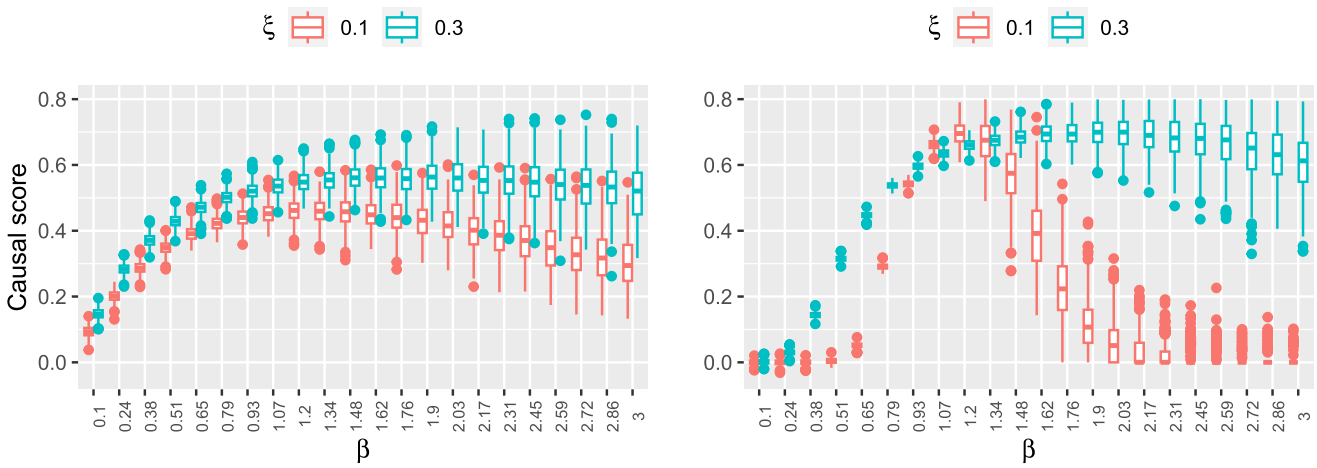

FIGURE 3 | Score (8) $s_{1\to 2}$ based on simulated data from LSCM (1) (left) and RMLM (2) (right). Results rely on 500 bootstrap replicates.

heavier-tailed than the cause, the causal score behaves similarly to Figure 3 when $\xi = 0.3$.

We now run logistic model-based simulations. They are based on simulating $n = 10^4$ data from a logistic or an asymmetric logistic extreme value copula. The bivariate asymmetric logistic copula introduced by Tawn (1990) is

$$C_{\alpha,\beta_1,\beta_2}(u,v) = \exp\left[-\left\{(-\beta_1 \log u)^{1/\alpha} + (-\beta_2 \log v)^{1/\alpha}\right\}^{\alpha} + (1-\beta_1)\log u + (1-\beta_2)\log v\right]$$

with $\beta_1, \beta_2 \in [0,1]$, the asymmetry parameters. The case $\beta_1 = \beta_2 = 1$ defines the (symmetric) logistic copula. The parameter

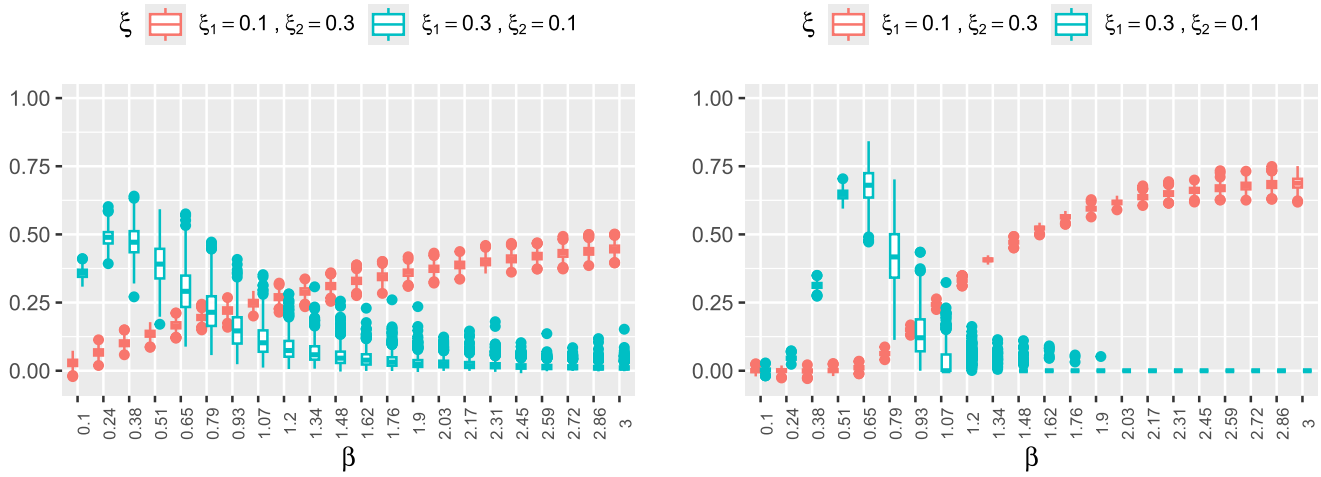

FIGURE 4 | Score (8) $s_{1\to 2}$ based on simulated data from LSCM (1) (left) and RMLM (2) (right) with different tail indices. Results rely on 500 bootstrap replicates.

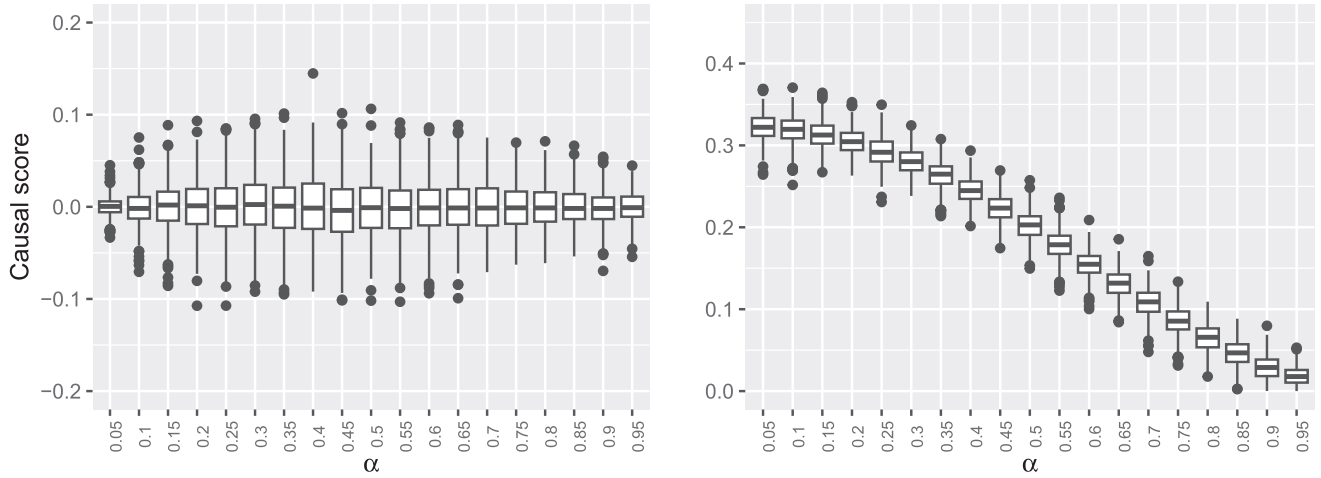

FIGURE 5 | Score (8) based on simulated data from logistic model with varying dependence parameter (left) and from asymmetric logistic model with varying dependence parameter (right). Results rely on 500 bootstrap replicates.

$\alpha \in (0, 1]$ specifies the strength of dependence with values close to zero corresponding to strong dependence and values close to one, to independence. For both the logistic and asymmetric logistic cases, we apply our extremal causality method on the MGPD distributed data above their marginal 95%-quantile. Figure 5 shows the resulting scores for the logistic data (left panel) and for the asymmetric logistic data with stronger asymmetry parameter ($\beta_1 = 0.8$) for the first component than for the second ($\beta_2 = 0.2$) (right panel) against different values of the tail dependence parameter α.

As expected, the score is close to zero in case of the symmetric logistic model and positive for the asymmetric model. The positiveness adequately suggests that the first component is the cause of the second. Again, the higher is the tail dependence (or equivalently the smaller is the value α), the stronger is the evidence for extremal causality. Our causal score can be related to the asymmetric tail Kendall's τ introduced by Deidda et al. (2023) where it is shown how it can be used to inform the direction of causality between the extreme observations that present asymmetric tail dependence structures. The findings from their simulated asymmetric copula and from ours for this model are aligned.

4.2 | Presence of Confounding

Building on the LSCM structure, we now consider two settings with a confounder variable Y_1 and the absence or presence of a direct causal link between the variables of interest Y_2 and Y_3. Precisely, the former is

$$\begin{cases} Y_1 &= \varepsilon_1, \\ Y_2 &= \beta Y_1 + \varepsilon_2, \\ Y_3 &= \beta Y_1 + \varepsilon_3 \end{cases}$$

and the latter is

$$\begin{cases} Y_1 &= \varepsilon_1, \\ Y_2 &= \beta Y_1 + \varepsilon_2, \\ Y_3 &= \beta Y_1 + \gamma Y_2 + \varepsilon_3 \end{cases}$$

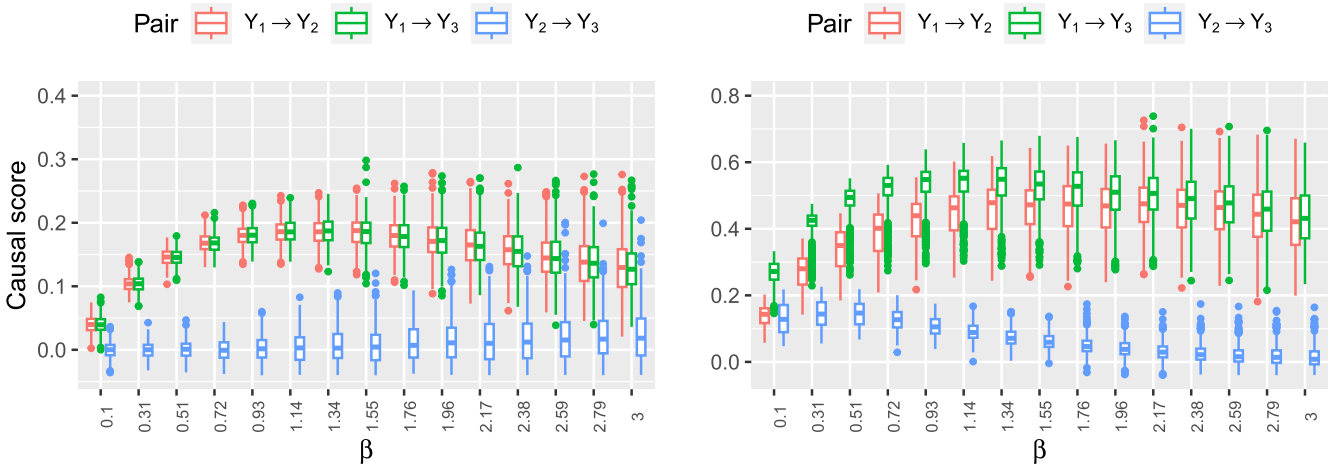

FIGURE 6 | Score (8) based on simulated data from LSCM (1) with confounder and no direct causal link (left) and with direct link (right). Results rely on 500 bootstrap replicates and the shape parameter ξ is set at 0.1 for all noise variables.

where γ is a random coefficient drawn between 0.1 and 3. The boxplots of the resulting causal scores (8) are represented in Figure 6. When Y_2 and Y_3 are not causally related, the presence of the confounder Y_1 has no impact on the causal score $s_{2\to3}$ which is centered around zero. The causal scores $s_{1\to2}$ and $s_{1\to3}$ behave similarly to the LSCM setting in Figure 3 (left panel), as expected. In the presence of a direct causal link between Y_2 and Y_3, $s_{1\to3}$ is on average higher than $s_{1\to2}$ due to the presence of two causal pathways between the variables Y_1 and Y_3. As the causal coefficient γ is randomly chosen, the causal score $s_{2\to3}$ seems to be pivoted by the coefficient β and is strictly positive for small values of β and decreases rapidly to zero with increasing β.

4.3 | Higher-Dimensional Setups

We assess the performance of our method in estimating a causal order of a graph induced by the LSCM (1) or the RMLM (2) structure. In both structural models, the heavy-tailed noise variables are assumed to have a Pareto distribution with different values for the shape parameter ξ. We consider random coefficients of the structural models that are uniformly selected between 0.1 and 4 and different combinations of the sample size n, the number of variables p, and the shape parameter ξ of the noise variables. For each setting, we simulate 100 random structural models, over which we compute the mean of a given performance metric. Similarly to Gnecco et al. (2021), we choose the structural intervention distance (SID) of Peters and Bühlmann (2015) to evaluate the performance of our method in estimating a causal order. As the range of values taken by the SID depends on the number of vertices of the DAG, we normalise the distances between the true DAG and the inferred one, to lie between 0 and 1. We compare our method to the EASE algorithm of Gnecco et al. (2021) where the number of exceedances is set at $\lfloor n^{0.4} \rfloor$, as recommended by the authors. In the sequel, we transform the data empirically to standard Pareto marginals and then to the standard MGP scale by setting the marginal thresholds at the 95% quantile of the Pareto distribution.

Figure 7 displays the results of the simulations with different tail heaviness of the noise variables of the structural models.

We observe that our proposed method for recovering a causal order performs better than the EASE algorithm of Gnecco et al. (2021) in all considered settings. We report better performances in heavier-tailed settings, that is, larger values of ξ, similarly to our findings in the bivariate setting. In addition to its competitive performances under different settings, our proposed methodology enjoys a computational advantage over EASE as it only requires ordering the estimated Wasserstein distances rather than a greedy search of root nodes; see Gnecco et al. (2021) for details about the EASE algorithm.

5 | Extremal Analysis of Swiss Hydrological Catchments

During the spring season, in particular, extreme precipitation events can directly lead to increased surface runoff and contribute to higher water levels in rivers and streams. These events can also result in saturation of soils, leading to increased infiltration and groundwater recharge, which can contribute to higher baseflow in rivers over time. In Switzerland, mountain regions have significant snowpack, and extreme snowmelt can occur due to rapid warming or rain-on-snow events in spring. Rapid snowmelt can in turn lead to increased surface runoff and contribute to higher river discharges, particularly in spring when snowpacks are melting. Extreme discharges in rivers and streams are therefore often the result of a combination of factors, including extreme precipitation, warm conditions resulting in rapid snowmelt, and high evapotranspiration. In the Plateau, atmospheric processes and topographical features can lead, under conditions such as condensation, to situations where precipitations are due to snowmelt.

We consider data simulated using the hydrological modelling system PREVAH (PREecipitation-Runoff-EVApotranspiration Hydrotope model) (Viviroli et al. 2009). The dataset consists of 307 catchments in Switzerland for which discharge, precipitation, snowmelt, temperature, and actual evapotranspiration values were simulated at a daily-resolution from 1981 to 2016 (Zappa and Brunner 2019). The catchments' flood events are mainly driven either by snowmelt (Alps) or rainfall (Jura,

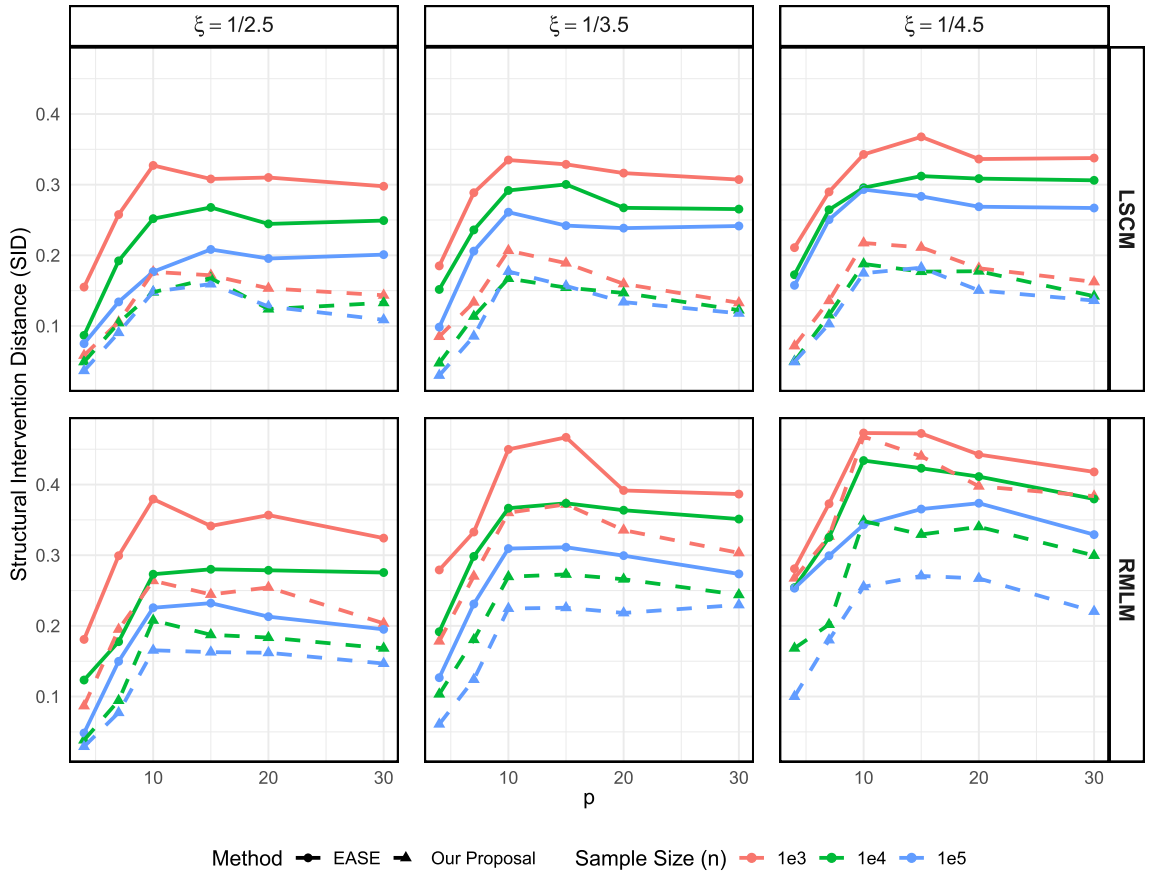

FIGURE 7 | SID values averaged over 100 simulations for different combinations of sample size n, dimension p, and shape parameter ξ. The first row corresponds to the LSCM model and the second to the RMLM model.

Plateau, and Southern Alps) or by their mixture (Pre-Alps), see Froidevaux et al. (2015). The system of the hydrological and (hydro)meteorological variables is spatially dynamic and the goal is to assess the extremal causal mechanisms over all catchments during spring (March–April–May).

Understanding the causal dynamics among the five variables (precipitation, discharges, snowmelt, temperature, and evapotranspiration) at their extreme levels is crucial for effective water resources management, flood risk assessment, climate change adaptation efforts, and for accurately predicting and mitigating the impacts of extreme hydrological events. As illustrated in Figure 8, Switzerland is a relief country where overall, altitude influences the spatial distribution of precipitation, the duration and timing of snow accumulation and melt, and the characteristics of river basins, all of which play critical roles in shaping the causal dynamics among extreme precipitation, extreme snowmelt, and extreme discharges in mountainous regions.

The objective of our study is the evaluation of extreme causal mechanisms between the five variables when at least one of them is extreme, that is higher than its 90% empirical marginal quantile, a situation that we will call "under extreme condition". To remove the time lag effects and temporally align the variables without imposing a (time) directional bias on their causal dynamics, we pre-process the data such that marginal events that might result in flooding are contemporaneous. For instance, we consider moving windows of three days[2] over which accumulated precipitation, accumulated snowmelt, accumulated (actual) evapotranspiration, mean temperatures, and discharge levels at the central point are computed. This results in a dataset of 3204 observations. While the pre-processing aims at aligning potential isolated or compound extreme events, we need to ensure that catchments experienced snowmelt over the considered time frame, that is, simulated values for this variable can be non-zero. This was not, for instance, the case for three out of the 307 catchments, where all simulated snowmelt values during the March–April–May period were zero. For these catchments that have small areas, though not the smallest in the dataset, we exclude snowmelt from the causal discovery in extreme events. For the rest of the catchments where non-zero values of snowmelt are present, we remove instances where no snowmelt was observed and perform causal discovery for the remaining variables. This way, we avoid biasing our results with observations under an extreme condition but with a complete absence of snowmelt. Figure 9 displays the pairs of variables on the standard Pareto vector scale and at a high altitude catchment in the Southern Alps.

For each of the 307 catchments over Switzerland, we consider events under extreme condition and apply our causal discovery method. That is, we empirically transform the data to the MGP

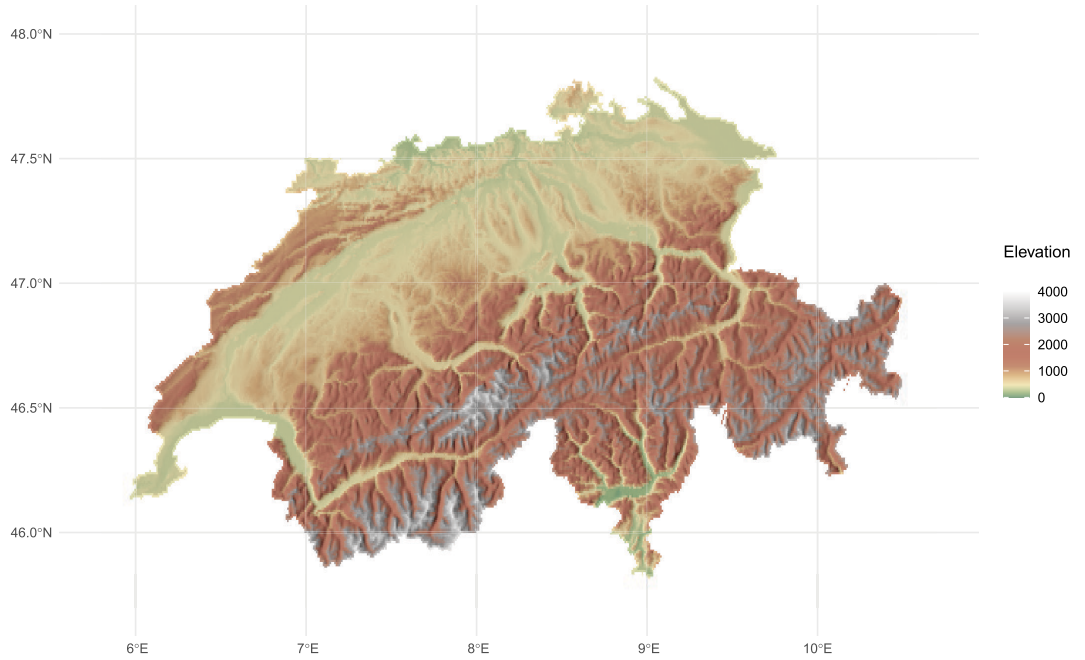

FIGURE 8 | Elevation map of Switzerland.

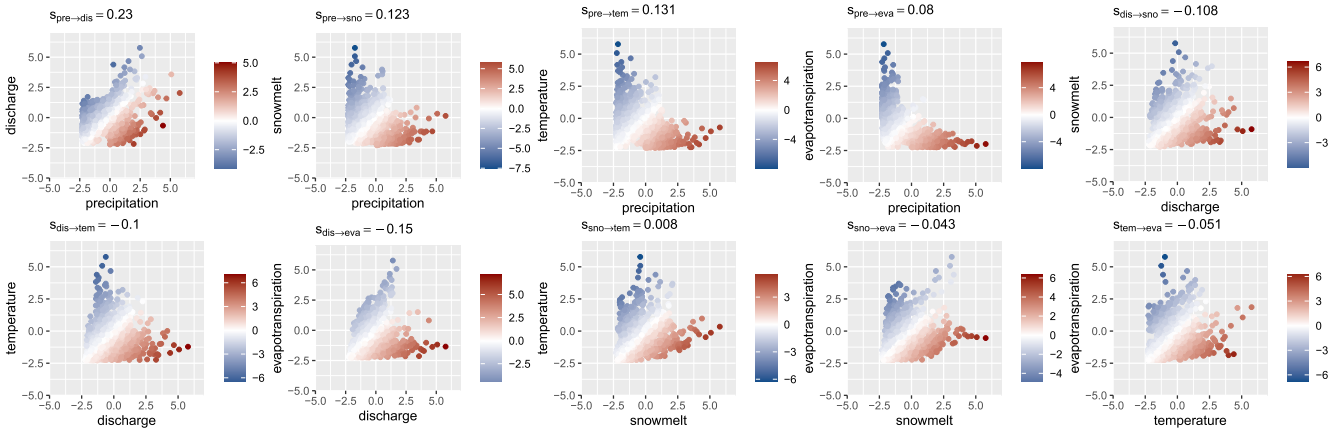

FIGURE 9 | Pairs of variables transformed empirically to the standard Pareto scale, at a high altitude catchment in the Southern Alps. Colouring scheme reflects the magnitude of the difference between the displayed margins.

scale using the probability integral transform and the marginal empirical distribution, compute the causal score (8) for each pair of variables considered, and retain significant scores only when zero is not in the 95% confidence interval obtained from bootstrapping the data 500 times. The panels of Figure 10 show the resulting values of the score over the different regions of Switzerland in spring (March–April–May). In all panels, grey areas represent catchments where the causal score is not significantly different from zero. Panels displaying the causal score for a pair of variables including evapotranspiration contain yellow catchments that should be excluded as the causal score had a high significant absolute value with the opposite sign than neighbouring low-altitude catchments.

The top two rows display the spatial behaviour of the causal score in pairs including the discharge variable. All panels highlight the causal effect of precipitation, snowmelt, temperature, and evapotranspiration on discharge values when the catchment is under extreme condition. The expected effects of precipitation and snowmelt on discharge is highlighted by higher scores at altitude. Indeed in mountainous regions, the interaction between extreme precipitation and extreme discharge in spring can be enhanced by extreme snowmelt situations in areas where snowpack accumulation can reach significant depths. Panels in the middle row identify a contrasting behaviour between low and high altitude catchments when it comes to the extremal causal links between precipitation and temperature as well as precipitation and snowmelt. For instance, the causal effect of precipitation on snowmelt under extreme condition is only present at very high altitude. During spring, warm fronts may bring rain to higher elevations where the snowpack exists. This rain falling onto the snowpack can accelerate the melting process. Rain-on-snow

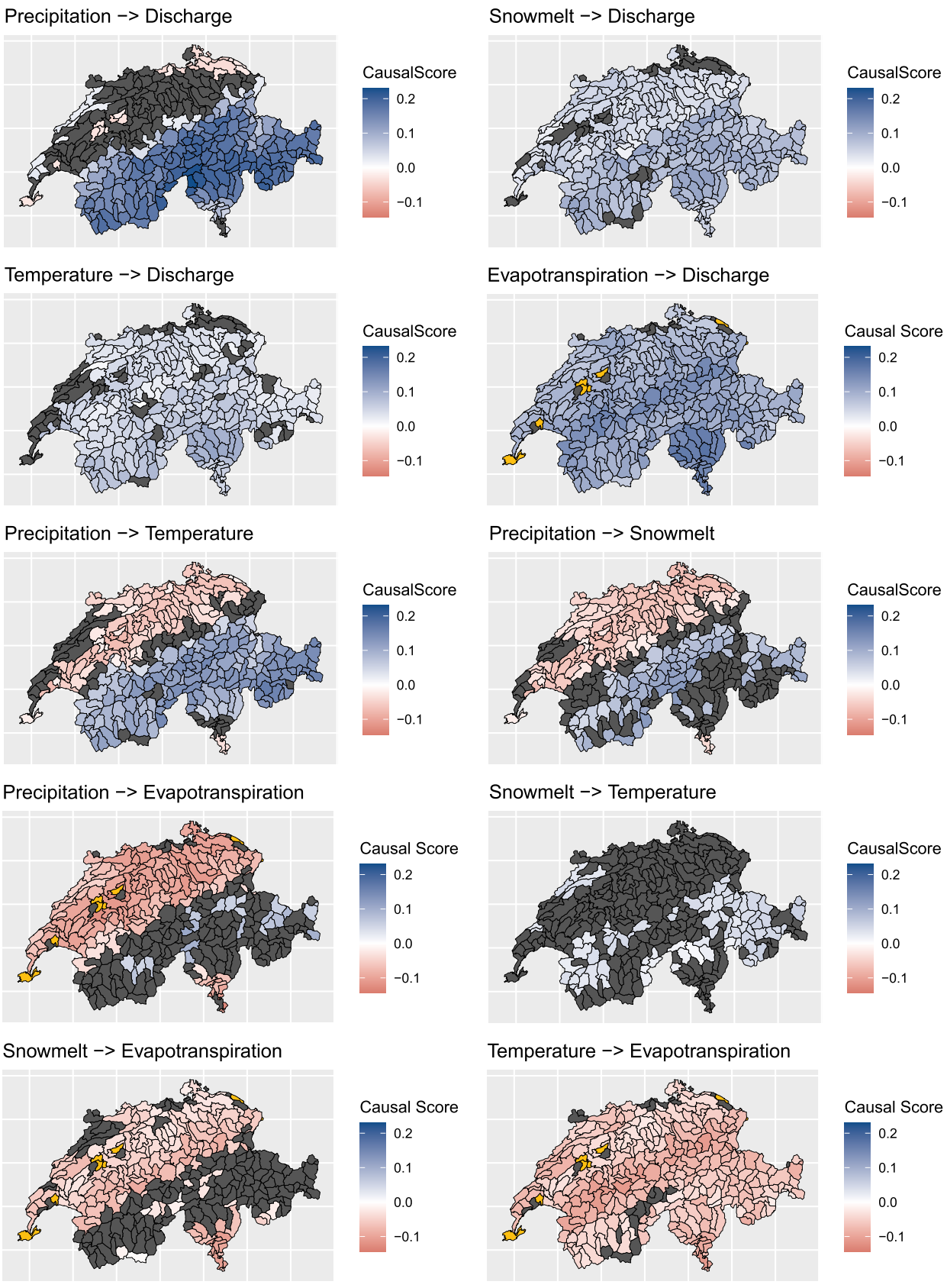

FIGURE 10 | Pairwise causal score (8) during spring for the five studied variables: precipitation, discharge, snowmelt, temperature, and evapotranspiration. Positive scores (in blue) highlight significant causal link suggested, negative scores (in red) represent the inverse causal link suggested, and dark grey areas show no significant causal link. Yellow catchments correspond to those removed.

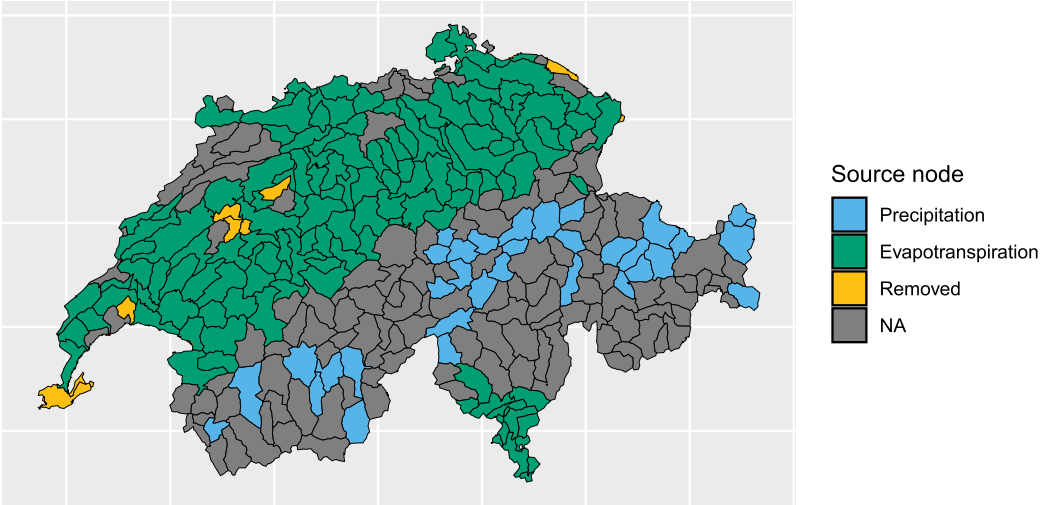

FIGURE 11 | Source node based on ordering the Wasserstein distance (7) among the five variables and based on majority vote among 500 resampling calculations. Dark grey highlights undefined majority vote and yellow catchments correspond to those removed.

events are particularly significant because rainwater has a higher density than snow, so it can infiltrate the snowpack more easily and increase the rate of melt. The red areas highlight situations where snowmelt and temperature cause precipitation. This happens at low elevations in the Plateau in Switzerland (negative scores) and is primarily due to atmospheric processes and topographical features. When warm, moist air masses move across the region, they are forced to rise as they encounter mountain ranges. This process is known as orographic lifting. This phenomenon is an example of how atmospheric and geographical factors interact spatially under extreme condition. Interestingly, the two bottom rows suggest that springtime evapotranspiration in Switzerland is primarily influenced by internal land–atmosphere interactions, rather than by precipitation, snowmelt, or temperature anomalies.

By ordering the Wasserstein distance (7) among the five variables, we can discover the source node over Switzerland under extreme condition. Contrarily to Gnecco et al. (2021) where the EASE algorithm requires a sequential source node discovering, our method recovers the causal ordering in one step. Our transformed variables are all on the same scale and comparable as the term $\max(U_1, \ldots, U_d)$ of (4) appears in the expression of all transformed margins. Figure 11 shows the dominating source nodes over the catchments. To assess uncertainty of our methodology, data are bootstrapped 500 times and we retain a variable as a source node only when this is the case at least 95% of the time. In accordance with findings in Figure 10, the contrast between high-relief and low ones underscores how orography flips the dominant source node between precipitation in the mountains and evapotranspiration on the Plateau. In the Alps and Pre-Alps, extreme events are primarily initiated by precipitation surges, whereas on the Plateau and Jura they usually stem from evapotranspiration losses. Again, few small basins with anomalous evapotranspiration patterns were excluded, likely reflecting local model artifacts or unique land-cover effects.

6 | Discussion

The Earth system is composed of many intricate linkages that make causal discovery a challenging task. This work addresses discovery of causal links at extreme levels between hydrological and (hydro)meteorological variables in different catchments in Switzerland. Relying on the multivariate extreme value theory, we propose a unifying definition of causality at extreme levels. This definition does not rely on interventionist notions and is unifying in the sense that it encompasses tail causality induced by the restricted model class of structural equations models for extremes, that is, with heavy-tailed errors, as well as tail causality induced by asymmetric strong tail dependencies. One advantage of our approach is the use of the asymptotically-motivated multivariate Generalised Pareto distribution to properly rescale the random vector of interest. This rescaling, in conjunction with the asymmetric nature of tail causality, whereby extreme events in the cause lead to extreme events in the effect, but not necessarily vice versa, facilitates the identification of the direction of causality when possible.

When applied to simulated variables in a network of catchments in Switzerland, our methodology unveiled tail causal relationships between precipitation, discharge, and snowmelt in high-elevation catchments, as well as a notable interaction between these hydrological variables and evapotranspiration in the Plateau. While the ground truth under extreme conditions remains somewhat unclear, our findings are partially aligned with the expected outcomes derived from physical laws. This emphasizes the need for expert knowledge when it comes to causal notions. For instance, while identifiable outcomes from causal discovery should be examined by domain experts or validated by laws of nature, unidentifiable outcomes might point towards data-related issues. Although not its primary use, our methodology can thus be used to detect limitations in meteorological simulations where the ground truth is dictated by physical laws.

The asymptotic independence (AI) case was not treated in this work and an extension of the score defined by (8) to this case will be valuable. A conceptual roadblock to extend Proposition 1 is the current lack of additive MGPD representations like (6) for the AI case. Further research is needed to define such models and then propose and study associated AI causal scores.

Acknowledgments

The authors would like to thank Massimiliano Zappa from the Swiss Federal Institute for Forest, Snow, and Landscape Research WSL for kindly providing the simulated (hydro)meteorological data used in this analysis. Open access publishing facilitated by Ecole polytechnique federale de Lausanne, as part of the Wiley - Ecole polytechnique federale de Lausanne agreement via the Consortium Of Swiss Academic Libraries.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Endnotes

[1] Throughout, operations involving vectors are to be interpreted componentwise.

[2] Results presented in this analysis were not sensitive to this choice of the length.

References

Brunner, M. I., B. Hingray, M. Zappa, and A.-C. Favre. 2019. "Future Trends in the Interdependence Between Flood Peaks and Volumes: Hydro-Climatological Drivers and Uncertainty." *Water Resources Research* 55: 4745–4759. https://doi.org/10.1029/20/19WR024701.

Chavez-Demoulin, V., and L. Mhalla. 2024. Causality and Extremes. https://arxiv.org/abs/2403.05331.

Davison, A. C., and R. L. Smith. 1990. "Models for Exceedances Over High Thresholds." *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 52: 393–442.

Deidda, C., S. Engelke, and C. De Michele. 2023. "Asymmetric Dependence in Hydrological Extremes." *Water Resources Research* 59: e2023WR034512. https://doi.org/10.1029/2023WR034512.

Dettinger, M. D. 2011. "Climate Change, Atmospheric Rivers, and Floods in California–A Multimodel Analysis of Storm Frequency and Magnitude Changes." *Journal of the American Water Resources Association* 47: 514–523.

Engelke, S., and A. S. Hitz. 2020. "Graphical Models for Extremes (with discussion)." *Journal of the Royal Statistical Society. Series B, Statistical Methodology* 82: 871–932.

Fatichi, S., M. J. Zeeman, J. Fuhrer, and P. Burlando. 2015. "An Overview of Current Applications, Challenges, and Future Trends in Distributed Process-Based Models in Hydrology." *Journal of Hydrology* 529: 849–859.

Ferreira, A., and L. de Haan. 2014. "The Generalized Pareto Process; With a View Towards Application and Simulation." *Bernoulli* 20: 1717–1737.

Froidevaux, P., J. Schwanbeck, R. Weingartner, C. Chevalier, and O. Martius. 2015. "Flood Triggering in Switzerland: The Role of Daily to Monthly Preceding Precipitation." *Hydrology and Earth System Sciences* 19: 3903–3924.

Gissibl, N., and C. Klüppelberg. 2018. "Max-Linear Models on Directed Acyclic Graphs." *Bernoulli* 24: 2693–2720. https://doi.org/10.3150/17-BEJ941.

Gnecco, N., N. Meinshausen, J. Peters, and S. Engelke. 2021. "Causal Discovery in Heavy-Tailed Models." *Annals of Statistics* 49: 1755–1778.

Kiriliouk, A., H. Rootzen, J. Segers, and J. L. Wadsworth. 2018. "Peaks Over Thresholds Modeling With Multivariate Generalized Pareto Distributions." *Technometrics* 61: 123–135. https://doi.org/10.1080/00401706.2018.1462738.

Monge, G. 1781. *Mémoire Sur la Théorie Des Déblais et Des Remblais*. De l'Imprimerie Royale.

Naveau, P., A. Hannart, and A. Ribes. 2020. "Statistical Methods for Extreme Event Attribution in Climate Science." *Annual Review of Statistics and Its Application* 7: 89–110.

Naveau, P., and J. Segers. 2024. Multivariate Extreme Value Theory. https://arxiv.org/abs/2412.18477.

Papastathopoulos, I., and K. Strokorb. 2016. "Conditional Independence Among Max-Stable Laws." *Statistics & Probability Letters* 108: 9–15. https://www.sciencedirect.com/science/article/pii/S0167715215002874.

Peters, J., and P. Bühlmann. 2015. "Structural Intervention Distance for Evaluating Causal Graphs." *Neural Computation* 27, no. 3: 771–799. https://doi.org/10.1162/NECO_a_00708.

Rootzén, H., J. Segers, and J. Wadsworth. 2018. "Multivariate Peaks Over Thresholds Models." *Extremes* 21: 115–145.

Rootzén, H., and N. Tajvidi. 2006. "Multivariate Generalized Pareto Distributions." *Bernoulli* 12: 917–930.

Schirmer, M., and T. Jonas. 2023. Enhance Existing Swiss Precipitation Products With Particular Regards to Snowfall, Technical Report, Meteoswiss. https://www.meteoswiss.admin.ch/dam/jcr:75a32bf0-c634-476f-a667-7dc3e23a3c8c/Final-report-Enhance-existing-Swiss-precipitation-products-with-particular-regards-to-snowfall.pdf.

Tawn, J. A. 1990. "Modelling Multivariate Extreme Value Distributions." *Biometrika* 77: 245–253.

Villani, C. 2008. *Optimal Transport: Old and New*, Grundlehren der Mathematischen Wissenschaften. Springer Berlin Heidelberg. https://books.google.ch/books?id=hV8o5R7_5tkC.

Viviroli, D., M. Zappa, J. Gurtz, and R. Weingartner. 2009. "An Introduction to the Hydrological Modelling System PREVAH and Its Pre- and Post-Processing-Tools." *Environmental Modelling & Software* 24: 1209–1222.

Zappa, M., and M. Brunner. 2019. Hydro-Meteorological Simulations for the Period 1981–2018 for Switzerland. https://www.envidat.ch/dataset/hydro-meteorological-simulations-1981-2018.

Appendix A

Proof of Proposition 1

Proof. Based on the definition of the Wasserstein distance, one can find an optimal coupling $\pi^* = (E', X_i', \pi)$ such that

$$W_1(X_i, E) = \mathrm{E}_\pi |E' - X_i'|$$

where $E' \overset{d}{=} E$ and $X_i' \overset{d}{=} X_i$. We now rely on the representation (4) and the fact that E stochastically dominates X_i, that is, $F_{X_i}(t) \geq F_E(t), \forall t \in \mathbb{R}$ to show that

$$t = F_{X_i}^{-1}\{F_{X_i}(t)\} \geq F_{X_i}^{-1}\{F_E(t)\}$$

and hence that $X_i' = F_{X_i}^{-1}\{F_E(E')\} \leq E'$. Thus, we have showed that the optimal coupling π^* satisfies the following property

$$X_i' = F_{X_i}^{-1}\{F_E(E)\} \leq E'$$

and that $W_1(X_i, E) = \mathrm{E}_\pi(E' - X_i')$. The desired equivalence is then straightforward

$$\begin{aligned} W_1(X_i, E) \leq W_1(X_j, E) &\Leftrightarrow W_1(X_i, E) - W_1(X_j, E) \\ &= \mathrm{E}(X_j) - \mathrm{E}(X_i) \leq 0 \end{aligned}$$

□