Modeling non-stationary extreme dependence with stationary max-stable processes and multidimensional scaling

Clément Chevalier , Olivia Martius & David Ginsbourger

Modeling non-stationary extreme dependence with stationary max-stable processes and multidimensional scaling

Clément Chevalier[*]

Institute of Statistics, University of Neuchâtel

Oeschger Centre for Climate Change Research, University of Bern

and

Olivia Martius

Institute of Geography and Oeschger Centre for Climate Change Research, University of Bern

and

David Ginsbourger

Institute of Mathematical Statistics and Actuarial Science and Oeschger Centre for Climate Change Research, University of Bern

Corresponding author Clément Chevalier clement.chevalier@unine.ch

Abstract

Modeling the joint distribution of extreme events at multiple locations is a challenging task with important applications. In this study, we use max-stable models to study extreme daily precipitation events in Switzerland. The non-stationarity of the spatial process at hand involves important challenges, which are often dealt with by using a stationary model in a so-called climate space, with well-chosen covariates. Here, we instead choose to warp the weather stations under study in a latent space of higher dimension using multidimensional scaling (MDS). Two methods are proposed to define target dissimilarity matrices, based respectively on extremal coefficients and on pairwise likelihoods. Results suggest that the proposed methods allow capturing complex spatial dependences of spatial extreme precipitations, enabling in turn to reliably extrapolate functionals such as extremal coefficients.

1 Introduction

Understanding the joint distribution of extremes at different locations is a challenging problem with important stakes. In particular, extreme rainfall can result in damage due to extensive overland flow and cause natural disasters such as mudslides (Guzzetti et al., 2008) or floods (Froidevaux et al., 2015). These rainfall events typically affect wider areas and are spatially dependent.

Spatial models for extremes (see, Davison et al. (2019) for a recent review) have received particular attention in the last few years, through the use of max-stable process modeling (De Haan, 1984; Smith, 1990; Schlather, 2002; Kabluchko et al., 2009). Although statistical inference is known to be challenging (Bacro and Gaetan, 2014), max-stable processes have been used for spatial modeling of extreme temperatures (Davison and Gholamrezaee, 2012), winds (Engelke et al., 2015), precipitation (Smith and Stephenson, 2009; Padoan et al., 2010; Huser and Davison, 2014; Shang et al., 2015) and snow depths (Blanchet and Davison, 2011; Gaume et al., 2013). The theory of max-stable processes generalizes (Ribatet, 2013) the now well-established univariate

extreme value theory (Coles, 2001). Although other approaches exist for constructing spatial models for extremes (see, Davison et al. (2012) for a review), this work focuses on the joint distributions of extremes through max-stable processes. Of the possible choices for max-stable processes, we chose to mainly focus on Brown-Resnick models (Kabluchko et al., 2009; Davison et al., 2012).

Max-stable models are often calibrated in a so-called climate space (see, Blanchet and Davison (2011) and references therein), which is a parametric transformation of the space of spatial coordinates with additional well-chosen covariates. This approach is used in, e.g., Cooley et al. (2007); Blanchet and Davison (2011) and has been shown to improve the fits compared to models that are merely calibrated in the 2 or 3-dimensional space of spatial coordinates. The use of additional covariates can be seen as a solution to deal with *non-stationarity*. Indeed, the dependence between extremes often cannot be seen as a simple function of the Euclidean distance between the stations in the (longitude, latitude) space or the (longitude, latitude, elevation) space, even if some geometrical anisotropy is added. Frei and Schär (1998) and Fukutome et al. (2015) have shown that the climatological spatial distribution of both mean and extreme precipitation in Switzerland is not a simple function of height and location because of slope and shielding effects. An example event is presented in Rossler et al. (2014), where the amount of precipitation falling on the north- and south-facing slope of a valley differed by a factor of three due to local circulation effects.

Overall, it appears that max-stable process modelling is at the same time enjoying solid theoretical foundations and posing severe implementation challenges. The fact that likelihoods are typically intractable, and that surrogate criteria such as pairwise log-likelihoods are numerically optimized instead, constitutes a substantial limitation. Furthermore, even when augmenting the geographical space using the climate space approach of Blanchet and Davison (2011), the model expressiveness may be insufficient to capture

complex dependency patterns such as reflected by extremal coefficients estimated from data by means of an F-madogram. This can be seen on the upper panel of Figure 1, where the so-called classical model delivers extremal coefficient isolines (upper right graph) that strongly differ from the ones estimated by F-madogram and interpolated (upper left graph). Note that while this example focuses on one arbitrary reference station (Aarberg), 219 stations are considered in our study, as precised below. Our main contribution is to introduce two methods that leverage multi-dimensional scaling and specificities of the Brown-Resnick model in order to better reproduce, and eventually extrapolate, complex dependency patterns such as illustrated here. The bottom panel of Figure 1 displays the extremal coefficient maps returned by our methods.

In our two proposed methods, we choose to deal with non-stationarity by using a stationary model in a different space. Instead of a climate space however, we rely on a latent space built non-parametrically in which the different stations under study are warped. We propose two different model fitting approaches, both of which use multidimensional scaling (MDS) to compute the coordinates of the stations in the latent space, yet with two different rationales regarding the target dissimilarity matrix. In the first method, the target dissimilarity matrix is constructed based on estimated extremal coefficients and on links between extremal coefficients, covariances, and distances under a Brown-Resnick model with a family of isotropic kernels monotonically decreasing with respect to Euclidean distance. A similar setting is employed in method 2, yet with target covariance values steeming from considerations on pairwise log-likelihoods rather than inherited from estimated extremal coefficients.

Max-stable models and MDS are both well-established concepts. The novelty of this work lies in the way we use MDS to produce new non-stationary max-stable model fitting approaches. In *non-extreme* spatial statistics, MDS has notably been used by Sampson and Guttorp (1992) to build a spatio-temporal kriging model. For the spatial part of their model, the dimension of the latent space built

with MDS was set to 2, allowing the creation of a map that warps any location of the input space to the latent space. More recently, Bornn et al. (2012) suggested dealing with non-stationarity by using a stationary model in an expanded space, which they defined as the product of the initial input space and a 1- or 2- dimensional space built with MDS. In *extreme* spatial statistics, non-stationarity has notably been addressed in Asadi et al. (2015) where the authors build a parametric hydrological distance which better assess dependencies in extreme river discharges than the Euclidean distance. Also, Huser and Genton (2016) dealt with non-stationarity using parametric transformations of stationary covariance functions, which might also be interpreted as a space transformation. Our approach share similarities with this idea, except that we do not transform an initial space, distance, or covariance function. Instead, we directly build latent spaces using MDS – a somewhat non parametric approach – and use standard distance and covariance functions. The calibration of the models presented in this work does not require to know the spatial coordinates or other covariates of the stations, as opposed to the other methods. However spatial coordinates are required in the additional procedure that we propose for bridging the gap between MDS and warping models defined over the whole domain of interest.

Throughout the paper we use an application test case pertaining to the study of extreme precipitations in Switzerland. In the Alpine area, extreme precipitation events are known to be more intense than in other parts of Europe (Frei and Schär, 1998). In Switzerland, extreme rainfall and associated flooding can have substantial socio-economic consequences (Hilker et al., 2009). A significant fraction of extreme daily rainfall events in Switzerland is associated with synoptic-scale weather systems (Martius et al., 2006; Pfahl et al., 2014; Giannakaki and Martius, 2016). Our dataset is provided by the Swiss Federal Office for Meteorology and Climatology (MeteoSwiss) and consists of daily measurements of rainfall from January 1st 1961 to December 31st 2013. We study the 219 stations that have no missing data. From the available daily precipitations we compute blockwise maxima over the autumn season, i.e. from 21 September to

20 December of each year. We focus on this season because daily precipitation is highest in summer and autumn in northern Switzerland (Umbricht et al., 2013; Giannakaki and Martius, 2016) and in autumn in southern Switzerland (Panziera et al., 2018). The blockwise maxima are henceforth referred to as yearly maxima. In this study, the number of stations in the dataset (219) is larger than in the previously cited references (at most approx. 100 stations in Blanchet and Davison (2011)).

The article is organized as follows: Section 2 provides a brief background in spatial extreme value statistics, and presents a short overview of MDS methods. Section 3 details our two different methods for fitting max-stable models using MDS and discusses parameter estimation. In Section 4, we compare the obtained models to a more classical max-stable model fitted in a (longitude, latitude, elevation) space with geometrical anisotropy and space rotation. In particular, we compare the ability of the models to reproduce the observed non-stationary dependencies. Finally, we discuss the advantages and drawbacks of the proposed approaches. For completeness, a short background in univariate extreme values statistics and more details on MDS methods are given as supplementary material, together with some details on the tested "classical" max-stable model. Further diagnostic plots and R code are also provided in the supplementary material.

2 Background

2.1 Max stable models and extremal coefficients

Max-stable processes are well-suited to model block maxima (Davison et al., 2012). Formally, a random process $(Y(x))_{x \in \mathbb{X}}$ is max-stable with unit Fréchet margins if the random process $(NY(x))_{x \in \mathbb{X}}$ has the same distribution as $(\max_{i=1,...,N} Y_i(x))_{x \in \mathbb{X}}$, where $N > 0$ and $Y_1,...,Y_N$ are N i.i.d copies of Y. A widely used representation is given in De Haan (1984), where a max-stable process $(Z^\star(x))_{x \in \mathbb{X}}$ is represented as follows:

$$Z^\star(x) := \max_{i \in \mathbb{N}} \eta_i W_i^\star(x), \qquad (1)$$

where the η_i's are drawn from a Poisson process on $\mathbb{R}_+$ with intensity $z^{-2}\mathrm{d}z$ and $W_1^\star, W_2^\star, \ldots$ are i.i.d copies of a non-negative random process $W^\star$ satisfying $\mathbb{E}W^\star(x) = 1$ for all $x \in \mathbb{X}$. Several choices of processes $W^\star$ lead to different models. Here, we focus on the geometric Gaussian process, $W^\star = \exp(\sigma W - \sigma^2 / 2)$ where $\sigma > 0$ and W is a centered Gaussian process with unit pointwise variance. This model will be referred to as the "Brown-Resnick model" even-though we do not impose a stationarity or intrinsic-stationarity condition on W.

For a max-stable process $Z^\star$ with unit Fréchet margins and for an arbitrary set of locations $\mathcal{D}_{1:N} := (x_1, \ldots, x_N)$, the finite-dimensional cumulative distributions of $Z^\star$ can be written as

$$\mathbb{P}(Z^\star(x_1) \le z_1, \ldots, Z^\star(x_N) \le z_N) = \exp(-V_{1:N}(z_1, \ldots, z_N)), \qquad (2)$$

for some function $V_{1:N} : (z_1, \ldots, z_N) \in \mathbb{R}^N \mapsto \mathbb{R}$ called the exponent function (Resnick, 1987) which satisfies $V_{1:N}(tz_1, \ldots, tz_N) = t^{-1}V_{1:N}(z_1, \ldots, z_N), t > 0$ and $V_{1:N}(z, \infty, \ldots, \infty) = z^{-1}$ for any permutation of its arguments. If we set $z_1 = \ldots = z_N := z$, Equation (2) can be rewritten as

$$\mathbb{P}(Z^\star(x_1) \le z, \ldots, Z^\star(x_N) \le z) = \exp(-V_{1:N}(1, \ldots, 1) / z)$$
$$:= \exp(-\theta_{1:N} / z),$$

where $\theta_{1:N}$ is the so-called extremal coefficient associated with the set of locations $\mathcal{D}_{1:N}$. Practitioners are often interested in the pairwise extremal coefficients θ_{ij} between stations x_i and x_j. For a Brown-Resnick model with parameter $\sigma > 0$, this extremal coefficient is given by

$$\theta_{ij} = 2\Phi\left(\sqrt{\frac{\sigma^2}{2}(1 - k(x_i, x_j))}\right), \qquad (3)$$

where k is the covariance function of the centered Gaussian process W and Φ is the cumulative distribution function (c.d.f.) of the standard normal distribution. Throughout this paper, all covariances $k(x_i, x_j)$ are also correlations, since W has unit variance.

2.2 Pairwise likelihood maximization and climate space

Let T denote the number of years of data, n the number of stations, and $z_{it}^\star$ the rescaled maximum precipitation at station x_i during year t (see supplementary material Section 1 for details on the data transformations to have unit Fréchet margins). The pairwise log-likelihood of some max-stable model with parameters γ is

$$\ell(\gamma) = \sum_{1 \leq i < j \leq n} \sum_{t=1}^{T} \log f_{ij}(z_{it}^\star, z_{jt}^\star; \gamma) := \sum_{1 \leq i < j \leq n} \ell_{ij} \quad (4)$$

where f_{ij} is the bivariate density of the random vector $(Z^\star(x_i), Z^\star(x_j))$ and $Z^\star$ is a max-stable process with unit Fréchet margins. In the particular case of Brown-Resnick models with parameter σ and covariance function k, f_{ij} is obtained by differentiating the following bivariate c.d.f. (see Davis et al. (2013) for detailed calculations):

$$F_{ij}(z_i, z_j) = \exp\left[-\left[\frac{1}{z_i}\Phi\left(v_{ij} + \frac{\log(z_j/z_i)}{2v_{ij}} \right) + \frac{1}{z_j}\Phi\left(v_{ij} + \frac{\log(z_i/z_j)}{2v_{ij}} \right) \right] \right], \quad (5)$$

where $v_{ij} = \sqrt{\sigma^2(1 - k(x_i, x_j))/2}$. The latter equation shows that the contribution ℓ_{ij} of the pair (i, j) of stations to the pairwise likelihood of Equation (4) is directly linked to the covariance $k(x_i, x_j)$ and hence to the parameters of the covariance functions and other parameters accounting for space transformation. An example is given in Blanchet and Davison (2011) where γ is a set of parameters that allows space transformation, this model is refered to as the "classical model" in this paper and is shortly described in the supplementary material, Section 2. The

models proposed in this work also rely on the notion of climate space, with the important difference that the space will be somehow built non-parametrically using Multidimensional Scaling (MDS).

2.3 Multidimensional scaling

Let D be a given $n \times n$ dissimilarity matrix and let $d > 0$ be an integer. Multidimensional scaling (MDS) techniques aim at finding a configuration of n points in $\mathbb{R}^d$ in such a way that the obtained $n \times n$ interpoint Euclidean distance matrix is as close as possible to D, with respect to some stress function. In many applications, MDS is used to provide graphical displays of n points, which ease the interpretation of an $n \times n$ distance or dissimilarity matrix D (Borg and Groenen, 2005). For visualization, the dimension d is thus often set to 2 or 3, although the algorithms usually remain valid for larger values of d. Different choices of stress functions, as well as different algorithms to minimize the stress, yield a large set of MDS techniques. In this work, we use the MDS algorithm of Sammon (1969) which minimizes the so called raw-stress using iterative gradient-descent in dimension nd. Details on this algorithm as well as on other MDS techniques are given in the supplementary material, Section 3.

As explained in the next section, we manage in this work to transform the Brown-Resnick model fitting problem into a MDS problem. In particular, for a given parameter choice γ, we obtain a dissimilarity matrix D and ultimately a placement of the station in a space of dimension d which yields a given goodness of fit (or "loss") that depends on γ. In our models, γ is merely a vector composed of the Brown-Resnick parameter σ and some covariance parameters of the Gaussian Process W, see Section 2.1. The dimension d of the latent space could actually be part of the vector of parameters γ. Instead, we choose $d = d^\star(\gamma)$ using the Bayesian Information Criterion (BIC) of Oh and Raftery (2001). The BIC has the advantage to be applicable to any MDS algorithm and to bring a solution to potential overfitting problems when d becomes too large.

3 Proposed model fitting approaches

We now present our two approaches that both consist in leveraging MDS in order to allow efficient non-stationary Brown-Resnick modelling via non-parametric warping. While both proposed methods aim at increasing the realism of the fitted model by capturing spatial non-stationarity reflected by sampled block maxima, they differ by the employed criteria of fit. In the first case (method 1), model parameters are tuned with the aim to best reproduce extremal coefficients estimated by F-madogram (Cooley et al., 2006), yet with a custom measure of fit that translates the problem on extremal coefficients into a problem on dissimilarity matrices which can be elegantly tackled by using off-the-shelf MDS algorithms. Following a different route, although based as well on MDS with a target dissimilarity matrix inherited from data assumed to have been generated by the Brown-Resnick model, method 2 relies instead on dissimilarities inherited from consideration on pairwise log-likelihoods. The crux of this second method is to remark indeed that to each pair of locations in the Brown-Resnick model corresponds an ideal covariance value that maximizes the corresponding contribution to the pairwise log-likelihood criterion. Assuming like method 1 an isotropic covariance function, written as a monotonically decreasing function of the underlying distance, method 2 hence also boils down to an MDS problem with a specific target dissimilarity matrix. One further issue in order to bring the applicability of these methods to a next level is to be able to warp arbitrary locations into the latent space induced by MDS, and not only the station locations used in the algorithm. For this we appeal to Gaussian Process interpolation (kriging) in the latent space, resulting in continuous formulations of the proposed non-stationary covariance functions. This is presented in Section 3.2. Furthermore, the estimation of model parameters such as σ and covariance parameters are adressed in Section 3.3 (the latent dimension d being taken care of via the employed kind of MDS procedure). Let us first focus in the next section on the way target covariance and distance matrices are obtained in the two proposed methods, respectively.

3.1 From data to target distance matrices

3.1.1 Method 1, based on fitting (estimated) pairwise extremal coefficients

The first introduced method adresses the issue of poorly reconstructed pairwise extremal coefficients (See Figure 1) by directly taking some misfit between model and reference extremal coefficients as an objective function, such as the " Extremal Coefficients Mean Squared Error" (ECMSE):

$$\text{ECMSE}(\theta) := \frac{1}{n(n-1)} \sum_{1 \leq i, j \leq n} (\theta_{ij} - \theta_{ij}^{\text{ref}})^2, \qquad (6)$$

where the θ_{ij}^{ref}'s are reference (estimated or prescribed) values for the pairwise extremal coefficients and $\theta = (\theta_{ij})$ form the extremal coefficient between stations x_i and x_j under the considered Brown-Resnick model. Such a criterion is looking at model fit from a very specific angle and one should keep aware of potential overfitting pitfalls (see for instance Brehmer et al. (2019) for theoretical results on why scoring functions cannot assess tail properties). Nevertheless, we will see in our numerical experiments that this first method turns out to not only deliver a non-stationary generative model with improved fitting of observed extreme precipitations but to also offer promising performances in predicting extremal coefficients involving unobserved locations. Futhermore, the latter is found to perform surprisingly better than the baseline in terms of pairwise log-likelihood. The rationale of method 1 is as follows. Assume first for expository purpose that perfect (noiseless) values of the pairwise extremal coefficients at locations $\mathbb{X} = (x_1, \ldots, x_n)$ (say the set of all $n = 219$ considered climatological stations) are given, so that θ^{ref} is assumed to stem from the Brown-Resnick model, yet with covariances to be determined. Recalling Equation 3, one can then write $\theta^{\text{ref}} = g(K)$ with $K = k(x_i, x_j)$ and $g(K)$ is the result of applying g entry-wise to K where

$$g : \rho \in [-1,1] \rightarrow g(\rho) = 2\Phi\left(\sqrt{\frac{\sigma^2}{2}(1-\rho)}\right) \in \left[0, 2\Phi\left(\frac{\sigma^2}{2}\right)\right]. \quad (7)$$

As *g* is a bijective mapping, one can hence express *K* in function of θ^{ref} as $K = g^{-1}(\theta^{\text{ref}})$. If furthermore *K* writes as the chaining between a function $k_o : [0,\infty) \rightarrow [-1,1]$ and some distance *d* on $\mathbb{X}$, then $K = k_o(D)$ for *D* the distance matrix with generic entry $d(x_i, x_j)$ $(i, j \in \{1,\dots,n\})$, so that in the end $\theta^{\text{ref}} = g(k_o(D))$. Assuming finally that k_o is positive and stricly decreasing, we have that $D = k_o^{-1}(g^{-1}(\theta^{\text{ref}}))$. In other words, for a broad class of models, and assuming known σ and k_o, there is a one-to-one correspondence between pairwise extremal coefficients and distances between pairs of stations. Method 1 relies on this correspondence in order to find a parametrization of stations in a higher-dimensional latent space so that the corresponding distance matrix be as close as possible to $k_o^{-1}(g^{-1}(\theta^{\text{ref}}))$. In practice, however, $\theta^{\text{ref}} = \hat{\theta}$ is the result of some estimation procedure (here using the F-madogram estimator of Cooley et al. (2006)), so that this workflow requires a few adaptations in order to be implementable. First, the inverse $\theta \rightarrow g^{-1}(\theta) = 1 - \frac{2}{\sigma^2}\left(\Phi^{-1}\left(\frac{\theta}{2}\right)\right)^2$ is extended to $[0,2]$, using the convention $\Phi^{-1}(1) = \infty$. Furthermore, since we work with covariance functions delivering strictly positive values, we floor the "target" pairwise covariances $g^{-1}(\theta_{i,j})$ with a minimum value $\varepsilon > 0$, leading to a target dissimilarity matrix

$$D^{\star(1)} = k_o^{-1}\left(\max\left(\varepsilon, g^{-1}(\hat{\theta})\right)\right). \quad (8)$$

The value of ε could be a parameter of our actual model, but for simplicity we fix it at $\exp(-3) \approx 0.05$. Note that the matrix $\tilde{K}^{(1)} := \max\left(\varepsilon, g^{-1}(\hat{\theta})\right)$ is not a covariance matrix because it has no reason to be positive definite. This is not a problem in practice, but there are two important consequences. First, it means that a *perfect* fit of the extremal coefficients $\hat{\theta}$ estimated from the data will not be possible. Second, it involves that the so-called "target dissimilarity matrix" will not always

be Euclidean, which prevents using classical scaling, as explained in the supplementary material, Section 3. Other MDS techniques which are better suited to non-Euclidean matrices are used. More specifically, we choose to use the algorithm of Sammon (1969) that relies on nd-dimensional gradient descent to minimize a given stress function. Our experiments suggested that the latter was giving better results than other MDS techniques.

3.1.2 Method 2, based on pairwise likelihood maximization

Our second approach shares some essential features with the first one, in the sense that it involves a target dissimilarity matrix derived from observed data. Yet it does not rely on the extremal coefficients nor on the function g. Instead, it directly prescribes a target for K, and the latter is obtained via considerations on the pairwise log-likelihood criterion. Consider indeed again as before a Brown-Resnick process with parameter σ, and focus for the moment on an arbitrary pair of distinct stations (x_i, x_j). The contribution of the Fréchet-transformed data from this pair of stations to the pairwise log-likelihood is given by the term ℓ_{ij} in Equation (4) and depends on the covariance or correlation value $K_{ij} = k(x_i, x_j)$ (see Equation (5) and the expression of v_{ij}).

When σ is fixed, it is possible to plot ℓ_{ij} values as a function of K_{ij} and see where ℓ_{ij} is maximized. An example is given in Figure 2 where ℓ_{ij} is plotted for 3 different pairs of stations (x_i, x_j). For each pair of stations, there exists an ideal value $K_{ij}^{\star}$ that maximizes ℓ_{ij}. In our work, the search for this ideal value is performed on the interval $[0, 0.99]$ to avoid the numerical issues caused by correlations of 1. The target covariance matrix $\tilde{K}^{\star(2)}$ is obtained by flooring the $K_{ij}^{\star}$'s with the same minimum value $\varepsilon > 0$ as in the previous method. The target dissimilarity to be fed into MDS is then defined analogously by

$$D^{\star(2)} = k_o^{-1}\left(\tilde{K}^{\star(2)}\right) = k_o^{-1}\left(\max\left(\varepsilon, K^{\star}\right)\right). \qquad (9)$$

The target dissimilarity matrices of Equation (8) and (9) depend on σ and on the covariance function k_o. The choice of these parameters is discussed in Section 3.3.

3.2 From MDS to warping models over the whole space

When the parameters of the proposed model are fixed a target dissimilarity matrix can be calculated (see, Section 3.1), and the stations are then placed in a latent space of dimension d in order to have Euclidean distances close to these dissimarities. Let $\mathbf{x_1},\ldots,\mathbf{x_n}$ be the coordinates of the n stations in the 3-dimensional space (longitude, latitude, elevation) and let $\mathbf{y_1},\ldots,\mathbf{y_n}$ be the coordinates of these stations in the latent d-dimensional space, obtained using MDS. Further, denote by ψ the MDS mapping, i.e. $\psi(\mathbf{x_i}) = \mathbf{y_i}, i = 1,\ldots,n$. The fitted Brown-Resnick process lives in the latent space of dimension d so that, at this stage, we do not have a *spatial* model yet. Indeed, the coordinates in the latent space of a station or location that was not used to fit the model are unknown. To adress this issue, we need a mapping ψ which is known everywhere, for any location $\mathbf{x}$ in the (longitude, latitude, elevation) space.

Following <u>Sampson and Guttorp (1992)</u> and <u>Borg and Groenen</u> (<u>2005</u>), this mapping is explicitly constructed everywhere using interpolation. The MDS mapping to be interpolated goes from $\mathbb{R}^3$ to $\mathbb{R}^d$. As many methods exist to interpolate *real-valued* functions, we then simply build d interpolators, each one being an Ordinary Kriging model (see <u>Roustant et al.</u> (<u>2012</u>) for a short description) associated to one of the coordinates in $\mathbb{R}^d$. For all i, we use the notation $\mathbf{y_i} = (y_{i,1},\ldots,y_{i,d})$, and $\psi(\mathbf{x}) := (\psi_1(\mathbf{x}),\ldots,\psi_d(\mathbf{x}))$. The ith kriging interpolators is trained from observations at locations $\mathbf{x_1},\ldots,\mathbf{x_n}$ and responses $y_{i,1},\ldots,y_{i,n}$. It enables to obtain the ith coordinate in the latent space of any location in Switzerland. These d kriging models are computed using the DiceKriging R package (<u>Roustant et al.</u>, <u>2012</u>) and with an anisotropic exponential covariance functions. The parameters of the covariance functions are estimated using maximum likelihood. An example of interpolation with d = 4

is given in Figure 3. Any location in Switzerland here has 4 coordinates in the latent space obtained with 4 kriging models. Each coordinate in the latent space can be represented with a map of Switzerland.

3.3 Parameter selection

In the previous section, MDS is performed assuming that the covariance function k of the Gaussian process W is known, and assuming that the Brown-Resnick parameter σ and the latent space dimension d are known. In practice, all of these parameters need to be estimated.

We first discuss the choice of the covariance function and, when applicable, of the parameters of this function. For our MDS application, among the family of stationary covariance functions, it is sufficient to consider only isotropic ones. Indeed, the use of range parameters (sometimes called correlation lengths), that would account for geometrical anisotropy in the latent space, would simply yield a configuration of points X_d with coordinates divided by the corresponding range parameter. In other words, all the models that we can produce with anisotropic covariance functions k can also be obtained with isotropic ones. In this work, the considered covariance function is the powered-exponential covariance:

$$k(h) := \exp(-h^{\alpha}), \alpha \in (0, 2]. \quad (10)$$

This covariance function can be used in spaces of arbitrary dimension d, a property that is mandatory for the application of our method. Many covariance functions (e.g., spherical, circular, cubic, Gneiting) are omitted here because they cannot be used in arbitrary dimensions. Experiments with the Matérn($\nu = 3/2$) and Matérn($\nu = 5/2$) covariance function have also been conducted, but underperformed the powered exponential covariance function. The Matérn(ν) covariance, with varying smoothness ν, might outperform the powered exponential covariance. However, this has not been investigated further because computing and inverting a large number of times the Matérn(ν) covariance for

arbitrary ν is too computer intensive. See also Section 4.4 for a detailed study on computation time.

We hence need to estimate 2 parameters: the Brown-Resnick parameter σ and the covariance parameter α in Equation (10). When these parameter are fixed, we mentionned in Section 2.3 that the MDS dimension d can be chosen automatically using the Bayesian Information Criterion of Oh and Raftery (2001). In addition to this BIC we impose a maximum value of 6 for d. This is motivated by the results of Perrin and Meiring (2003), which indicate that a non-stationary field in space of dimension ν can be represented as second-order stationary in dimension 2ν. This result applied to a non-stationary random field in the (longitude, latitude, elevation) space yields a maximum latent space dimension of $d = 6$.

As the choice of d is now automatically embedded in the MDS procedure, selection of (σ, α) can simply be performed by optimizing a well-chosen criterion. For method 1, we choose (σ, α) by minimizing the ECMSE of Equation (6). For method 2, we maximize the pairwise likelihood of Equation (4). The optimization of the criterion could be performed with a numerical optimization algorithm, but the objective function can be highly multimodal, specially given that the chosen MDS dimension d varies with (σ, α). Hence, given that we only have 2 parameters, we simply evaluate the criterion on a grid of (σ, α) values in $[2, 4] \times [1, 2]$. The solution has the advantage to be simple and to allow parallel computing.

4 Results and comparisons

4.1 Introduction: competing models

In this section, we compare the max-stable models obtained using the two proposed model fitting methods with a more classical model fitted in a climate space, following Blanchet and Davison (2011). The classical model uses the

space (longitude, latitude, elevation) and a parametric space deformation. The parameters of the model are the Brown-Resnick parameter σ, the angle β and geometrical anisotropy parameters c_1, c_2, c_3. In addition, we use the isotropic powered exponential covariance function in the climate space, which has an exponent parameter α. The 6 parameters of the model are estimated by maximizing the pairwise-likelihood obtained with the data from all $n = 219$ stations. As in Blanchet and Davison (2011), we maximize the pairwise likelihood using a simple line-search algorithm. The likelihood is sequentially maximized with respect to one parameter while keeping the other parameters fixed, and the procedure is iterated until convergence. For the classical model, we considered adding additional covariates to the climate space, like the mean precipitation. However, the additional covariates did not improve the obtained likelihoods. The parameters of the fitted models, which are compared in the next sections, are summarized in Table 1.

4.2 Extremal coefficients fits and pairwise log-likelihoods

The fitted max-stable models should be able to reproduce the pairwise extremal coefficients estimated from the data. They should further give good overall pairwise log-likelihoods. In our experiments, both methods 1 and 2 yield better ECMSE than the classical model, as well as better pairwise likelihoods. To determine the significance of these results and to challenge the proposed methods, we decided to perform 50 additional random experiments. In each random experiment, n_2 test stations are not used to fit our models and thus the models are fitted based on only $n_1 := n - n_2$ training stations, in the spirit of a cross-validation analysis. The obtained ECMSE and pairwise likelihoods for *all* training and tests stations are then obtained by using the MDS mapping for the n_1 training stations and by interpolating this mapping for the n_2 test stations (see Section 3.2). In the random experiments, n_2 is chosen uniformly in the set of integers between 25 and 50. The n_2 test stations are then chosen using the

space-filling algorithm of the BalancedSampling R package (Grafström, 2016; Grafström et al., 2012).

Figure 4 gives the obtained pairwise log-likelihoods (left plot) and ECMSE (right plot) for our full models (methods 1 and 2 using all $n = 219$ stations), the classical model, and all 50 random experiments. The dashed curve indicates the pairwise likelihood or ECMSE of the classical model. The triangles indicate the pairwise likelihood or ECMSE of the proposed full models, and the boxplots indicate the results of the 50 random experiments. With method 1, 100% of the random experiment yield a better ECMSE than the classical model, even though the classical model is fitted based on all n stations. With method 2 (optimizing the pairwise likelihood) 100% of the random experiments yield a better pairwise likelihood than the classical model and – more surprisingly – they also always outperform it with respect to the ECMSE. Based on these experiments, we conclude on the significance of the improvements with respect to the classical model.

Next Section shows that the other advantage of the flexible MDS-based model fitting is its ability to reproduce complex dependence structures.

4.3 Extremal coefficients maps

We show in this section that the main advantage of the proposed model fitting methods is their natural ability to handle complex non-stationary dependencies. Figure 1 in the introduction shows a map of pairwise extremal coefficients between a given station (Aarberg) and any other point in Switzerland. The first map is computed from the extremal coefficients estimated from the data, using F-madogram. For visualization only, this map requires a spatial interpolation, which is performed here using Ordinary Kriging in the (longitude, latitude, elevation) space, with an anisotropic exponential covariance function. The other three maps are computed from the three competing models (classical model, method 1, method 2). One clearly sees that methods 1 and 2 better reproduce the strong

non-stationarity observed in the data. For many stations in Switzerland, these map of extremal coefficients cannot be well reproduced with ellipses in the (longitude, latitude, elevation) space, as assumed by the classical model. Method 1 aims at fitting extremal coefficients and thus the nice fit observed with the data is not a surprise. However, notice that method 2 does *not* aim at fitting extremal coefficients and relies instead on pairwise likelihood. The good performance of method 2 with respect to extremal coefficients and its ability to reproduce a complex non-stationary dependence structure is thus very encouraging.

The maps in Figure 1 can be produced for all n = 219 stations, and all 219 pictures are released as supplementary material in a file which can be easily downloaded. In Figure 5, we produce these maps for the 4 stations after Aarberg, in alphabetical order. Notice that for the Adelboden station ("ABO", i.e. the 4 maps at the top left), method 1 is able to reproduce complex dependencies, where the set of locations with extremal coefficients lower than 1.6 is not a connected set. The same phenomenon can be observed for many other stations. Additional extremal coefficients maps are presented and commented in the Supplementary Material. In addition, we produce diagnostic plots quantifying dependencies between more than 2 stations.

4.4 Computation time

The use of Sammon's MDS algorithm on matrices of size $n \times n$ is relatively fast. With a standard computer with 3.40Ghz cpu and 8Gb of RAM, Sammon's algorithm takes a bit less than 0.1 second for the considered number of stations n = 219. However, in the fitting procedure of Section 3, MDS is used a large number of times for different values of the parameters α, σ, d. To estimate the parameters, we optimize a criterion on a grid of (σ, α) values of size 21 × 21. For each (σ, α), an MDS is performed with dimension $d = 2, d = 3$ and continues to at most d = 6 as long as the BIC of Oh and Raftery (2001) recommends to increase d. Fitting a model with method 1 or 2 thus requires at most $21^2 \times 5 \approx 2'200$ calls to the Sammon MDS algorithm, which takes a bit less than 4 minutes.

Computationnal costs with method 2 are larger than with method 1 for two reasons. First computing a pairwise likelihood is slower than computing an ECMSE. Second, when (σ, α) are fixed, computing the target dissimilarity matrix $D^{\star(2)}$ (see, Equation (9)) given to the MDS is not straightforward. This dissimilarity matrix actually depends on the matrix $K^\star$ containing "optimal" pairwise covariances, which depends on σ, but not on α nor d. As we use a grid of size 21 for σ we then simply store all these 21 matrices. This initial step does not involve any MDS. Storing these 21 matrices takes approximately 20 minutes for method 2. For method 1, no storage is performed.

Notice that in Section 4.2 we perform 50 random experiments by removing n_2 test stations from the dataset. To avoid multiplying the previous computation times by 50, the value of α is "profiled" based on the full model. In other words for a given σ we automatically choose $\alpha^\star := \mathrm{argmin}_\alpha \mathrm{Crit}(\sigma, \alpha)$ where Crit is the criterion computed for the full model. Hence, in these 50 random experiments, optimizations are performed only with respect to σ, at the cost of at most 21 × 5 calls to the MDS algorithm instead of $21^2 \times 5$ calls.

Compared to the proposed methods, fitting the classical model requires the maximization of a pairwise likelihood with respect to six parameters, and takes approximately 4 minutes. The improvement provided by the proposed models thus comes at the price of higher computational costs. The computational cost could be reduced by using faster procedures to optimize with respect to (σ, α). Our methodology could be used with a larger number n of stations since Sammon's MDS mapping is still relatively fast. The main limiting factor is the diagonalization of an $n \times n$ matrix, which is needed to obtain an initial placement of the stations through classical scaling, before using Sammon's gradient-descent.

5 Summary and discussion

5.1 The full process: using these methods with your own data

We provide a summary of the proposed techniques, with the goal of allowing practitioners to use the proposed methods with their own dataset. The Supplementary Material indeed presents some R code and data which can be downloaded and which enables to reproduce every result presented in this paper. The code is also meant to be used with new datasets. Here we discuss the different steps involved to fit a full model, based on all stations, for methods 1 and 2.

5.1.1 Model fitting

The *only* input required to fit a model with one of the two presented methods is the Fréchet transformed block-maxima, stored as a $n \times T$ matrix, where n is the number of stations and T the number of years of data. Both methods 1 and 2 will then estimate the parameter (σ, α) by calculating a loss function, the "criterion", which depends on these parameter and by selecting $(\sigma^\star, \alpha^\star)$ optimizing the criterion. Before detailing the equations applied to compute the criterion for a given (σ, α) value, a preliminary step needs to be computed, for Method 2 only. Indeed, $n \times n$ matrices containing pairwise optimal covariances need to be stored for every possible value of σ, which represents 21 matrices to store with our grid. For a given σ, and for a pair (i, j) of stations, the optimal pairwise covariance between the stations is obtained by maximizing the pairwise log-likelihood of stations i, j alone, by playing with the covariance value $k(x_i, x_j) \in [0,1]$, see Figure 2. Pairwise log-likelihoods are computed by differentiating the c.d.f. of Equation (5) and depend of $k(x_i, x_j)$ because v_{ij} in Equation (5) depends on $k(x_i, x_j)$.

Now, for a given (σ, α) value the criterion is computed as follows. For Method 1, we compute a target dissimilarity matrix $D(\sigma, \alpha)$ by applying Equation (8) to the extremal coefficients estimated from the data, using F-madogram. We then use MDS and choose the dimension $d^\star(\sigma, \alpha)$ suggested by the BIC of Oh and Raftery (2001). Once the MDS is run, the n stations have coordinates in the latent space from which we compute a Euclidean distance matrix $\hat{D}(\sigma, \alpha)$. Then,

from the (σ, α) values, we compute a covariance matrix $k(\hat{D}(\sigma, \alpha))$ and modeled extremal coefficients $g(k(\hat{D}(\sigma, \alpha)))$ where g is given by Equation (7). The modeled extremal coefficients are compared to the ones estimated from the data, which gives our ECMSE, see Equation (6).

For method 2, the target dissimilarity matrix $D(\sigma, \alpha)$ is obtained by applying the function k^{-1}, which depends on α, to the stored matrix with optimal pairwise covariances, see Equation (9). The MDS procedure is then the same than with Method 1. Once the stations have coordinates in the latent space of dimension $d^{\star}(\sigma, \alpha)$, we have a Brown-Resnick model in the latent space with fixed σ and known covariance function k. A pairwise likelihood of the data is calculated. This criterion depends on (σ, α) and is used to choose these parameters.

5.1.2 Warping any location in Switzerland to the latent space

Once a model is fitted with either method 1 or 2, one can quantify dependencies **only** between locations that were used to fit the model, i.e. between stations where data are recorded. This is sufficient to compute the criteria (ECMSE, pairwise likelihoods) described previously. However, if one is interested in dependencies between **any** pair of location in Switzerland (e.g. to build extremal coefficients maps) then it is necessary to have the coordinates of all these locations in the latent space. In this case, the MDS mapping needs to be interpolated, see Section 3.2. In addition to the Fréchet-transformed data, this interpolation requires the following inputs:

- The coordinates (longitude, latitude, elevation) of the stations used to fit the model. We remind that this information was NOT needed to fit the model with either method 1 or 2.
- The coordinates (longitude, latitude, elevation) of the pixels that need to be mapped in the latent space. In our application, we have a total of $77'256$ pixels in Switzerland.

- The coordinates of the stations used to fit the model in the latent space of dimension d, obtained with MDS.

Interpolating the MDS mapping consists in interpolation a mapping from $\mathbb{R}^3$ to $\mathbb{R}^d$. As explained in Section 3.2, this can be easily done by interpolating d mappings from $\mathbb{R}^3$ to $\mathbb{R}$. Many interpolation methods could be used here. In this work we used the Ordinary Kriging models of the DiceKriging R package. In these models, we used an anisotropic exponential covariance function with parameters estimated by maximum likelihood, see Figure 3 to have 4 examples of these interpolated $\mathbb{R}^3$ to $\mathbb{R}$ mappings. Once the interpolation is performed, all the selected pixels have coordinates in the latent space and the computation of dependencies is possible. Extremal coefficients maps between one fixed location and any location in Switzerland (see, Figures 1 and 5) can be computed from Euclidean distances and by applying $(g°k)$ to the distances; i.e. by transforming distances into modeled extremal coefficients.

5.2 Capturing full independence: unbounded variograms

In this work, we chose to build max-stable models from **stationary** log-Gaussian processes. A drawback of this approach is the impossibility to reproduce full independence between precipitations at any pair of stations, even if the considered stations are far away in the latent space. For example, with method 1 we obtain the value $\sigma = 2.9$ which means that the model cannot reproduce extremal coefficients larger than $2\Phi(\sigma/\sqrt{2}) \approx 1.96$, see Equation 3. To overcome this drawback, the two model-fitting methods presented in this paper could be applied to models involving an intrinsic-stationary Gaussian process with unbounded variogram. Indeed, if $\gamma(\cdot)$ is the semivariogram of the considered Gaussian process, the modeled extremal coefficient between two locations with distance h is $2\Phi\left(\sqrt{\dfrac{\gamma(h)}{2}}\right)$ and can converge to 2 if γ is unbounded.

In that case, the proposed methods can be applied provided that the chosen semivariogram can be "inverted", meaning that it is possible to compute a target dissimilarity matrix which would reproduce – with this model – the extremal coefficients estimated from the data. A parametric semivariogram of the form $\gamma(h) = \| h \|^{\alpha}, 0 < \alpha \le 2$ (fractional Brownian motion) satisfies this invertibility condition. Extensions like the Generalized Cauchy model of Gneiting and Schlather (2004) could also be applied. The performances of models with unbounded variograms has not been investigated in this work and is an avenue for future research.

5.3 Uncertainty quantification

The models presented in this work involve parameters – namely α, σ, d – which are estimated from the data. Classical pairwise-likelihood maximization in a climate space enables to estimate variances for the model parameters (see, e.g. Blanchet and Davison (2011)). However, with the presented models, this cannot straightforwardly be reproduced as the criterion which is optimized involves an MDS which is itself performed with a numerical optimization algorithm. Uncertainty quantification is thus challenging. To quantify uncertainties on the parameters σ, α, d fitted with methods 1 and 2, we here performed a parametric bootstrap as follows. For each method, we generate a set of 50 datasets using max-stable process simulation. Each dataset consist on 53 years of block-maximum precipitations, for 219 stations. The method to simulate max-stable processes is decribed in the Supplementary Material. For Method 1, the max-stable model lives in a latent space of dimension d = 5 and has parameters $\sigma = 2.9, \alpha = 2$. For method 2, we have $d = 6, \sigma = 2.7, \alpha = 2$, see Table 1. For all 50 datasets generated from the "method 1" model, we refit a max-stable model based on method 1. We do the same for method 2.

The results of these experiments are summarized in Table 2 where mean values and standard-deviation (in bracket) over the 50 experiments are computed for every parameter. For method 1, we notice that the obtained latent space

dimension is always of 5 or 6, with approx. equal probability. Method 1 performs quite well at recovering the parameters σ, α used to generate the datasets. On the other hand method 2 chooses more often $d = 6$ and, rarely $d = 5$. Method 2 seems to have more variability with respect to the estimates of σ, α. In addition, method 2 here tends to overestimate the true value of σ (here: 2.7). Overall, the results of these bootstraped experiments seem more satisfying with method 1.

5.4 Conclusion and additional perspectives

In this work, we introduce new max-stable model fitting procedures that rely on multidimensional scaling (MDS). Instead of relying on a stationary model in a climate space, obtained with some parametric space deformation of the (longitude, latitude, elevation) space, we use a stationary model in a latent space built with MDS. Compared to more classical approaches, the proposed methods are better able to reproduce non-stationary spatial dependencies, as shown by the extremal coefficients map in Figure 1. User-friendly commented R code is available to let practitioners apply the presented work on their own data.

To complement the possible future research avenues mentioned in the previous sections, we add that one could also implement a tailored MDS algorithm where one could directly play with the coordinates X_d of the n stations in the latent space to directly maximize a log-likelihood or minimize an ECMSE, without simplying the problem through the computation of a target dissimilarity matrix. One of the main challenges for such a procedure would be its computation time, since MDS is currently performed many times the parameters of the model.

References

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

Bacro, J.-N. and Gaetan, C. (2014), "Estimation of spatial max-stable models using threshold exceedances," *Statistics and Computing,* 24, 651–662.

Blanchet, J. and Davison, A. C. (2011), "Spatial modeling of extreme snow depth, " *The Annals of Applied Statistics,* 5, 1699–1725.

Borg, I. and Groenen, P. J. (2005), *Modern Multidimensional Scaling: Theory and Applications,* Springer.

Bornn, L., Shaddick, G., and Zidek, J. (2012), "Modeling nonstationary processes through dimension expansion," *Journal of the American Statistical Association,* 107, 281–289.

Brehmer, J. R., Strokorb, K., et al. (2019), "Why scoring functions cannot assess tail properties," *Electronic Journal of Statistics,* 13, 4015–4034.

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

Cooley, D., Naveau, P., and Poncet, P. (2006), "Variograms for spatial max-stable random fields," in *Dependence in Probability and Statistics,* eds. P. Bertail, P. Doukhan, and P. Soulier, Springer, New York, volume 187 of *Lecture Notes in Statistics,* pp. 373–390.

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

Davis, R. A., Klüppelberg, C., and Steinkohl, C. (2013), "Statistical inference for max-stable processes in space and time," *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* 75, 791–819.

Davison, A. and Gholamrezaee, M. (2012), "Geostatistics of extremes," *Proceedings of the Royal Society A,* 468, 581–608.

Davison, A., Padoan, S., and Ribatet, M. (2012), "Statistical modeling of spatial extremes," *Statistical Science,* 27, 161–186.

Davison, A. C., Huser, R., and Thibaud, E. (2019), "Spatial extremes," in *Handbook of Environmental and Ecological Statistics,* eds. A. E. Gelfand, M. Fuentes, J. A. Hoeting, and R. L. Smith, CRC Press, pp. 711–744.

De Haan, L. (1984), "A spectral representation for max-stable processes," *The Annals of Probability,* 12, 1194–1204.

Engelke, S., Malinowski, A., Kabluchko, Z., and Schlather, M. (2015), "Estimation of Hüsler–Reiss distributions and Brown–Resnick processes," *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* 77, 239–265.

Frei, C. and Schär, C. (1998), "A precipitation climatology of the Alps from high-resolution rain-gauge observations," *International Journal of Climatology,* 18, 873–900.

Froidevaux, P., Schwanbeck, J., Weingartner, R., Chevalier, C., and Martius, O. (2015), "Flood triggering in Switzerland: the role of daily to monthly preceding precipitation," *Hydrology and Earth System Sciences,* 19, 3903–3924.

Fukutome, S., Liniger, M. A., and Süveges, M. (2015), "Automatic threshold and run parameter selection: a climatology for extreme hourly precipitation in Switzerland," *Theoretical and Applied Climatology,* 120, 403–416.

Gaume, J., Eckert, N., Chambon, G., Naaim, M., and Bel, L. (2013), "Mapping extreme snowfalls in the French Alps using max-stable processes," *Water Resources Research,* 49, 1079–1098.

Giannakaki, P. and Martius, O. (2016), "Synoptic-scale flow structures associated with extreme precipitation events in northern Switzerland," *International Journal of Climatology,* 36, 2497–2515.

Gneiting, T. and Schlather, M. (2004), "Stochastic models that separate fractal dimension and the hurst effect," *SIAM review,* 46, 269–282.

Grafström, A. (2016), "Balancedsampling: Balanced and spatially balanced sampling. R package version 1.5.1." URL `http://www.antongrafstrom.se/balancedsampling`.

Grafström, A., Lundström, N. L., and Schelin, L. (2012), "Spatially balanced sampling through the pivotal method," *Biometrics,* 68, 514–520.

Guzzetti, F., Peruccacci, S., Rossi, M., and Stark, C. P. (2008), "The rainfall intensity-duration control of shallow landslides and debris flows: an update," *Landslides,* 5, 3–17.

Hilker, N., Badoux, A., and Hegg, C. (2009), "The Swiss flood and landslide damage database 1972-2007," *Natural Hazards and Earth System Sciences,* 9, 913–925.

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

Huser, R. and Genton, M. G. (2016), "Non-stationary dependence structures for spatial extremes," *Journal of Agricultural, Biological, and Environmental Statistics,* 21, 470–491.

Kabluchko, Z., Schlather, M., and De Haan, L. (2009), "Stationary max-stable fields associated to negative definite functions," *The Annals of Probability,* 37, 2042–2065.

Martius, O., Zenklusen, E., Schwierz, C., and Davies, H. C. (2006), "Episodes of alpine heavy precipitation with an overlying elongated stratospheric intrusion: A climatology," *International Journal of Climatology,* 26, 1149–1164.

Oh, M.-S. and Raftery, A. E. (2001), "Bayesian Multidimensional Scaling and Choice of Dimension," *Journal of the American Statistical Association,* 96, 1031–1044.

Padoan, S., Ribatet, M., and Sisson, S. (2010), "Likelihood-based inference for max-stable processes," *Journal of the American Statistical Association,* 105, 263–277.

Panziera, L., Gabella, M., Germann, U., and Martins, O. (2018), "A 12-year radar-based climatology of daily and sub-daily extreme precipitation over the Swiss Alps," *International Journal of Climatology,* 38, 3749–3769.

Perrin, O. and Meiring, W. (2003), "Nonstationarity in $\mathbb{R}^n$ is second-order stationarity in $\mathbb{R}^{2n}$," *Journal of applied probability,* 40, 815–820.

Pfahl, S., Madonna, E., Boettcher, M., Joos, H., and Wernli, H. (2014), "Warm Conveyor Belts in the ERA-Interim Dataset (1979-2010). Part II: Moisture Origin and Relevance for Precipitation," *Journal of Climate,* 27, 27–40.

Resnick, S. I. (1987), *Extreme Values, Regular Variation, and Point Processes,* Springer, New York.

Ribatet, M. (2013), "Spatial extremes: Max-stable processes at work," *Journal de la Société Française de Statistique,* 154, 156–177.

Rossler, O., Froidevaux, P., Borst, U., Rickli, R., Martius, O., and Weingartner, R. (2014), "Retrospective analysis of a nonforecasted rain-on-snow flood in the Alps - a matter of model limitations or unpredictable nature?" *Hydrology and Earth System Sciences,* 18, 2265–2285.

Roustant, O., Ginsbourger, D., and Deville, Y. (2012), "Dicekriging, Diceoptim: Two R packages for the analysis of computer experiments by kriging-based

metamodelling and optimization," *Journal of Statistical Software,* 51, URL
`http://www.jstatsoft.org/v51/i01`.

Sammon, J. W. (1969), "A nonlinear mapping for data structure analysis," *IEEE Transactions on computers,* 18, 401–409.

Sampson, P. D. and Guttorp, P. (1992), "Nonparametric estimation of nonstationary spatial covariance structure," *Journal of the American Statistical Association,* 87, 108–119.

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

Shang, H., Yan, J., and Zhang, X. (2015), "A two-step approach to model precipitation extremes in California based on max-stable and marginal point processes," *The Annals of Applied Statistics,* 9, 452–473.

Smith, E. and Stephenson, A. (2009), "An extended Gaussian max-stable process model for spatial extremes." *Journal of Statistical Planning and Inference,* 139, 1266–1275.

Smith, R. (1990), "Max-stable processes and spatial extremes. Unpublished manuscript.".

Umbricht, A., Fukutome, S., Liniger, M., Frei, C., and Appenzeller, C. (2013), " Seasonal variation of daily extreme precipitation in Switzerland," Technical report, MeteoSwiss.

SUPPLEMENTARY MATERIALS

Supplementary Material: A short background in univariate extreme values statistics and more details on MDS methods are provided, together with some details on the tested "classical" max-stable model of Blanchet and Davison (2011). Additional extremal coefficients maps are commented. We also

provide a link to download the Fréchet-transformed dataset used in this work, as well as commented R code, allowing one to easily reproduce all of the results and figures in this article. A link for downloading all 219 extremal coefficients maps is also provided. Finally additional diagnostic plots for the proposed models are given. (PDF file)
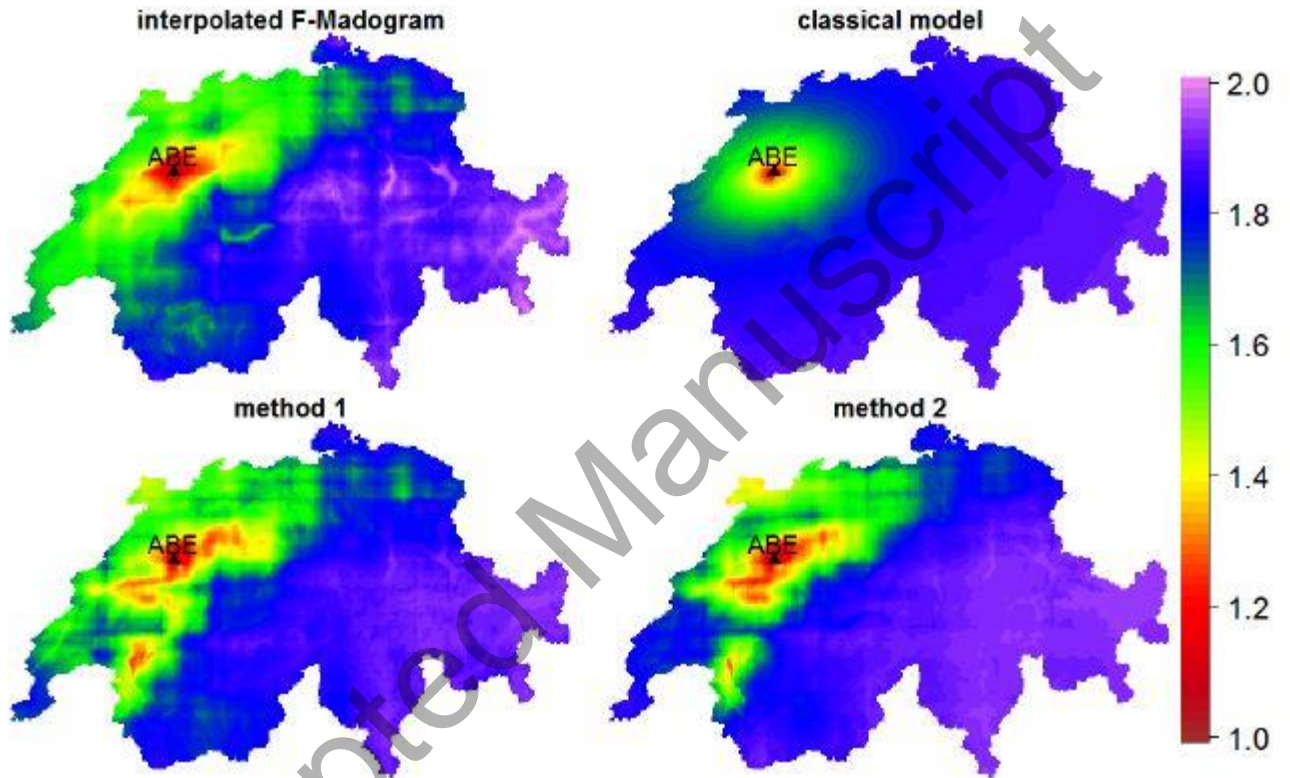

Fig. 1 Maps of pairwise extremal coefficients between the Aarberg station and other points in Switzerland. The top left map is obtained from the observed extremal coefficients (estimated with the F-madogram estimator) and with interpolation via kriging. The top right map represents extremal coefficients returned by the model of Blanchet and Davison (2011) while the two bottom maps represents those returned by our two proposed methods.
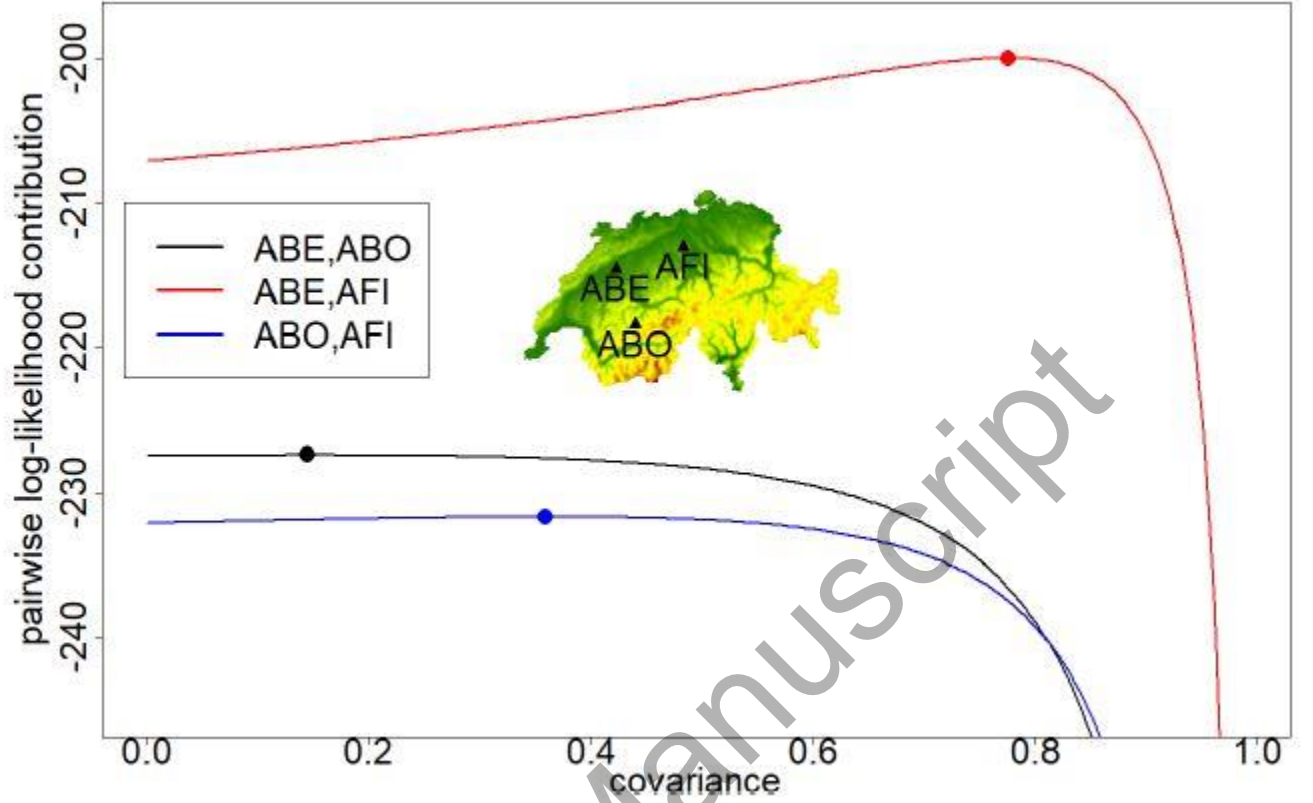
Fig. 2 Contribution ℓ_{ij} of 3 different pairs of stations to the pairwise log-likelihood as a function of their covariance K_{ij}, using $\sigma = 2.4$. The locations of the 3 stations forming the 3 pairs are indicated on the map. For each pair of stations, the covariance maximizing ℓ_{ij} is indicated with a dot. Since the covariances are positive and are also correlations, they are in the interval $[0,1]$.
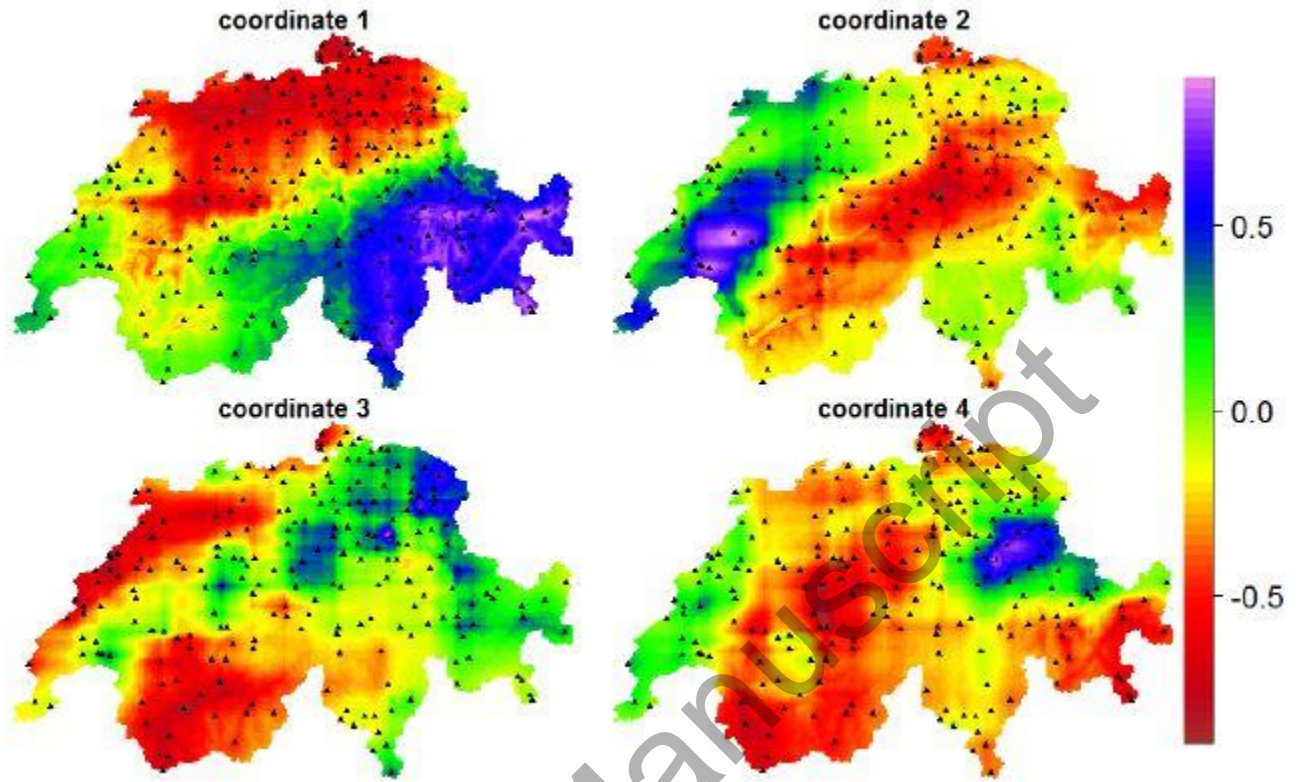
Fig. 3 Ordinary Kriging predictors $\psi_1^*, \ldots, \psi_4^*$ obtained by interpolating the MDS mapping. Here, we fitted the model using method 1, $\sigma = 2.8$. The covariance function of the Gaussian process W (see, section 2.1) is a stationary isotropic powered exponential covariance function k with exponent $\alpha = 2$.
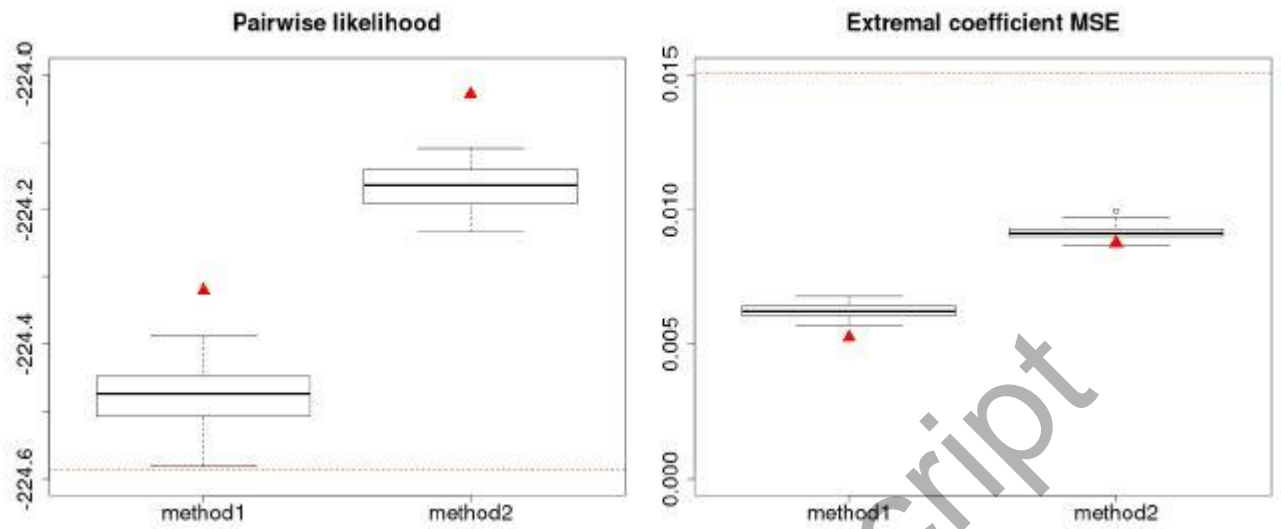
Fig. 4 Pairwise log-likelihoods (left plot) and ECMSE (right plot) obtained using the proposed full models (red triangles) and the classical model (red dashed curve). The boxplots indicate the results of 50 random experiments in which n_2 stations are randomly selected and are not used to fit the proposed models.
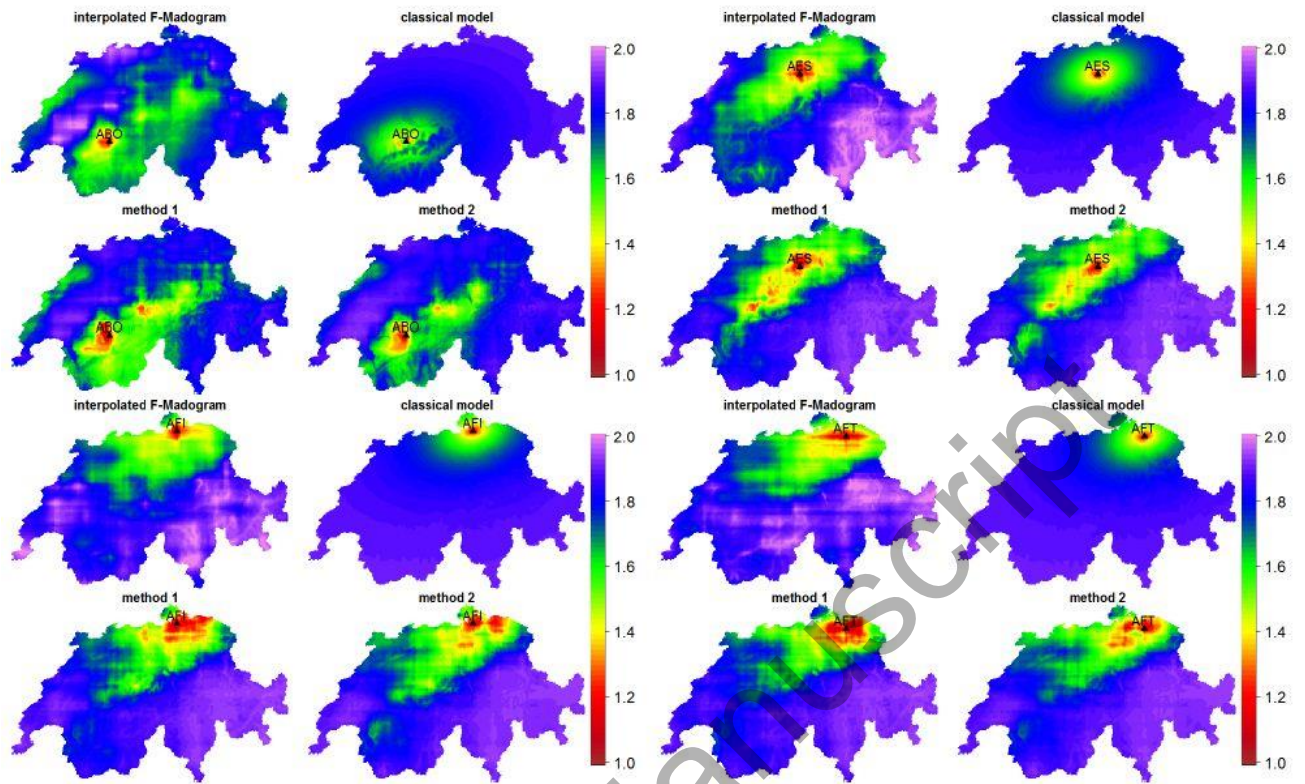
Fig. 5 Maps of pairwise extremal coefficients, similar to those in Figure 1. The 4 maps are plotted for the Adelboden (4 maps at the top left), Aesch (top right), Andelfingen (bottom left) and Affeltrangen (bottom right) stations.

Table 1 Description of the three fitted models that are compared. All of these models are Brown-Resnick models with an isotropic (in the latent or climate space) powered exponential covariance function with parameter α. For the classical model, the estimated space-deformation parameters are $(c_1, c_2, c_3, \beta) \approx (0.70, 1.52, 3.3e-4, -0.11)$, see Equation 2 in the supplementary material.

	method 1	method 2	classical model
latent/climate space dimension d	5	6	3
Brown-Resnick parameter σ	2.9	2.7	2.37
Powered exponential exponent α	2	2	1.09

Table 2 Mean and standard deviation (bracket) of the parameters d, σ, α, computed over 50 random experiments. In each experiment, a new dataset is generated from an initial model, using max-stable process simulation. For method 1, the parameters used to simulate the max-stable process are $d = 5, \sigma = 2.9, \alpha = 2$. For method 2, they are $d = 6, \sigma = 2.7, \alpha = 2$.

	method 1	method 2
latent space dimension d	5.54(0.5)	5.92(0.27)
Brown-Resnick parameter σ	2.91(0.27)	3.07(0.51)
Powered exponential exponent α	2.00(0.03)	1.91(0.13)