Regionalization of the extremal dependence structure using spectral clustering

Véronique Maume-Deschamps, Pierre Ribereau, Manal Zeidan

Regionalization of the extremal dependence structure using spectral clustering

Véronique MAUME-DESCHAMPS[1], Pierre RIBEREAU[1] and Manal ZEIDAN[1,2]

[1]Universite Claude Bernard Lyon 1, ICJ UMR5208, CNRS, Ecole Centrale de Lyon, INSA Lyon, Université Jean Monnet, 69622 Villeurbanne, France.

[2]Department of Operation Research and Intelligent techniques, University of Mosul, Mosul,Iraq.

Contributing authors: veronique.maume@univ-lyon1.fr; pierre.ribereau@univ-lyon1.fr; manal.zeidan@univ-lyon1.fr;

Abstract

The influence of an extreme event depends on the geographical features of the region where the event occurs. In order to understand the behavior of an extreme event, we consider statistical models capable of capturing the extremes and their spatial dependence. Max-stable processes are widely used in studying extreme events. However, assuming a fixed extremal dependence for a max-stable process may not be reasonable, depending on the topology of the region under study. In environmental extreme events, different types of extremal dependencies can appear across the spatial domain. In this study, we present an adapted spectral clustering algorithm for max-stable processes. This algorithm combines spectral clustering with extremal concurrence probability to cluster locations into k regions, each with an homogeneous extremal dependence. Additionally, we propose an approach to model the entire region based on the clustered zones. In order to validate the proposed methodology, we tested it in two simulation cases using a non-stationary max-stable mixture model. The accuracy of the results encouraged us to apply it to two datasets: rainfall data on the east coast of Australia and rainfall over France.

Keywords: Max-stable processes, Extremal dependence, Extremal concurrence probability, Spectral clustering

1 Introduction

Constructing a statistical model for environmental extreme events, such as rainfall, temperature, and so on, is very important for understanding their behavior and accurately predicting their occurrence. Max-stable processes are natural models for spatial extremes, as they are natural extensions of the Extreme Value Theory (EVT) to spatial domains. They are powerful statistical models for extreme events in a continuous space and can assess the risk in areas that do not have stations. One common assumption used for max-stable models is that the extremal dependence is fixed across spatial domains. However, this assumption may be incorrect and can lead to the construction of meaningless models, particularly when the data sets are taken from a large region or regions with complex geographical features (Richards and Wadsworth (2021)). For instance, consider the rainfall data from the east coast of Australia, a region characterized by mountain ranges. These mountains are known to significantly influence rainfall events and can give rise to complex rainfall patterns. Similarly, when considering rainfall data in France, which features extensive coastal areas, extreme rainfall can be produced by disturbances from both the Mediterranean Sea and the Atlantic Ocean.

Researching a max-stable model capable of capturing changing extremal dependence is a recent topic of study. For instance, Huser and Genton (2016) developed an approach that captures non-stationary patterns in spatial extremes using covariates. They developed a non-stationary extremal-t model by combining the extremal-t model with a non-stationary correlation function. This model is satisfactory in capturing the extremal dependence. Huser and Genton (2016) presented another non-stationary max-stable model without using it in their paper. This model is defined as a max-stable mixture, and the details of this model are presented in Section 5. In short, this model is a mixture of two max-stable models with a proportion of mixing that varies spatially. This proportion can be modeled as a function depending on covariates. This model could capture different extremal dependence in different spatial regions. However, it is computationally difficult and requires prior knowledge of relevant covariates.

Recently, clustering was used to create regionalisations of extreme events. Clustering is an unsupervised machine learning tool that is widely used in data analysis to identify subgroups with similar features. It has a wide range of applications in fields such as computer science, statistics, biology, and climate science.

In the context of spatial extremes, only a few studies use clustering to partition an entire region into homogeneous sub-regions based on similarities in extremal dependence. For instance, Bernard et al (2013) presented a novel clustering algorithm for maxima, using the F-madogram introduced by Cooley et al (2006). By combining the F-madogram with a partitioning around medoids (PAM) algorithm, they were able to cluster the extremes based on dependence strength. The algorithm was applied to analyze rainfall patterns over France. Afterward, Bador et al (2015) applied this algorithm to large

regions and different variables, analyzing the maxima of summer temperatures across Europe. Saunders et al (2021) demonstrated that the PAM algorithm presented by Bernard et al (2013) is sensitive to stations density. To address this issue, they proposed the use of hierarchical clustering with F-madogram. They applied their proposed algorithm to rainfall stations in Australia. Then, for each of the resulting regions, they fitted a stationary max-stable model. In this case, its not possible to model the dependence between two locations that belong to different clusters. However, in all of these studies, there is no clear idea about which model for the entire region can be estimated using these clusters.

Our contribution in this paper has two paths. Firstly, we proposed a clustering approach for max-stable processes. We adapted spectral clustering for max-stable processes by combining it with the extremal concurrence probability introduced by Dombry et al (2018). The extremal concurrence probability for a max-stable process is the probability that the maximum value of the process occurs at two or more locations simultaneously. This clustering algorithm aims to identify regions with similar extremal dependence. We demonstrated the applicability of this algorithm in clustering the max-stable mixture process presented by Huser and Genton (2016) into k regional clusters, each with a similar extremal dependence. Additionally, we considered different sizes of blocks and discussed the effect of the block size on the clustering maps.

The second path involves using the clustered regions in modeling the entire area. To the best of our knowledge, this paper is the first to address this. For this purpose, we proposed a censored pairwise likelihood based on clusters. We validated our clustering and estimation approaches through a simulation study and then applied them to two datasets. The first dataset consists of rainfall data from the east coast of Australia, while the second dataset includes rainfall data from France provided by Météo-France.

The paper is organized as follows. An overview of spectral clustering is provided in Section 2. Section 3 presents Max-stable processes. Section 4 describes the adapted spectral clustering for max stable process. Section 5 presents the applicability of the adapted spectral clustering . Section 6 presents the inference using composite liklihood approach. A simulation study is presented in Section 7. Section 8 applies the methodology to data: rainfall over east coast of Australia and rainfall over France. Finally, Section 9 presents the discussion and conclusions of our study.

2 Spectral clustering : an overview

Spectral clustering is a technique used in machine learning and data analysis for clustering data points into groups based on the similarity between them. It is based on the concept of spectral graph theory, which is the study of the properties of graphs using linear algebra.

Spectral clustering has several advantages, as it can handle high-dimensional data, which is often a limitation for other clustering algorithms. This is done by

reducing the high-dimensional data to a lower-dimensional space using eigenvalue decomposition. Furthermore, it can handle different kinds of similarity measures, which makes it flexible and adaptable to different types of data. Also, it does not make any assumptions on the shape or size of clusters.
Spectral clustering considers the dataset as a graph, where each data point $P_i, i = 1, \cdots, N$ represents a vertex in an undirected weighted graph. An undirected graph $G = (V, E, W)$ is generally defined by a set of vertices $V = \{v_1, v_2, \cdots, v_n\}$, a set of edges $E = \{(v_i, v_j)|v_i, v_j \in V\}$ between these vertices, and a similarity matrix W. An element $w_{ij} \in W$ represents the amount of similarity between the vertices v_i,v_j and the weight that will be assigned to each edge. It is important to note that since the graph is undirected, the similarity matrix should be symmetric. If $w_{ij} = 0$, this means that there is no edge between the vertices v_i,v_j. Each vertex v_i in the graph has a degree d_i:

$$d_i = \sum_{j=1}^{N} w_{ij}. \tag{1}$$

The degrees $d_1, \cdots, d_N$ represent the elements of a diagonal matrix called the degree matrix of the graph D.
Spectral clustering aims to separate the main graph G into sub-graphs so that the weights of the edges between these sub-graphs are small, indicating dissimilarity between the clusters, while the weights of the edges connecting nodes within each sub-graph are relatively high, indicating similarity within the clusters.

2.1 Steps of spectral clustering algorithm

In general, any spectral clustering algorithm involves the following three steps.

1. **Pre-processing**
   Construct the similarity matrix W from the dataset using a measure that takes into account the aim of clustering, and then construct the similarity graph. There are different ways to do this depending on the pairwise similarity w_{ij}. The aim is to model the neighborhood relation among the data points $x_1, \cdots, x_n$. These ways are summarized as follows:

   - ε **-neighborhood graph:** In this graph, the vertices v_i, v_j will be connected by an edge if they are similar enough, i.e if $w_{ij} > \varepsilon$, ε is a pre-defined non-negative real number. Usually, this graph is considered as an unweighted graph.
   - k**-nearest neighbor graphs:** In this graph, the distance between each pair of vertices is computed using the Euclidean distance. Then, the vertices v_i, v_j are connected by an edge if v_j is among the k nearest neighbors of v_i or vice versa, and the edge is weighted by the similarity w_{ij}. The neighborhood relationship among data points is controlled by a pre-defined integer number k.

- **The fully connected graph:** In this graph, each vertex is connected to all other vertices by edges, and these edges are weighted by the similarities w_{ij}. This type of graph is useful only if the similarity function can model the neighborhood relation among the data points. The commonly used similarity function is the Gaussian similarity function, which is defined as $w_{ij} = \exp(-\|P_i - P_j\|^2/2\sigma^2)$, where the neighborhood relation is controlled by σ.

For further information on similarity graphs, we refer to Von Luxburg (2007) and Parodi (2012).

2. **Spectral representation**
Compute the Laplacian matrix of the graph, which is an essential tool to identify clusters in the data using spectral clustering. It is a matrix that characterizes the connectivity of a graph. It captures the relationships between the nodes, and can be used to identify the nodes that are most closely connected to each other. There are two different definitions for this matrix, depending on the degree matrix D and the similarity matrix S of the graph, as follows.
(a) Unnormalized Laplacian matrix L: $L = D - W$.
(b) Normalized Laplacian matrix L^{norm}: $L^{norm} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$.
The choice of Laplacian matrix type to use with spectral clustering depends on the application and the problem to be solved. Spectral clustering is often used to optimize two objective functions: Ratio Cut (Rcut) and Normalized Cut (Ncut). Both of these objective functions measure the quality of the partition of a graph into clusters. Let C_i be a subset of vertices i.e $C_i \subset V, i = 1, \cdots, k$ and its complement $\overline{C}_i := V \backslash C_i$, the Ratio Cut function (Rcut) (Hagen and Kahng (1992)) is defined as:

$$\text{Rcut}(C_1, \cdots, C_k) = \sum_{i=1}^{k} \frac{\text{cut}(C_i, \overline{C}_i)}{|C_i|}. \tag{2}$$

Where

$$\text{cut}(C, \overline{C}) := \sum_{i \in C, j \in \overline{C}} w_{ij}$$

$$|C_i| := \text{number of vertices in } C_i$$

In this function, the size of a subset C_i is measured by its number of vertices. Using the unnormalized Laplacian matrix L with spectral clustering leads to minimizing the Ratio Cut function.
In contrast, the Normalized cut function(Ncut)(Shi and Malik (2000)) is defined as:

$$\text{Ncut}(C_1, \cdots, C_k) = \sum_{i=1}^{k} \frac{\text{cut}(C_i, \overline{C}_i)}{\text{vol}(C_i)}. \tag{3}$$

where

$$\text{vol}(C) := \sum_{i \in C} d_i$$

In the Normalized cut function, the size of a subset C_i is measured by the weights of its edges. Using the normalized Laplacian matrix L^{norm} with spectral clustering leads to minimizing the Normalized cut function. For more details see Von Luxburg (2007).
The matrices L and L^{norm} have some important properties: they are symmetric and positive semi-definite matrices; the N eigenvalues $\lambda_1, \cdots, \lambda_N$ of these matrices are non negative real-valued, so $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$; the multiplicity k of the value 0 as an eigenvalue of these matrices is equal to the number of connected components $C_1, \cdots, C_k$ in the graph. (for more details, see Mohar et al (1991), Mohar (1997) and Chung (1997)).
The eigenvalues of the graph Laplacian matrix and its associated eigenvectors are computed. Then, the eigenvectors are used to constitute a low-dimensional representation of the data, where the clusters are more separated. Typically, the k eigenvectors corresponding to the k smallest eigenvalues are used to construct a k-dimensional representation of the data, as they capture the structure of the graph and important features of the data (see Wierzchoń and Kłopotek (2018)). Reducing dimension can reveal hidden patterns in the data that may be difficult to distinguish in higher dimension.

3. **clustering**
   Apply the k-means clustering algorithm to the low-dimensional representation to group the data points into k clusters.

In spectral clustering, a specific heuristic method has been proposed for choosing the number of clusters k. This method relies on the gap between two consecutive eigenvalues, with the number of clusters determined by the value of k that maximizes the eigengap δ_k: $\delta_k = \mid \lambda_{k+1} - \lambda_k \mid$, $k \geq 2$ (see Von Luxburg (2007)). This method is effective in determining the number of clusters when the dataset is well separated.

3 Max-stable processes

In this section, we will provide a brief overview of max-stable processes and define the extremal concurrence probability, which is a critical tool for our approach. We essentially follow Dombry et al (2018) and refer the reader to this reference for further details.

3.1 Definition of Max-stable processes

Let $Z_1(s), Z_2(s) \cdots$ be a sequence of independent replications of a spatial process $\{Z(s), s \in \mathcal{S}\}, \mathcal{S} \subset \mathbb{R}^d, d \geq 1$. If there exist continuous functions $A_n(s) > 0$ and $B_n(s) \in \mathbb{R}$ such that

$$\frac{\max_{i=1,\cdots,n} Z_i(s) - B_n(s)}{A_n(s)} \stackrel{d}{=} X(s), s \in \mathcal{S}, n \to \infty, \tag{4}$$

is non-degenerate, then $\{X(s), s \in \mathcal{S}\}$ is a max stable process (see De Haan and Pereira (2006)). The univariate maxima $X(s)$ at any location s, follows a Generalized Extreme Value distribution (GEV), i.e, for all $x \in \mathbb{R}$,

$$\mathbb{P}(X(s) \leq x) = \exp[-(1 + \xi(s)\frac{x - \mu(s)}{\sigma(s)})^{-1/\xi(s)}], \tag{5}$$

where $\mu(s) \in \mathbb{R}$ is the location parameter, $\sigma(s) > 0$ is the scale parameter and $\xi(s) \in \mathbb{R}$ is the shape parameter. These parameters are possibly different from one location to another. Setting $\mu(s) = \sigma(s) = \xi(s) = 1$, leads to consider unit Fréchet distributions, i.e, $\mathbb{P}(X(s) \leq x) = \exp[-1/x], x > 0$, and $\{X(s), s \in \mathcal{S}\}$ is called a simple max-stable process (see Ribatet (2017) and Ribatet et al (2016)). De Haan (1984) provided the spectral representation for simple max-stable processes $\{X(s), s \in \mathcal{S}\}$ as follows:

$$X(s) = \max_{i \geq 1} \zeta_i Y_i(s), s \in \mathcal{S}, \mathcal{S} \subset \mathbb{R}^d, d \geq 1 \tag{6}$$

where $\{\zeta_i, i \geq 1\}$ is a Poisson point process on $(0, \infty)$ with intensity $\zeta^{-2}d\zeta$ and $Y_1(s), Y_2(s), \cdots$ denote a sequence of independent replications of a non negative stochastic process $\{Y(s), s \in \mathcal{S}\}$ with $\mathbb{E}[Y(s)] = 1$ for all $s \in \mathcal{S}$. Equation (6) may be written as follows:

$$X(s) = \max_{\varphi \in \Phi} \varphi(s), s \in \mathcal{S} \tag{7}$$

where $\Phi = \{\varphi_i(s) = \zeta_i Y_i(s) : s \in \mathcal{S}, i \geq 1\}$ is a Poisson point process on $\mathbb{C}_0$, the space of non-negative continuous functions on $\mathcal{S}$ (see Ribatet (2017)). Let $\mathbf{S}$ be a set of m spatial locations : $\mathbf{S} = \{s_1, \cdots, s_m\} \subset \mathcal{S}$, then the multivariate maxima distribution is given by

$$\mathbb{P}\{X(s_1) \leq x_1, \cdots, X(s_m) \leq x_m\} = \exp\left\{-\mathbb{E}\left[\max_{j=1,\cdots,m} \frac{Y(s_j)}{x_j}\right]\right\} \tag{8}$$

where $\{Y(s), s \in \mathcal{S}\}$ is the process appearing in Equation (6). The exponent function

$$V_{\mathbf{S}}(x_1, \cdots, x_m) = \mathbb{E}\left[\max_{j=1,\cdots,m} \frac{Y(s_j)}{x_j}\right], \tag{9}$$

is called the exponent measure. It characterizes the dependence structure of $X(s_1), \cdots, X(s_m)$. Since the exponent measure is homogeneous of order -1, we can obtain a useful relation by setting $x_j = x$ for all $j = 1, \cdots, m$ such that $V_{\mathbf{S}}(1, \cdots, 1) = \theta_{\mathbf{S}}$ where $\theta_{\mathbf{S}}$ is the extremal coefficient that provides a summary of the dependence structure (see Schlather and Tawn (2003) and Smith (1990)). Particularly, when $\mathbf{S} = \{s_1, s_2\}$ the extremal coefficient satisfies $\theta_{\mathbf{S}} = V_{\mathbf{S}}(1, 1) \in [1, 2]$. The lower bound corresponds to the variables $X(s_1)$ and $X(s_2)$ which are completely dependent, while the upper bound corresponds to the case where they are independent.

Several models for max-stable processes have been presented based on this spectral representation, including the Brown-Resnick model (see Brown and Resnick (1977)), the Smith model (see Smith (1990)), the Schlather model (see Schlather (2002)), and the Extremal-t model (see Opitz (2013)).

3.2 Extremal concurrence probability

Other indices in order to measure the dependence between extremes exist in the literature. Dombry et al (2018) introduced the extremal concurrence probability for the analysis of extremal dependence, which was especially designed for max-stable processes. It has properties similar to the pairwise extremal coefficient, but it has the advantage of being a probability measure, which makes it more interpretable and axiomatic. The extremal concurrence probability focuses on the occurrence times of extremes, which means whether the record maxima occurs simultaneously, i.e., at the same time for all locations. It can be interpreted as the chance of a single extreme event affecting all the locations and being responsible for the record maximum.
It is based on the spectral representation of the max-stable processes. The idea behind this metric can be explained as follows.
Recall the spectral representation in Equation (7). We say that the extremes are concurrent at locations $s_1, \cdots, s_m \in \mathcal{S}$ if

$$X(s_j) = \varphi_\ell(s_j), j = 1, \cdots, m \tag{10}$$

for some $\ell \geq 1$. This means that the values of the process $\{X(s), s \in \mathcal{S}\}$ at locations $s_1, \cdots, s_m$ come from the same spectral function φ_ℓ.
The extremal concurrence probability is defined as

$$p_r(s_1, \cdots, s_m) = \mathbb{P}\{\text{for some } \ell \geq 1 : X(s_j) = \varphi_\ell(s_j),\ j = 1, \cdots, m\} \tag{11}$$

According to Theorem 3 in Dombry et al (2018), the bivariate extremal concurrence probability estimation coincides with Kendall's τ statistic:

$$\hat{p}_r(s_1, s_2) \equiv \tau = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \text{sign}\{X_i(s_1) - X_j(s_1)\}\text{sign}\{X_i(s_2) - X_j(s_2)\}, \tag{12}$$

where $\{X_i(s), s \in \mathcal{S}, i = 1, \cdots, n\}$ are n independent copies of $\{X(s), s \in \mathcal{S}\}$. The bivariate extremal concurrence probability for max-stable processes satisfies $p_r(s_1, s_2) = 0$ if and only if $X(s_1)$ and $X(s_2)$ are independent, and $p_r(s_1, s_2) = 1$ if and only if $X(s_1)$ and $X(s_2)$ are almost surely equal. These properties were stated and proved in Proposition 1 of Dombry et al (2018).

4 Adapting spectral clustering for max stable process

Let $X_i(s_j), s_j \in \mathcal{S}$, $\mathcal{S} \subset \mathbb{R}^d$, $d = 2$, $i = 1, \cdots, n$ be a sequence of n independent and identically distributed max stable processes at different locations $s_j, j = 1, 2, \cdots, m$. In order to apply spectral clustering in extremal dependence context, locations $s_1, \cdots, s_m$ are considered as vertices in a fully connected graph. Each vertex is connected to all others by edges, and the weights of these edges represent the similarity values among the locations. Figure 1 represents a fully connected graph for a max-stable process consisting of 15 locations. For viewing purposes, the self-edges were not shown. Select-

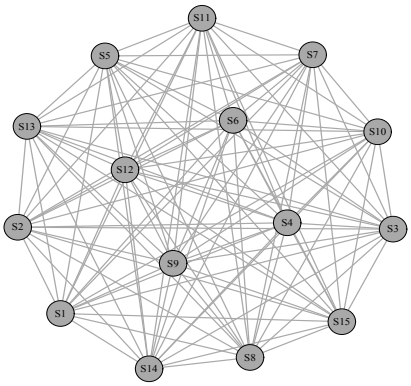

Fig. 1 Fully connected graph with 15 vertices. Each vertex represents a location in the max-stable process

ing an appropriate metric to construct the similarity matrix is essential in the spectral clustering algorithm, especially when using a fully connected graph. It is important to choose an extremal dependence measure that can accurately model the neighborhood relations among the locations. In this study, we used the extremal concurrence probability, as introduced by Dombry et al (2018)(see Section 3.2). The similarity matrix represents the pairwise extremal concurrence probability matrix, denoted by $CP \in \mathbb{R}^{m \times m}$, where m is the number of locations. For a pair $(s, s') \in \mathcal{S} \times \mathcal{S}$, the element of the matrix CP is given by:

$$\hat{p}_r(s, s') = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} \text{sign}\{X_i(s) - X_j(s)\}\text{sign}\{X_i(s') - X_j(s')\} \quad (13)$$

After constructing the similarity matrix CP according to Equation (13), it is used to compute the graph Laplacian matrix. Using the normalized graph Laplacian matrix L^{norm} helps to achieve our goal of making the size of the resulting clusters dependent on the concurrence of extremes (i.e., the weights of the graph edges). The spectrum λ and the eigenvectors of L^{norm} are then computed. The eigenvectors $q_1, \cdots, q_k$ are used to constitute a k-dimensional

representation for the data. This is done by representing these eigenvectors as columns of an $m \times k$ matrix denoted Q. Each row in Q represents a location s_j: $s_j \rightarrow (q_{j,1}, \cdots, q_{j,k})$, $j = 1, \cdots, m$, this is called spectral mapping (see Wierzchoń and Kłopotek (2018)). Normalizing each row of Q to norm 1 results in a matrix denoted $Y \in \mathbb{R}^{m \times k}$. According to Ng et al (2001), this last step improves the performance of the clustering algorithm. Instead of using k-means, which is usually used at this step, we used a Gaussian Mixture Model (GMM) to cluster the rows of Y. GMM clusters the datapoints based on probability distribution, considering that the datapoints come from a Gaussian mixture. Each cluster has a Gaussian distribution model with parameters mean and covariance. Taking the covariance into account makes GMM more robust than k-means, which depends only on the cluster mean. For more details about GMM, see for example Bouveyron et al (2019). We summarize these steps in Algorithm 1.

Algorithm 1 Adapted spectral clustering

Require: The similarity matrix $CP \in \mathbb{R}^{m \times m}$, constructed according to Equation (13) .

Ensure: Clusters $\{C_1, \cdots, C_k\}$.

1: Compute the normalized Laplacian matrix $L^{norm} = D^{-\frac{1}{2}}(D - CP)D^{-\frac{1}{2}}$.
2: Compute the spectrum of L^{norm}.
3: Compute the eigengap δ_k and determine the potential number of clusters which correspond to relatively large values of δ_k. Then for each of them repeat the steps 4 to 7.
4: Compute the k smallest eigenvectors $q_1, q_2, \cdots, q_k$ of L^{norm}, and arrange these vectors in columns to be a matrix Q, where $Q \in \mathbb{R}^{m \times k}$.
5: Normalize the rows of Q to norm 1, resulting the matrix $Y \in \mathbb{R}^{m \times k}$: $Y_{jl} = Q_{jl}/(\sum_l Q_{jl}^2)^{\frac{1}{2}}$, $j = 1, \cdots, m$, $l = 1, \cdots, k$.
6: Consider each row of Y as a point in $\mathbb{R}^k$ and implement Gaussian Mixture Model (GMM) to cluster them into k clusters.
7: Assign the location s_j to cluster l if and only if row j of the matrix Y is assigned to cluster l.

5 Applicability of the adapted spectral clustering

The adapted Spectral clustering algorithm, which utilizes the extremal concurrence probability as a similarity matrix, aims to identify regions where the pairwise concurrence probability between locations is high. As a result, we obtain regions where the locations exhibit a similar extremal dependence. This can reduce the possibility that the extremal dependence will be non-stationary in the clusters.

As a preliminary step, we need to determine the spatial structure of the k clusters. This stems from the expectation that locations geographically close to each other may exhibit similar extremal dependence. We obtained the spatial structure of these k clusters by applying the Partitioning Around Medoids (PAM) clustering algorithm proposed by Kaufman and Rousseeuw (1990) to the spatial coordinates of these locations. As a result, we obtained k clusters in which the locations are spatially closer to each other, as measured by their euclidean distances.

We illustrate the application of spectral clustering by examining a non-stationary max-stable process. We construct this process using an approach presented by Huser and Genton (2016), named max-stable mixtures. Let $X^1(s)$ and $X^2(s)$ be independent max stable processes that have unit Fréchet margins defined on the same space $\mathcal{S}$. Then the process $X(s) = \max\{\pi(s)X^1(s), (1 - \pi(s))X^2(s)\}$ is a simple max stable process, where $\pi(s)$ is any function with $0 \leq \pi(s) \leq 1$. The bivariate distribution is given by

$$F(x_1, x_2;\ \Psi) = \mathbb{P}\{X(s) \leq x_1, X(s') \leq x_2\} = \exp\Big\{-V_{\mathbf{S}}^{mix}(x_1, x_2)\Big\} \tag{14}$$

and

$$V_{\mathbf{S}}^{mix}(x_1, x_2) = V_{\mathbf{S}}^1\Big(\frac{x_1}{\pi(s)}, \frac{x_2}{\pi(s')}\Big) + V_{\mathbf{S}}^2\Big(\frac{x_1}{1-\pi(s)}, \frac{x_2}{1-\pi(s')}\Big) \tag{15}$$

where, $V_{\mathbf{S}}^1$ and $V_{\mathbf{S}}^2$ are the exponent measure of $X^1(s)$ and $X^2(s)$, respectively, and $\mathbf{S} = \{s, s'\}$. The proportion $\pi(s)$ determines which of the processes $X^1(s)$ and $X^2(s)$ prevails at location s. If $\pi(s)$ is constant across space, then the process $X(s)$ is stationary. However, $X(s)$ becomes non-stationary when $\pi(s)$ varies spatially, such as if it depends on covariates. This variation may lead to capturing different extreme behaviors in different spatial regions.

For clustering purposes, we will consider that each spatial cluster k has its own mixing proportion π_k. To make the process more realistic, we take into account the presence of regression in the values of π_k between the clusters. Figure 2 represents an example of realization of such a process by mixing a Schlather and an extremal-t process over the $[0, 1]^2$. In the simulation study (Section 7), we will explore how spectral clustering can recover these clusters of locations. Assuming we have two clustered regions, depending on these clusters, we can rewrite the bivariate distribution in Equations 14 and 15 as

$$F(x_1, x_2;\ \Psi) = \begin{cases} \exp\Big\{-\big\{V_{\mathbf{S}}^1(\frac{x_1}{\pi_1} + \frac{x_2}{\pi_1})) + V_{\mathbf{S}}^2(\frac{x_1}{1-\pi_1} + \frac{x_2}{1-\pi_1})\big\}\Big\}, & \text{if } s, s' \in C_1 \\ \exp\Big\{-\big\{V_{\mathbf{S}}^1(\frac{x_1}{\pi_2} + \frac{x_2}{\pi_2})) + V_{\mathbf{S}}^2(\frac{x_1}{1-\pi_2} + \frac{x_2}{1-\pi_2})\big\}\Big\}, & \text{if } s, s' \in C_2 \\ \exp\Big\{-\big\{V_{\mathbf{S}}^1(\frac{x_1}{\pi_1} + \frac{x_2}{\pi_2})) + V_{\mathbf{S}}^2(\frac{x_1}{1-\pi_1} + \frac{x_2}{1-\pi_2})\big\}\Big\}, & \text{if } s \in C_1, s' \in C_2 \end{cases} \tag{16}$$

For simplicity we will denote the distribution function components in Equation 16 by $F_{C_1}(x_1, x_2;\ \Psi)$ if $s, s' \in C_1$, $F_{C_{21}}(x_1, x_2;\ \Psi)$ if $s, s' \in C_2$ and $F_{C_{12}}(x_1, x_2;\ \Psi)$ if $s \in C_1, s' \in C_2$. Using the distribution in Equation 16, we can model the whole region under study depending on clustered regions.

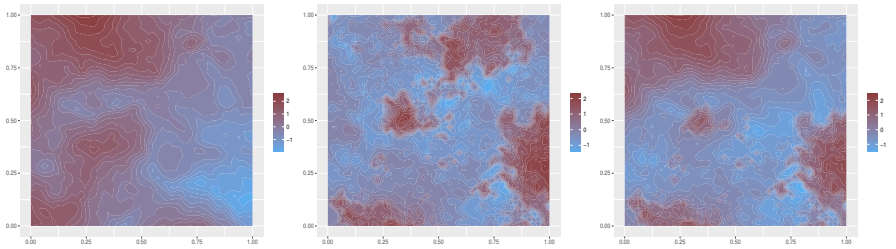

Fig. 2 Example of realization of a max-stable mixtures process on the logarithm scale. Left panel: realization of an Extremal-t process with whittle-matérn correlation function $\rho(h) = (2^{1-c_1}/\Gamma(c_1))(\|h\|/c_2)^{c_1} K_{c_1}(\|h\|/c_2)$ with degree of freedom equal to 5, range $c_2 = 0.7$ and smooth $c_1 = 1$. Middle panel: realization of a Schlather process with exponential correlation function $\rho(h) = \exp(-\|h\|/\phi)$, with range $\phi = 0.7$. Right panel: corresponding realization of the max-stable mixture with $\pi(s) = 3x - 1$ if $x \in [0.4, 0.6]$, if $x < 0.4$ $\pi(s) = 0.2$ and $\pi(s) = 0.8$ if $x > 0.6$, where x represents the value of the X coordinate corresponding to location s.

6 Inference: composite likelihood approach

The full likelihood inference for max-stable models is computationally intractable (Castruccio et al (2016)). The most widely used approach is the pairwise composite likelihood (Padoan et al (2010)). Let Ψ represent the vector of unknown parameters; the pairwise composite log-likelihood can be expressed as follows:

$$\mathcal{P}(\Psi) = \sum_{i=1}^{n} \sum_{j=1}^{m-1} \sum_{j'=j+1}^{m} \mathcal{W}_{jj'} \log \mathcal{L}(x_{ij}, x_{ij'};\ \Psi) =: \sum_{i=1}^{n} \mathcal{P}_i(\Psi) \tag{17}$$

where x_{ij} represents the block maxima i which recorded at station j, $\mathcal{L}(x_{ij}, x_{ij'};\ \Psi)$ is the likelihood of the pair $(x_{ij}, x_{ij'})$ and $\mathcal{W}_{jj'}$ are non negative weights that specify the contribution of each pair. Therefore the maximum pairwise liklihood estimator is given by $\hat{\Psi} = \arg\max \mathcal{P}(\Psi)$. Under some regularity conditions (see Padoan et al (2010)), $\hat{\Psi}$ obtained by maximizing the pairwise composite log-likelihood in Equation (17) is consistent and asymptotically normally distributed, such that

$$\hat{\Psi} \sim \mathcal{N}(\Psi, \mathcal{I}(\Psi)^{-1}) \tag{18}$$

with

$$\mathcal{I}(\Psi) = \mathcal{H}(\Psi)\, \mathcal{J}(\Psi)^{-1}\, \mathcal{H}(\Psi) \tag{19}$$

where $\mathcal{I}(\Psi)$ is the sandwish information matrix, $\mathcal{H}(\Psi) = \mathbb{E}\big[-\nabla^2 \mathcal{P}(\Psi)\big]$ is the sensitivity matrix and $\mathcal{J}(\Psi) = \mathbb{V}\big[\nabla \mathcal{P}(\Psi)\big]$ is the variability matrix. Therefore, the variance of the estimated parameters can be assessed by estimating the matrices $\mathcal{H}(\Psi)$ and $\mathcal{J}(\Psi)$. In this context, the empirical estimates of the

matrices $\mathcal{H}(\Psi)$ and $\mathcal{J}(\Psi)$ are given by

$$\hat{\mathcal{H}}(\Psi) = -\sum_{i=1}^{n} \nabla^2 \mathcal{P}_i(\hat{\Psi}) \tag{20}$$

and

$$\hat{\mathcal{J}}(\Psi) = \sum_{i=1}^{n} \nabla \mathcal{P}_i(\hat{\Psi}) \nabla \mathcal{P}_i(\hat{\Psi})^T \tag{21}$$

In practice, $\hat{\mathcal{H}}(\Psi)$ can be obtained directly from the optimization algorithm since it equals the negative of the Hessian matrix.
In this paper, our interest is to use the clustered regions in modeling the whole region. We define the censored pairwise likelihood contribution $\mathcal{L}_{CL}(x_{ij}, x_{ij'};\ \Psi)$ for a pair $(x_{ij}, x_{ij'})$ as follows

$$\mathcal{L}_{CL}(x_{ij}, x_{ij'};\ \Psi) = \begin{cases} \partial^2_{12} F_{C_1}(x_{ij}, x_{ij'}), & \text{if } s_j, s_{j'} \in C_1 \\ \partial^2_{12} F_{C_2}(x_{ij}, x_{ij'}), & \text{if } s_j, s_{j'} \in C_2 \\ \partial^2_{12} F_{C_1 2}(x_{ij}, x_{ij'}), & \text{if } s_j \in C_1, s_{j'} \in C_2 \end{cases} \tag{22}$$

where ∂_i is the differentiation with respect to variable x_i.
Finally, selecting the best-fitted model under the composite likelihood approach is performed using the Composite Likelihood Information Criterion (CLIC)(Varin and Vidoni (2005)), which is defined as CLIC $= -2\,\mathcal{P}(\hat{\Psi}) + 2\,\mathrm{tr}(\hat{\mathcal{H}}(\hat{\Psi})^{-1}\,\hat{\mathcal{J}}(\hat{\Psi}))$. The lowest value of CLIC correspond to best fit model.

7 Simulation study

In order to assess the accuracy of our algorithm, we tested it on two simulation cases. In each cases, the coordinates of the locations were generated randomly and uniformly in $[0, 1]^2$. In practice, we usually do not have two completely separated and independent clusters, some smoothness and dependence from one to the other may be present, this is why we will use the max-stable mixture model in two different simulation case.

- **Case 1:** In this case, we assumed the existence of two clusters, one in the north and the other in the south. To identify the spatial structure of these clusters, we applied the PAM algorithm to the coordinates of 80 locations. The left panel in Figure 3 displays the two clusters, with the first cluster in the north (depicted by red points) and the second in the south (depicted by blue points). We then established the boundary area between the clusters arbitrarily, represented by the gray points located between two black lines (latitude ranging from 0.4 to 0.55), as shown in the right panel of Figure 3. In this context, we have named the cluster in the north as A, the cluster in the south as B, and the boundary area as middle. Then, we simulated a non-stationary process, as described in Section 5. Denoted by M_1 , the model represents a mixture of the Schlather model with a powered exponential

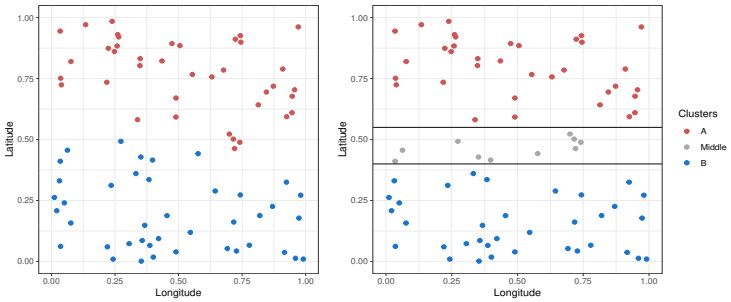

Fig. 3 The left panel shows the spatial structure of the two clusters, while the right panel displays cluster A in the north, cluster B in the south, and the middle, which is the boundary area between clusters A and B.

correlation function $\rho(h) = \exp(-(\|h\|/\phi)^{\alpha})$, and the Brown-Resnick model with a semivariogram $\gamma(h) = (\|h\|/\phi)^{\alpha}$. The parameters of the Schlather model are set to $\phi_1 = 0.3$ and $\alpha_1 = 1$, where the subscript index 1 refers to the first mixing component. The parameters of the Brown-Resnick model are set to $\phi_2 = 0.7$ and $\alpha_2 = 1.2$, where the subscript index 2 refers to the second mixing component. The number of observations is fixed at 500, and the spatially varying proportion was set as follows:

$$\pi(s) = \begin{cases} 0.3, & \text{if } s \in \text{A} \\ \frac{53}{30} - \frac{8}{3} \times \text{latitude}, & \text{if } s \in \text{Middle} \\ 0.7, & \text{if } s \in \text{B} \end{cases} \tag{23}$$

in order to get continuous realizations of the underlying process. With this choice, we get two different areas (A and B) where the underlying process is stationary; moreover the areas A and B are dependent leading to a more realistic model for environmental phenomena.

Afterward, we applied adapted spectral clustering to this process. The eigengap heuristic indicated that the number of clusters equals 2, as illustrated in the left panel of Figure 4. In order to further confirm this choice, we employed the widely used Silhouette Index (Rousseeuw (1987)) for clustering validation. The Silhouette Index is a valuable metric for assessing the separation of clusters, with values ranging from -1 to 1. The highest Silhouette Index value corresponds to the most reliable clustering results. We considered potential numbers of clusters, including 2, 3, and 4, which corresponded to relatively large eigengap values, the right panel of Figure 4 clearly shows that the optimal number of clusters is 2. The spectral clustering result for k equals 2 is depicted in the left panel of Figure 5, with red points representing cluster A, blue points representing cluster B, and black triangles representing the points that lie in middle. The pairwise extremal concurrence probability matrix for the clustering result is displayed in the right panel of Figure 5. Since we know the true clusters A, Middle and B, We

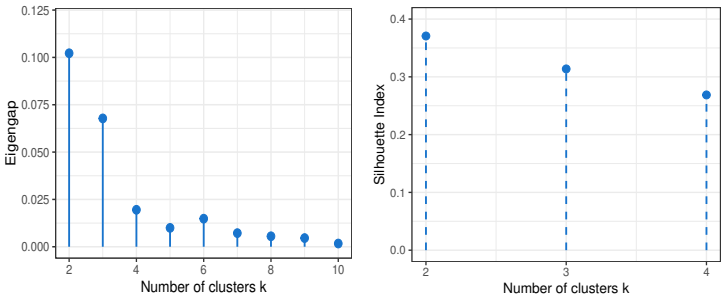

Fig. 4 The left panel shows the eigengap values against number of clusters k, while the right panel displays Silhouette Index values against number of clusters k.

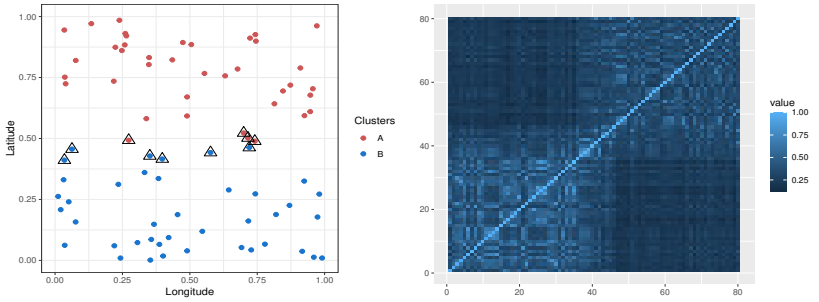

Fig. 5 The left panel shows the spectral clustering result for k=2 where the red points represent cluster A, blue points represent cluster B and the black traingles represent the points that lie in the middle, while the right panel shows the pairwise extremal concurrence probability matrix for the clustering result.

can calculate the precision of the spectral clustering algorithm for recovering the true clusters by determining the ratio of locations in the true cluster shared in common with the cluster resulting from spectral clustering. We will denote this ratio as $\mathcal{A}_{C_t,C_{sp}}$ for simplicity:

$$\mathcal{A}_{C_t,C_{sp}} = \frac{|C_t \cap C_{sp}|}{|C_t|} \tag{24}$$

where C_t represents the true cluster, C_{sp} represents the cluster resulting from spectral clustering, and $|.|$ denotes the number of elements in a set. The values of $\mathcal{A}_{C_t,C_{sp}}$ for all combinations between the true clusters and the spectral clustering clusters are presented in Table. 1. Its clear that the true clusters A and B are correctly recovered by spectral clustering. Regarding the Middle cluster, 4 locations are clustered with A which has $\pi(s)$ values closer to 0.3 and 6 locations are clustered with B which has $\pi(s)$ values closer to 0.7. To assess whether the number of observations affects the algorithm performance, we implemented it with different numbers of observations: n = 500, 200, 100, and 50. For each number of observations, we simulated 10 processes on the same coordinates with the same parameters as mentioned

Table 1 The values of $\mathcal{A}_{C_t,C_{sp}}$ for one simulation of case 1 with the number of observation equals 500

	Spectral clustering clusters	
True clusters	A	B
A	1	0
Middle	0.4	0.6
B	0	1

above. The mean of $\mathcal{A}_{C_t,C_{sp}}$ for the 10 simulations is shown in Table 2. From Table 2, we can observe that a higher number of observations leads

Table 2 The mean values of $\mathcal{A}_{C_t,C_{sp}}$ for 10 simulations of case 1 with the number of observations set to n = 500, 200, 100, and 50.

	Spectral clustering clusters							
	$n = 500$		$n = 200$		$n = 100$		$n = 50$	
True clusters	A	B	A	B	A	B	A	B
A	1	0	1	0	1	0	0.98	0.02
Middle	0.41	0.59	0.43	0.57	0.39	0.61	0.51	0.49
B	0	1	0	1	0	1	0.07	0.93

to better recovery of the true clusters. However, the algorithm performs well even when the number of observations is relatively small.

Its worth noting that for the 40 simulations in Table 2, the accuracy of determining the number of clusters (k = 2) using the eigengap heuristic is 0.8. However, the eigengap is still useful in providing us with the potential number of clusters, which corresponds to the relatively largest eigengap. When we calculate the Silhouette Index only for this potential number of k, we achieve an accuracy of 1 in determining the correct number of clusters.

Now, we will assess the ability of the proposed censored pairwise likelihood, as defined in Section 6, Equation 22, to estimate the parameters of the whole region based on clusters. We simulate $N = 100$ max stable mixture processes (model M_1) with the same conditions as explained previously, with the number of observations fixed at 100. Then, we use Equation 22 to obtain the estimator $\hat{\Psi}$. After that, we will create a boxplot for the estimated parameters and calculate the mean estimate and the root mean square error (RMSE) for each estimated parameters. Let $\hat{\Psi}_i$ denoted the ith estimation,

then

$$\text{RMSE} = \left[\frac{1}{N}\sum_{i=1}^{N}(\hat{\Psi}_i - \Psi)^2\right]^{1/2} \tag{25}$$

The boxplots of the errors of the estimated parameters, i.e., $(\hat{\Psi} - \Psi)$, are displayed in the left panel of Figure 6, and the RMSE values are shown in the right panel of Figure 6. The subscript index 1 (2) refers to the first (second) mixing model. The mean estimates of the parameters are $\hat{\pi}_1 = 0.29$, $\hat{\pi}_2 = 0.67$, $\hat{\phi}_1 = 0.28$, $\hat{\alpha}_1 = 0.98$, $\hat{\phi}_2 = 0.66$ and $\hat{\alpha}_2 = 1.23$. Generally, the

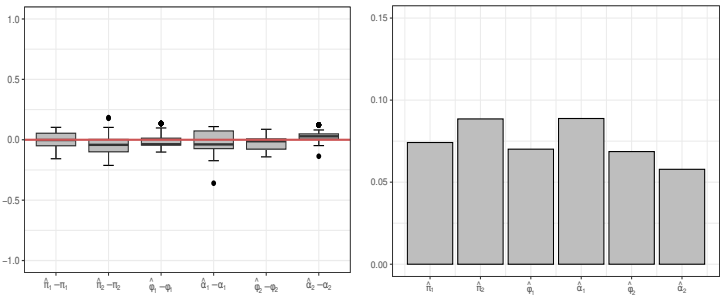

Fig. 6 The left panel shows the boxplots displaying $(\hat{\Psi} - \Psi)$ of the estimated parameters using censored pairwise likelihood Equation 22 based on 100 simulations of model M_1 with parameters $\pi_1 = 0.3$, $\pi_2 = 0.7$, $\phi_1 = 0.3$, $\alpha_1 = 1$, $\phi_2 = 0.7$ and $\alpha_2 = 1.2$. The number of observations fixed at 100. The red horizontal line represents the zero value. Meanwhile, the right panel displays the barplots for the RMSE of $\hat{\Psi} = \{\hat{\pi}_1, \hat{\pi}_2, \hat{\phi}_1, \hat{\alpha}_1, \hat{\phi}_2, \hat{\alpha}_2\}$ for the same simulations.

proposed censored pairwise likelihood appears to work well in estimating the parameters for the whole region depending on clustered regions.

- **Case 2:** Here, similar to case 1, we assumed the existence of two clusters, one in the west and the other in the east. The spatial structure of these clusters is depicted in the left panel of Figure 7 after applying the PAM algorithm to the coordinates of 40 locations. The red points represent the west cluster, while the blue points represent the east cluster. We then established the boundary area between the clusters arbitrarily, represented by the gray points located between two black lines (longitude ranging from 0.4 to 0.6), as shown in the right panel of Figure 7. We have named the cluster in the west as A, the cluster in the east as B, and the boundary area as middle. Next, we mixed the Brown-Resnick model with a semivariogram $\gamma(h) = (\|h\|/\phi)^{\alpha}$, with an isotropic Smith model, as described in Section 5, we denote this model by M_2. In this model, the parameters of the Brown-Resnick model are set to $\phi_1 = 0.3$ and $\alpha_1 = 1$, where the subscript index 1 refers to the first mixing component. The parameter of the isotropic Smith model are set to $\phi_2 = 0.5$, where the subscript index 2 refers to the second mixing component. The number of observations is set to 500, and the spatially varying proportion

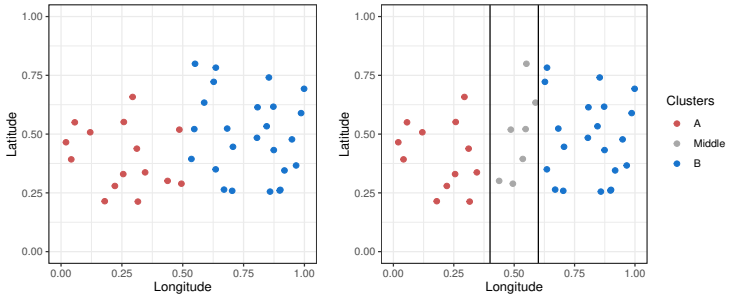

Fig. 7 The left panel shows the spatial structure of the two clusters, while the right panel displays cluster A in the west, cluster B in the east, and the middle, which is the boundary area between clusters A and B.

was defined as follows:

$$\pi(s) = \begin{cases} 0.2, & \text{if } s \in \text{A} \\ -1 + 3 \times \text{longitude}, & \text{if } s \in \text{Middle} \\ 0.8, & \text{if } s \in \text{B} \end{cases} \tag{26}$$

Then, we applied adapted spectral clustering to this process. The eigengap heuristic and Silhouette Index indicate that the number of clusters equals 2, as illustrated in the left and right panels of Figure 8 respectively. The

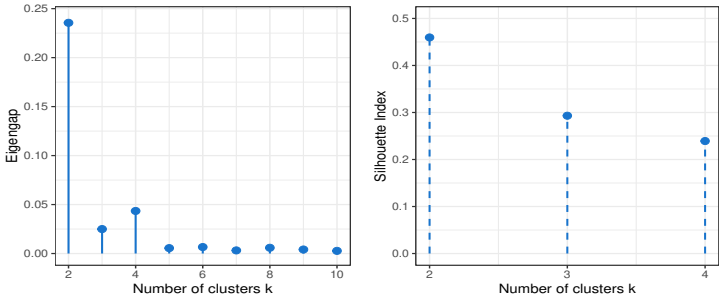

Fig. 8 The left panel shows the eigengap values against number of clusters k, while the right panel displays Silhouette Index values against number of clusters k.

spectral clustering result for k equals 2 is depicted in the left panel of Figure 9, with red points representing the west cluster A, blue points representing the east cluster B, and black triangles representing the points that lie in the boundary area. The pairwise extremal concurrence probability matrix for the clustering result is displayed in the right panel of Figure 9. Table 3 represents the values of $\mathcal{A}_{C_t,C_{sp}}$ for all combinations between the true clusters and the spectral clustering clusters. It's clear that the true clusters A and B are correctly recovered by spectral clustering. As for the Middle cluster, 3 out of 7 locations are clustered with A, which has $\pi(s)$ values closer to 0.2, and 4 out

Fig. 9 The left panel shows the spectral clustering result for k=2 where the red points represent cluster A, blue points represent cluster B and the black traingles represent the points that lie in the middle, while the right panel shows the pairwise extremal concurrence probability matrix for the clustering result.

of 7 locations are clustered with B, which has $\pi(s)$ values closer to 0.8. Table 4 presents the algorithms performance for different numbers of observations: n = 500, 200, 100, and 50. For each number of observations, we simulated 10 processes on the same coordinates with the same parameters as mentioned above. The table displays the mean of $\mathcal{A}_{C_t,C_{sp}}$ for the 10 simulations. From

Table 3 The values of $\mathcal{A}_{C_t,C_{sp}}$ for one simulation of case 2 with the number of observation equals 500

	Spectral clustering clusters	
True clusters	A	B
A	1	0
Middle	0.43	0.57
B	0	1

Table 4 The mean values of $\mathcal{A}_{C_t,C_{sp}}$ for 10 simulations of case 2 with the number of observations set to n = 500, 200, 100, and 50.

	Spectral clustering clusters							
	$n = 500$		$n = 200$		$n = 100$		$n = 50$	
True clusters	A	B	A	B	A	B	A	B
A	1	0	1	0	1	0	1	0
Middle	0.46	0.54	0.46	0.54	0.46	0.54	0.46	0.54
B	0	1	0	1	0	1	0	1

Table 4, its clear that the algorithm successfully recovers the true clusters in all the tested numbers of observations, also, the number of observations does not affect the clustering of points that lie in the boundary area. Its worth noting that, in the 40 simulations presented in Table 4 the accuracy of determining the number of clusters (k = 2) using the eigengap heuristic and the Silhouette Index is equal to 1.

In order to asses the performance of the the proposed censored pairwise likelihood, we simulate $N = 100$ models the same conditions as M_2 and with the number of observations fixed at 100. Afterwards, we estimate the parameters use Equation 22. Figure 10 display in its left panel the boxplots of the errors of the estimated parameters, while its right panel shown the RMSE values. As before, the subscript index 1 (2) refers to the first (second) mixing model. The mean estimates of the parameters are $\hat{\pi}_1 = 0.25$, $\hat{\pi}_2 = 0.76$, $\hat{\phi}_1 = 0.29$, $\hat{\alpha}_1 = 1.05$ and $\hat{\phi}_2 = 0.47$. Regardless of a slight overfitting or underfitting for some parameters, the proposed censored pairwise likelihood works well.

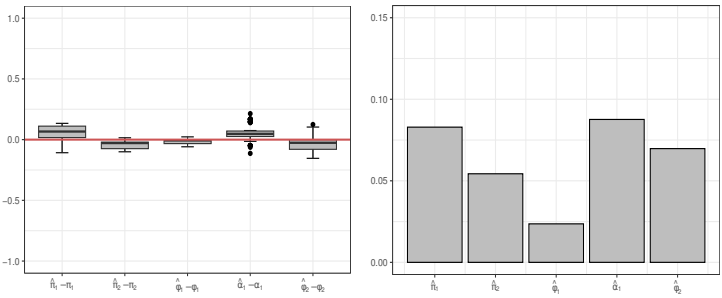

Fig. 10 The left panel shows the boxplots displaying $(\hat{\Psi} - \Psi)$ of the estimated parameters using censored pairwise likelihood Equation 22 based on 100 simulations of model M_2 with parameters $\pi_1 = 0.2$, $\pi_2 = 0.8$, $\phi_1 = 0.3$, $\alpha_1 = 1$ and $\phi_2 = 0.5$. The number of observations fixed at 100. The red horizontal line represents the zero value. Meanwhile, the right panel displays the barplots for the RMSE of $\hat{\Psi} = \{\hat{\pi}_1, \hat{\pi}_2, \hat{\phi}_1, \hat{\alpha}_1, \hat{\phi}_2\}$ for the same simulations.

Since the results of the simulations appear satisfactory, we can use this technique to clustering and modeling the maxima for rainfall in East Australia and France.

8 Application on data

This section is devoted to two data applications: one on rainfall in Australia's east coast, and the other on rainfall in France.

8.1 Rainfall over east coast of Australia

We will begin with a brief description of the data, followed by the application of our clustering method and a discussion of the results.

8.1.1 Description of the data

This data represents the daily rainfall totals (in millimeters) measured over a 24-hour period at 40 stations on the east coast of Australia during the winter season (April to September) from 1972 to 2019, resulting in a total of 183 * 48 = 8,784 observations at each station. The altitude of these stations ranges from 2 to 540 meters. The distance between the stations ranges approximately from 11 km to 2058 km. The geographic locations of the 40 stations are illustrated in the left panel of Figure 11. More information about this data can be found in references such as Ahmed et al (2022), Bacro et al (2016), Ahmed et al (2017), and Abu-Awwad et al (2020). The data is freely available on the website http://www.bom.gov.au.

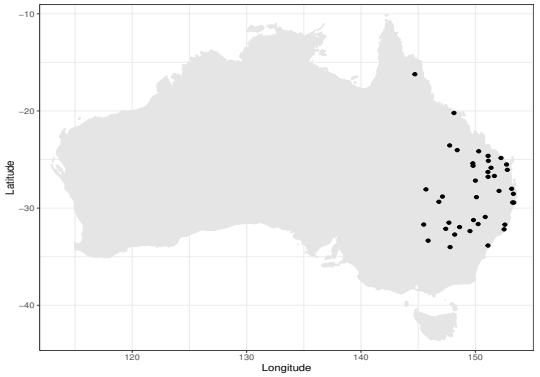

Fig. 11 Geographic locations of 40 stations on the east coast of Australia.

8.1.2 Regionalizing the maxima for rainfall over east coast of Australia

We will apply the adapted spectral clustering algorithm, described in Section 4 Algorithm 1, to clustering the extremal dependence. To demonstrate the effect of the block size on the resulting clusters, we will use different sizes of block. Specifically, we will test block sizes of 183 days, 30 days, 15 days and 10 days. Figure 12 shows the results of spectral clustering for Australia's rainfall data for each block size. We can observe from the first column of Figure 12 that the eigengap heuristic indicates that the number of clusters is 2 for all the tested block sizes. To reinforce this result, we calculated the Silhouette Index for the potential number of clusters, and the results confirm that the best number of clusters is 2 for all the tested block sizes, as shown in the middle column of Figure 12. In the right column of Figure 12 it is evident that the spectral clustering result is the same for all the tested block sizes, with the exception of one location in a block size of 10 days. This variation is expected since it is located in the boundary area between the clusters. For this data, we can say that the size of the block does not affect the clustering map.

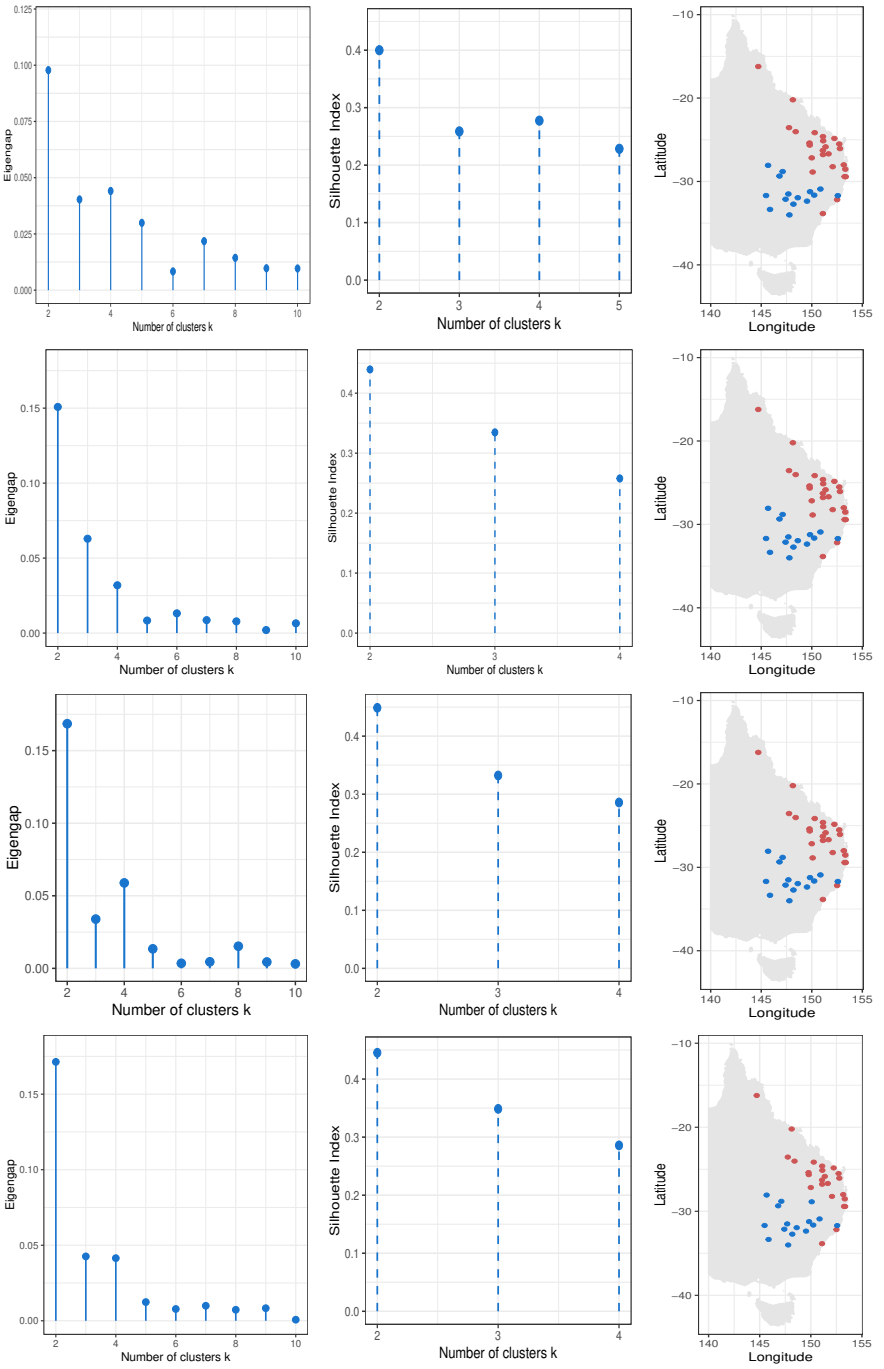

Fig. 12 The left column displays Eigengap values, and the middle column shows the Silhouette Index against the number of clusters k, while the right column represents the spectral clustering result for the east coast of Australia rainfall data over the period of 1972-2019. The first (second, third, and fourth) row is for a block size of 183 days (30 days, 15 days, and 10 days) respectively.

8.1.3 Modelling the rainfall over east coast of Australia

In order to determine whether the clustering of extremal dependence is significant in modeling the maxima of rainfall over the east coast of Australia, we will assess this by modeling with a 30 day block size.
We consider nine arbitrary models that belong to three classes: max-stable, stationary max-stable mixture (constant mixing proportion), and non-stationary max-stable mixture (different mixing proportions for each cluster). The descriptions of the models are as follows.
Class I: consists of max stable models I1-I3.

- I1: Schlather model with an powered exponential correlation function $\rho(h) = \exp(-(\|h\|/\phi)^\alpha)$, $\phi > 0$ and $0 < \alpha \leq 2$.
- I2: Brown-Resnick model with semivariogram $\gamma(h) = (\|h\|/\phi)^\alpha$, $\phi > 0$ and $0 < \alpha \leq 2$.
- I3: Extremal-t model with an exponential correlation functions $\rho(h) = \exp(-\|h\|/\phi)$, $\phi > 0$.

Class II: consists of stationary max stable mixture models II1-II2.

- II1: max stable mixture model which combines I1 and I2.
- II2: max stable mixture model which combines I3 and I2.
- II3: max stable mixture model which combines I2 and isotropic Smith model with a covariance matrix $\Sigma = \phi\,\mathrm{Id}_2$

Class III: consists models III1-III3 which are the non stationary max stable mixture models as defined in Equation 16 for models II1-II3.
For all the models above, the univariate must be unit Fréchet. Therefore, we use the empirical distribution function to transform the data to a unit Fréchet distribution. After that, we estimate the dependence parameters using the composite likelihood approach, assuming equal weights. For model selection, we calculate CLIC. The results are summarized in Table 5, where ϕ, α, df and π represent the range, smooth, degree of freedom and spatial mixing proportion parameters. The subscript index 1 (2) refers to the first (second) mixing model.
Regarding the results in Table 5, the best fitting model for the data is the max stable mixture with different spatial proportions for each cluster.

8.2 Rainfall over France

This subsection is devoted to the study of rainfall data in France.

8.2.1 Description of the data

This data is provided by Météo-France and represents the hourly precipitation recorded at 80 French monitoring stations. The data was measured during the fall season (September, October and November) over the period 1993 - 2021. Each station has 24 * 91 * 29 = 63336 observations. The distance between the stations ranges approximately from 12 km to 1248 km. The geographic locations of these stations were chosen to cover all the French metropolitan

Table 5 Summary of the fitted models based on the 30-day maxima of rainfall over the east coast of Australia. The bold number for CLIC indicates the lowest value.

Class I	ϕ	α	df					CLIC
I1	169.11	1.06	-					1871755
I2	134.23	0.63	-					1866939
I3	661.62	-	3.23					1865962
Class II	ϕ_1	α_1	df_1		π	ϕ_2	α_2	CLIC
II1	380.17	1.37	-		0.25	115.73	0.61	1865515
II2	1716.5	-	4.66		0.54	48.66	0.54	1865280
II3	224.04	0.77	-		0.88	28.74	-	1866104
Class III	ϕ_1	α_1	df_1	π_1	π_2	ϕ_2	α_2	CLIC
III1	100.02	0.34	-	0.11	0.53	90.11	0.44	1866179
III2	604.90	-	2.44	0.41	0.88	125.31	0.48	**1864845**
III3	224.77	0.77	-	0.89	0.87	53.26	-	1865982

regions. Figure 13 illustrates the geographic locations of the 80 stations. This data was studied by Bernard et al (2013) during the period 1993 - 2011.

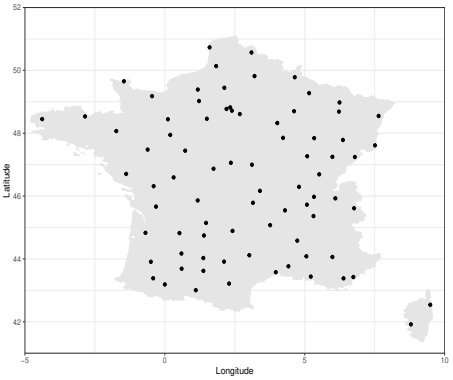

Fig. 13 Geographic locations of 80 stations over France region

8.2.2 Regionalizing the maxima for rainfall over France

In order to clustering the rainfall extremal dependence, we implemented the adapted spectral clustering on the data. Additionally, we studied the effect of the block size on the clustering map of extremal dependence. The block sizes we considered are monthly, 2 weeks, and weekly. The results obtained with the spectral clustering on this data for each block size are shown in Figure 14. The eigengap heuristic indicated that the number of clusters is 2 for all tested block sizes, as shown in the left panels of Figure 14. This was reinforced by the Silhouette Index, as is evident in the middle panels of Figure 14.

We will begin our discussion with the weekly block size since it was studied by Bernard et al (2013). Implementing spectral clustering on the data with this

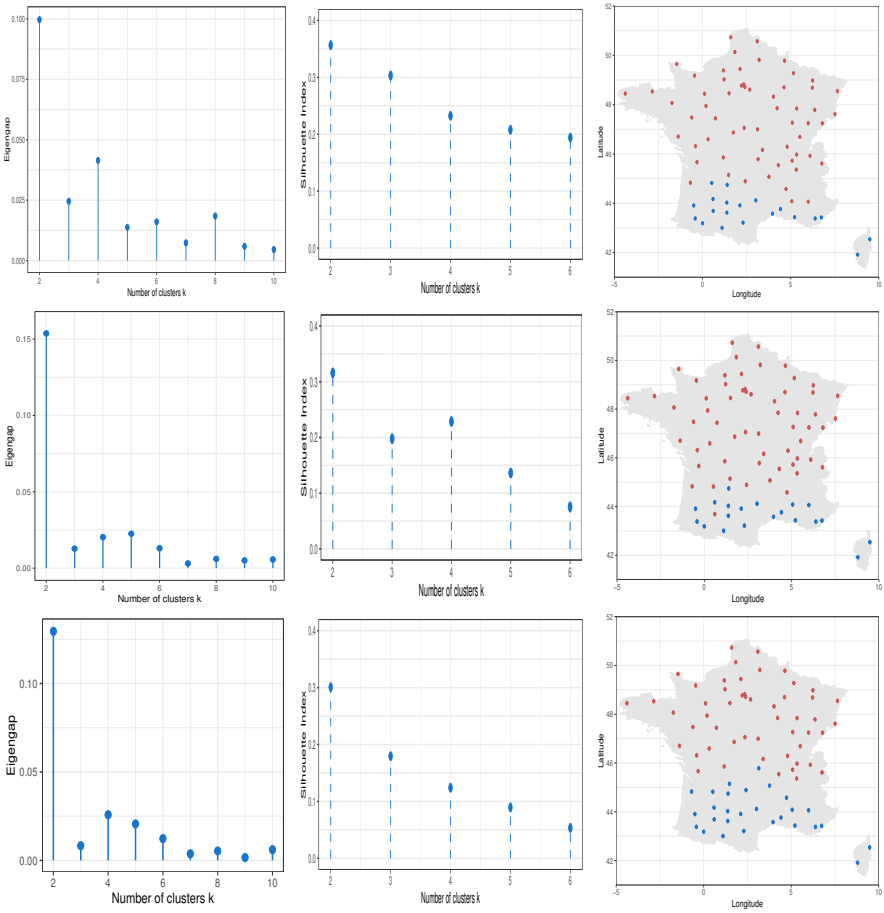

Fig. 14 The left column displays Eigengap values, the middle column shows the Silhouette Index against the number of clusters k, and the right column represents the spectral clustering result for the France rainfall data over the period of 1993-2021. The first, second, and third rows are for a monthly block size, a two-week block size, and a weekly block size, respectively.

block size divides France into two regional areas, north and south, along the Loire valley line. Each of these regions has a different extremal dependence, as explained in the right panel of the third row of Figure 14. This can be interpreted easily. The extreme rainfall in the north of France is produced by disturbances from the Atlantic, which affect Brittany, Paris and other areas in the north of France. In contrast, the extreme rainfall in the south of France is caused by the Mediterranean sea, which affects the coastal areas, particularly Cévennes and the Montagne Noire. The results are similar to those obtained

by Bernard et al (2013), where the selection criterion for the number of clusters indicated that $k = 2$. The locations clustering result is relatively close to Bernard et al (2013), where France was divided into north and south regions. This indicates that the behavior of the data has not changed since 2011.
The right panels of Figure 14 show the location clustering results for all tested block sizes. It is clear that the south region becomes smaller as the block size becomes larger. For this data, we can say that the size of the block may affect the clustering map.

8.2.3 Modelling the rainfall over France

Once again, we will check if the clustering of extremal dependence is valuable in modeling the maxima of rainfall over France. We will use the weekly block size and consider the same models proposed for Australian data, following the same modeling steps as in section 8.1.3. The results are summarized in Table 6. The lowest value for CLIC in Table 6 indicates that the best-fitting model for

Table 6 Summary of the fitted models based on the weekly maxima of rainfall over France. The bold number for CLIC indicates the lowest value.

Class I	ϕ	α	df					CLIC
I1	164.89	0.96	-					9900596
I2	86.58	0.55	-					9918058
I3	344.08	-	1.88					9890874
Class II	ϕ_1	α_1	df_1		π	ϕ_2	α_2	CLIC
II1	408.38	1.05	-		0.48	56.89	0.55	9886465
II2	100.28	-	1		0.53	90.15	0.47	9894430
II3	100.06	0.42	-		0.87	90	-	9909144
Class III	ϕ_1	α_1	df_1	π_1	π_2	ϕ_2	α_2	CLIC
III1	426.87	1.02	-	0.50	0.43	57.05	0.56	9885185
III2	840.75	-	1.90	0.59	0.49	30.38	0.55	**9884561**
III3	331.47	0.87	-	0.76	0.66	316.06	-	9895379

the France data is the max-stable mixture with different spatial proportions for each cluster. Although the difference in the value of π between the clusters in the north and in the south is not too large, it appears significant in modeling.

9 Discussion and Conclusion

Modeling environmental extreme events requires an understanding of the extremal dependence structure. In many studies, it is assumed that extremal dependence is fixed. However, this assumption may be incorrect, especially in large regions or areas with complex geographical or climatic patterns. Therefore, finding a method that can detect regions with similar extremal

dependence is valuable.

In this study, we combined spectral clustering with extremal concurrence probability to develop a simple clustering method for max-stable processes. Additionally, we modeled the entire region based on the regions resulting from spectral clustering. We validated our approach through a simulation study based on a non-stationary max-stable process. Then, we applied it to two environmental datasets.

The first dataset consists of daily rainfall data over the east coast of Australia. Spectral clustering divided the east coast area with respect to rainfall maxima into two regional clusters, where assuming fixed extremal dependence for each cluster is reasonable. We found that, for this data, the block size does not affect the clustering map.

The second dataset consists of hourly precipitation data over France. Spectral clustering divided the France region into two regional clusters: one in the north and the other in the south, each exhibiting similar extremal dependence. However, for this data, the clustering maps are not the same for all the tested block sizes. Therefore, we conclude that the block size can result in changes in regional clusters and should be taken into consideration.

Modeling these two datasets indicates that the best-fitted model for each of them is the max-stable mixture model with different spatial mixing proportions for each clustered region. This leads us to consider regionalizing extremal dependence to help model the non-stationary max-stable mixture model, instead of relying on covariates that are not always provided to model spatial proportions.

Finally, despite the simplicity of the adapted clustering algorithm, it is powerful. As a future direction of this study, one can study other variables like temperature. Another direction is to test the efficiency of this clustering algorithm when applied to a larger region, such as the whole of Australia or the continent of Europe.

Acknowledgments We acknowledge partial support from the PAUSE program, which is operated by the Collège de France. Also, we would like to thank the French meteorological service (Météo-France) for providing us with the data.

References

Abu-Awwad A, Maume-Deschamps V, Ribereau P (2020) Fitting spatial max-mixture processes with unknown extremal dependence class: an exploratory analysis tool. Test 29(2):479–522. https://doi.org/10.1007/s11749-019-00663-5

Ahmed M, Maume-Deschamps V, Ribereau P, et al (2017) A semi-parametric estimation for max-mixture spatial processes. arXiv preprint arXiv:171008120

Ahmed M, Maume-Deschamps V, Ribereau P (2022) Recognizing a spatial extreme dependence structure: A deep learning approach. Environmetrics 33(4):e2714. https://doi.org/10.1002/env.2714

Bacro JN, Gaetan C, Toulemonde G (2016) A flexible dependence model for spatial extremes. Journal of Statistical Planning and Inference 172:36–52. https://doi.org/10.1016/j.jspi.2015.12.002

Bador M, Naveau P, Gilleland E, et al (2015) Spatial clustering of summer temperature maxima from the cnrm-cm5 climate model ensembles & e-obs over europe. Weather and climate extremes 9:17–24. https://doi.org/10.1016/j.wace.2015.05.003

Bernard E, Naveau P, Vrac M, et al (2013) Clustering of maxima: Spatial dependencies among heavy rainfall in france. Journal of climate 26(20):7929–7937. https://doi.org/10.1175/JCLI-D-12-00836.1

Bouveyron C, Celeux G, Murphy TB, et al (2019) Model-based clustering and classification for data science: with applications in R, vol 50. Cambridge University Press

Brown BM, Resnick SI (1977) Extreme values of independent stochastic processes. Journal of Applied Probability 14(4):732–739. https://doi.org/10.2307/3213346

Castruccio S, Huser R, Genton MG (2016) High-order composite likelihood inference for max-stable distributions and processes. Journal of Computational and Graphical Statistics 25(4):1212–1229

Chung FR (1997) Spectral graph theory, vol 92. American Mathematical Soc.

Cooley D, Naveau P, Poncet P (2006) Variograms for spatial max-stable random fields, Springer New York, New York, NY, pp 373–390. https://doi.org/10.1007/0-387-36062-X_17

De Haan L (1984) A spectral representation for max-stable processes. The annals of probability 12(4):1194–1204

De Haan L, Pereira TT (2006) Spatial extremes: Models for the stationary case. The annals of statistics 34(1):146–168. https://doi.org/10.1214/009053605000000886

Dombry C, Ribatet M, Stoev S (2018) Probabilities of concurrent extremes. Journal of the American Statistical Association 113(524):1565–1582. https://doi.org/10.1080/01621459.2017.1356318

Hagen L, Kahng AB (1992) New spectral methods for ratio cut partitioning and clustering. IEEE transactions on computer-aided design of integrated circuits and systems 11(9):1074–1085. https://doi.org/10.1109/43.159993

Huser R, Genton MG (2016) Non-stationary dependence structures for spatial extremes. Journal of agricultural, biological, and environmental statistics 21(3):470–491. https://doi.org/10.1007/s13253-016-0247-4

Kaufman L, Rousseeuw PJ (1990) Finding Groups in Data: An Introduction to Cluster Analysis. Wiley

Mohar B (1997) Some applications of Laplace eigenvalues of graphs, Springer Netherlands, Dordrecht, pp 225–275. https://doi.org/10.1007/978-94-015-8937-6_6

Mohar B, Alavi Y, Chartrand G, et al (1991) The laplacian spectrum of graphs. Graph theory, combinatorics, and applications 2(871-898):12

Ng A, Jordan M, Weiss Y (2001) On spectral clustering: Analysis and an algorithm. Advances in neural information processing systems 14

Opitz T (2013) Extremal t processes: Elliptical domain of attraction and a spectral representation. Journal of Multivariate Analysis 122:409–413. https://doi.org/10.1016/j.jmva.2013.08.008

Padoan SA, Ribatet M, Sisson SA (2010) Likelihood-based inference for max-stable processes. Journal of the American Statistical Association 105(489):263–277

Parodi P (2012) Computational intelligence with applications to general insurance: a review: I–the role of statistical learning. Annals of Actuarial Science 6(2):307–343. https://doi.org/10.1017/S1748499512000036

Ribatet M (2017) Modelling spatial extremes using max-stable processes

Ribatet M, Dombry C, Oesting M (2016) Spatial extremes and max-stable processes. Extreme Value Modeling and Risk Analysis: Methods and Applications pp 179–194

Richards J, Wadsworth JL (2021) Spatial deformation for nonstationary extremal dependence. Environmetrics 32(5):e2671. https://doi.org/10.1002/env.2671

Rousseeuw PJ (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics 20:53–65

Saunders K, Stephenson A, Karoly D (2021) A regionalisation approach for rainfall based on extremal dependence. Extremes 24(2):215–240. https://doi.org/10.1007/s10687-020-00395-y

Schlather M (2002) Models for stationary max-stable random fields. Extremes 5(1):33–44. https://doi.org/10.1023/A:1020977924878

Schlather M, Tawn JA (2003) A dependence measure for multivariate and spatial extreme values: Properties and inference. Biometrika 90(1):139–156. https://doi.org/10.1093/biomet/90.1.139

Shi J, Malik J (2000) Normalized cuts and image segmentation. IEEE Transactions on pattern analysis and machine intelligence 22(8):888–905. https://doi.org/10.1109/34.868688

Smith RL (1990) Max-stable processes and spatial extremes. Unpublished manuscript 205:1–32

Varin C, Vidoni P (2005) A note on composite likelihood inference and model selection. Biometrika 92(3):519–528

Von Luxburg U (2007) A tutorial on spectral clustering. Statistics and computing 17(4):395–416. https://doi.org/10.1007/s11222-007-9033-z

Wierzchoń ST, Kłopotek MA (2018) Modern algorithms of cluster analysis, vol 34. Springer