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. 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 a homogeneous extremal dependence. In addition, we propose an approach to model the entire region based on the clustered zones. 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 considered the 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 in 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 for 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. This region is characterized by mountain ranges, which significantly influence rainfall events and can give rise to complex rainfall patterns, as known. Similarly, when considering rainfall data in France, which features extensive coastal areas, extreme rainfall can be produced by disturbances from the Mediterranean Sea and the Atlantic Ocean.

Searching for a max-stable model able to capture the changes in 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 presented a non-stationary extremal-t model, which is 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, named a max-stable mixture. In brief, this model represents a mix of two max-stable models where the mixing proportion varies spatially and is modeled as a function depending on covariates. This model could capture different extremal dependencies in different spatial regions. However, this model requires prior knowledge of relevant covariates.

Recently, clustering has been used to create regionalizations of extreme events, where it is an unsupervised machine-learning tool widely used in data analysis to identify subgroups with similar features. It has a wide range of applications in several fields such as computer science, statistics, biology, and climate science (see for example Dhanachandra et al (2015), Talebi et al (2020), Wright (2024) and Gaetan et al (2024)).

In the spatial extremes field, only a few studies used clustering to partition an entire region into homogeneous sub-regions. For instance, Bernard et al (2013) presented a novel clustering algorithm for maxima, using the F-madogram introduced by Cooley et al (2006). By merging the F-madogram with a partitioning around medoids (PAM) algorithm, they clustered 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 using hierarchical clustering with F-madogram. They applied their proposed algorithm to rainfall stations in Australia. Then, they fitted a stationary max-stable model for each resulting region. In this case, it's not possible to model the dependence between two locations belonging to different clusters. However, in all the above-presented literature there is no clear idea about which model for the entire region can represented using these clusters.

Regarding the regionalization for extremal dependence modeling, Castro-Camilo and Huser (2020) proposed a spatial modeling framework for a non-stationary dependence structure of extremes. They presented a non-stationary exponential factor copula model with spatially varying parameters. Also, they suggested a censored local likelihood by fitting a stationary version of this model locally, assuming that the stationarity might be valid in small regions. This framework is applied to model the dependence structure of extreme precipitation over the U.S. region. Majumder and Reich (2023) introduced a non-stationary process mixture model, consisting of a mix between a max-stable process with a transformed Gaussian process, where the mixing proportion was allowed to vary spatiotemporally. They considered two hydrologic regions in the studied area so that each has its mixing proportion modeled as a function of the yearly precipitation at that region, resulting in a mixing proportion for the model varying over space and time. They fitted their model to annual streamflow maxima in the central United States. Hector and Reich (2024) presented a spatial regionalization approach to fit the Brown-Resnick model in high dimension. This approach consists of partitioning the whole spatial domain into subsets of the spatial domain and estimating its marginal and dependence parameters locally using censored pairwise composite likelihood, then integrating them using a modified Generalized Method of Moments(GMM). They showed this approach leads to less bias in parameter estimations compared to the full-data approach. They applied this procedure to modeling the streamflow data in the United States.

Our contribution 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 strong and homogeneous extremal dependence, which may reduce the possibility of non-stationary. We demonstrated the applicability of this algorithm in clustering the non-stationary max-stable mixture process presented by Huser and Genton (2016) into k regional clusters.

The second path involves using the clustered regions to model the entire area. For this purpose, we proposed a composite 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 machine learning technique used for data analysis to cluster the data points into groups based on similarity. It depends 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. It used the eigenvalue decomposition to reduce the high-dimensional data to a lower-dimensional space. Furthermore, it can handle different similarity measures, making it flexible and adaptable to different kinds of data. Also, it does not make any assumptions about the shape or size of clusters.
Spectral clustering considers the dataset like 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, which is the weight assigned to each edge. It is important to note that since the graph is undirected, the similarity matrix should be symmetric. If $w_{ij} = 0$, 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, denoted by 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, and 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 considers the aim of clustering. Then, construction of the similarity graph to model the neighborhood relation among the data points $P_1, \cdots, P_N$. There are different ways, which 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, $w_{ij} > \varepsilon$, ε is a pre-defined non-negative real number. Usually, this graph is unweighted.
   - k**-nearest neighbor graphs:** In this graph, the distance between each pair of vertices represents the Euclidean distance. 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 weight represents the similarity w_{ij}. Here, k is a pre-defined integer number that controls the neighborhood relationship among data points.
   - **The fully connected graph:** In this graph, each vertex is connected to all other vertices by edges, and the weights of these edges represent the similarities w_{ij}. This type of graph is beneficial only if the similarity function can model the neighborhood relation among the data points. The commonly used similarity function is the Gaussian similarity function 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. It is an essential tool to identify clusters in the data using spectral clustering. It characterizes the connectivity of a graph, where it can capture the relationships between the nodes and identify the most connected nodes. There are two definitions for this matrix, which are described 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 in 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 functions measure the quality of the partition of a graph into sub-graphs (clusters). Let C_l be a subset of vertices i.e $C_l \subset V, l = 1, \cdots, k$ and its complement $\overline{C}_l := V \backslash C_l$, the Ratio cut function (Rcut)(Hagen and Kahng (1992)) is defined as:

$$\text{Rcut}(C_1, \cdots, C_k) = \sum_{l=1}^{k} \frac{\text{cut}(C_l, \overline{C}_l)}{|C_l|}, \tag{2}$$

where

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

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

In this function, the size of a subset C_l is measured by its number of vertices. Using the unnormalized Laplacian matrix L in spectral clustering is equivalent to minimizing the Ratio cut function.

In contrast, the Normalized cut function(Ncut)(Shi and Malik (2000)) is defined as:

$$\text{Ncut}(C_1, \cdots, C_k) = \sum_{l=1}^{k} \frac{\text{cut}(C_l, \overline{C}_l)}{\text{vol}(C_l)}, \tag{3}$$

where

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

In the Normalized cut function, the size of a subset C_l is measured by the weights of its edges. Using the normalized Laplacian matrix L^{norm} in spectral clustering is equivalent to minimizing the Normalized cut function. For more details about the Laplacian matrix 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)).

After computing the Laplacian matrix, calculate its eigenvalues and eigenvectors. The eigenvectors constitute a low-dimensional data representation, where the clusters are more-separated. Typically, the k eigenvectors corresponding to the k smallest eigenvalues construct a k-dimensional representation of the data, as they capture the structure of the graph and the main features of the data (see Wierzchoń and Kłopotek (2018)). Reducing dimensions can reveal hidden patterns in the data that may be difficult to distinguish in higher dimensions.

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

Spectral clustering has a specific heuristic method to choose 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 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 $V_{\mathbf{S}}(1, \cdots, 1) = \theta(s_1, \cdots, s_m)$ where $\theta(.)$ is the extremal coefficient that provides a summary of the dependence structure (see Schlather and Tawn (2003) and Smith (1990)). Particularly, when $\mathbf{S} = \{s, s'\} \subset \mathcal{S}$ and s, s' are separated by a spatial lag $h : h = s - s'$, the extremal coefficient satisfies $\theta(s, s') = V_{\mathbf{S}}(1, 1) \in [1, 2]$. The lower bound corresponds to the variables $X(s)$ and $X(s')$ that are completely dependent, while the upper bound corresponds to the case where they are independent. For simplicity $\theta(s, s')$ can be denoted by $\theta(h)$.

Several models for max-stable processes have been presented based on this spectral representation, including the Smith model (see Smith (1990)), the Schlather model (see Schlather (2002)), the Brown-Resnick model (see Kabluchko et al (2009)), and the Extremal-t model (see Opitz (2013)). The details about the construction of these models and their parametric forms are shown in Appendix A.

3.2 Extremal concurrence probability

Other indices to measure the dependence between extremes exist in the literature. Dombry et al (2018) introduced the extremal concurrence probability for measuring the extremal dependence, specially 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 at all locations occurred simultaneously. It can be interpreted as the chance of a single extreme event affecting all the spatial locations and being responsible for the record maximum. The extremal concurrence probability 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:

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

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 $p_r(s,s') = 0$ if and only if $X(s)$ and $X(s')$ are independent, and $p_r(s,s') = 1$ if and only if $X(s)$ and $X(s')$ 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 $s_j, j = 1, 2, \cdots, m$. To apply spectral clustering in extremal dependence context, locations $s_1, \cdots, s_m$ are considered 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 consists 15 locations. For viewing purposes, the self-edges are not shown. Selecting an appropriate met-

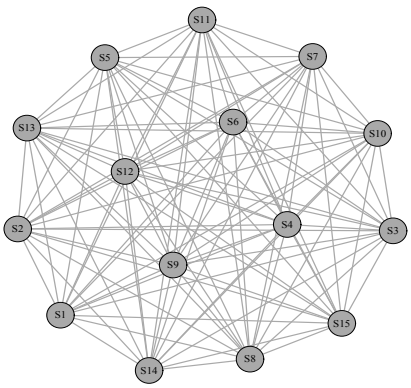

Fig. 1 Fully connected graph with 15 vertices. Each vertex represents a location in the max-stable process

ric to construct the similarity matrix is essential in the spectral clustering algorithm, especially when using a fully connected graph. For our aim of clustering, choosing an extremal dependence measure that can accurately model the neighborhood relations among the locations is very important. 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 Equation (12).

After constructing the similarity matrix CP, 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). Then, the spectrum λ and the eigenvectors of L^{norm} are computed. The eigenvectors $q_1, \cdots, q_k$ constitute a k-dimensional data representation. 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 $U \in \mathbb{R}^{m\times k}$. According to Ng et al (2001), this last step improves the performance of the clustering algorithm. Instead of using k-means usually used at this step, we used a Gaussian Mixture Model (GMM) to cluster the rows of U. GMM clusters the data points based on probability distribution, considering that the data points 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 (ASC)

Require: The similarity matrix $CP \in \mathbb{R}^{m\times m}$ constructed according to Equation (12), and the number of clusters k.

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 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 $Q \in \mathbb{R}^{m\times k}$.
4: Normalize the rows of Q to norm 1, resulting the matrix $U \in \mathbb{R}^{m\times k}$: $U_{jl} = Q_{jl}/(\sum_l Q_{jl}^2)^{\frac{1}{2}}$, $j = 1, \cdots, m$, $l = 1, \cdots, k$.
5: Consider each row of U as a point in $\mathbb{R}^k$ and implement Gaussian Mixture Model (GMM) to cluster them into k clusters.
6: Assign the location s_j to cluster l if and only if row j of the matrix U is assigned to cluster l.

5 Applicability of the adapted spectral clustering (ASC)

The adapted spectral clustering algorithm, which uses the extremal concurrence probability as a similarity matrix, aims to identify regions with high concurrence probability between locations. As a result, we obtain regions where the locations exhibit a strong and homogenous extremal dependence since their pairwise extremal concurrence probabilities are similar in value. That can reduce the possibility that the extremal dependence will be non-stationary in the clusters.

We assume the spatial structure of locations is relatively homogeneous, and expect that the geographically closer locations exhibit strong extremal dependences and have a high extremal concurrence probability. So, as a preliminary step, we need to determine the spatial structures of the k clusters. We obtain these spatial structures by applying the partitioning around medoids (PAM) clustering algorithm proposed by Kaufman and Rousseeuw (1990), where its general steps are shown in Algorithm 2. We applied Algorithm 2, where data points and the distance matrix are represented by the locations and the Euclidean distances matrix between the spatial coordinates of these locations, respectively. As a result, we obtained k clusters, in which the locations are spatially closer when measured by their Euclidean distances.

We illustrate the application of the adapted 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

$$F(x_1, x_2;\ \Psi) = \mathbb{P}\{X(s) \leq x_1, X(s') \leq x_2\} = \exp\Big\{-V_{\mathbf{S}}^{mix}(x_1, x_2)\Big\} \qquad (13)$$

Algorithm 2 partitioning around medoids (PAM) clustering

Require: The distance matrix and the number of clusters k.
Ensure: Clusters $\{C_1, \cdots, C_k\}$.

1: Select k datapoints randomly and set them as the initial medoids.
2: Form the clusters by assigning each of the remaining datapoints to the closest medoid.
3: Update the medoids of each cluster, by choosing the datapoint that minimizes the total intracluster distance.
4: If there is no change in the medoids then the algorithm stop, otherwise goback to to step 2.

and

$$V_{\mathbf{S}}^{mix}(x_1, x_2) = V_{\mathbf{S}}^1\Big(\frac{x_1}{\pi(s)}, \frac{x_2}{\pi(s')}\Big) + V_{\mathbf{S}}^2\Big(\frac{x_1}{1-\pi(s)}, \frac{x_2}{1-\pi(s')}\Big) \tag{14}$$

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 has its mixing proportion i.e $\pi(s) = \pi_l$ for all $s \in C_l, l = 1, \cdots, k$. We will demonstrate in the simulation study (Section 7) the ability of ASC to recover these clusters. Assuming we have k clustered regions, depending on these clusters and letting $C_l, C_{l'}$ refer to two different clusters, we can rewrite the bivariate distribution in Equations 13 and 14 as

$$F(x_1, x_2;\ \Psi) = \begin{cases} \exp\Big\{-\big\{V_{\mathbf{S}}^1(\frac{x_1}{\pi_l} + \frac{x_2}{\pi_l})) + V_{\mathbf{S}}^2(\frac{x_1}{1-\pi_l} + \frac{x_2}{1-\pi_l})\big\}\Big\}, & \text{if } s, s' \in C_l \\ \exp\Big\{-\big\{V_{\mathbf{S}}^1(\frac{x_1}{\pi_l} + \frac{x_2}{\pi_{l'}})) + V_{\mathbf{S}}^2(\frac{x_1}{1-\pi_l} + \frac{x_2}{1-\pi_{l'}})\big\}\Big\}, & \text{if } s \in C_l, s' \in C_{l'} \end{cases} \tag{15}$$

For simplicity we will denote the distribution function components in Equation 15 by $F_{C_l}(x_1, x_2;\ \Psi)$ if $s, s' \in C_l$, and $F_{C_{ll'}}(x_1, x_2;\ \Psi)$ if $s \in C_l, s' \in C_{l'}$. Using the distribution in Equation 15, we can model the whole region under study depending on clustered regions.

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:

$$\mathcal{P}(\Psi) = \sum_{i=1}^{n} \sum_{j=1}^{m-1} \sum_{j'=j+1}^{m} \mathcal{W}_{jj'} \log \mathcal{L}(x_{ij}, x_{ij'};\ \Psi) =: \sum_{i=1}^{n} \mathcal{P}_i(\Psi) \tag{16}$$

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 (16) is consistent and asymptotically normally distributed, such that

$$\hat{\Psi} \sim \mathcal{N}(\Psi, \mathcal{I}(\Psi)^{-1}) \tag{17}$$

with

$$\mathcal{I}(\Psi) = \mathcal{H}(\Psi)\,\mathcal{J}(\Psi)^{-1}\,\mathcal{H}(\Psi) \tag{18}$$

where $\mathcal{I}(\Psi)$ is the sandwich 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

$$\hat{\mathcal{H}}(\Psi) = -\sum_{i=1}^{n} \nabla^2\,\mathcal{P}_i(\hat{\Psi}) \tag{19}$$

and

$$\hat{\mathcal{J}}(\Psi) = \sum_{i=1}^{n} \nabla\mathcal{P}_i(\hat{\Psi})\,\nabla\mathcal{P}_i(\hat{\Psi})^T \tag{20}$$

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 composite pairwise likelihood contribution $\mathcal{L}_{CL}(x_{ij}, x_{ij'};\ \Psi)$ for a pair $(x_{ij}, x_{ij'})$ as follows

$$\mathcal{L}_{CL}(x_{ij}, x_{ij'};\ \Psi) = \begin{cases} \partial^2_{12} F_{C_l}(x_{ij}, x_{ij'}), & \text{if } s_j, s_{j'} \in C_l \\ \\ \partial^2_{12} F_{C_{ll'}}(x_{ij}, x_{ij'}), & \text{if } s_j \in C_l, s_{j'} \in C_{l'} \end{cases} \tag{21}$$

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 $\text{CLIC} = -2\,\mathcal{P}(\hat{\Psi}) + 2\,\text{tr}(\hat{\mathcal{H}}(\hat{\Psi})^{-1}\,\hat{\mathcal{J}}(\hat{\Psi}))$. The lowest value of CLIC corresponds to the best-fit model. In this paper, we empirically estimated the matrices $\mathcal{H}(\Psi)$ and $\mathcal{J}(\Psi)$ as in Equations (19) and (20), respectively.

7 Simulation study

We consider two simulation cases to assess the accuracy of the adapted spectral clustering (ASC) and compare it with partitioning around medoids (PAM) clustering. In each case, the coordinates of the locations were generated randomly and uniformly in $[0, 1]^2$. We consider one case with 40 locations while the other with 80 locations, and in both cases the number of observations is fixed to 300 to have simulated data closed in the dimensions to the data used in the application (see Section 8).

- **Case 1:** This case consists of 40 locations with coordinates shown in the left panel of Figure 2. We assumed the existence of four regions, each with

different extremal dependence. We applied the PAM algorithm to the coordinates of the locations, as mentioned in Section 5, to have the spatial structures of these regions. The right panel of Figure 2 shows the spatial structures of the four regions. Then, we simulated a non-stationary process,

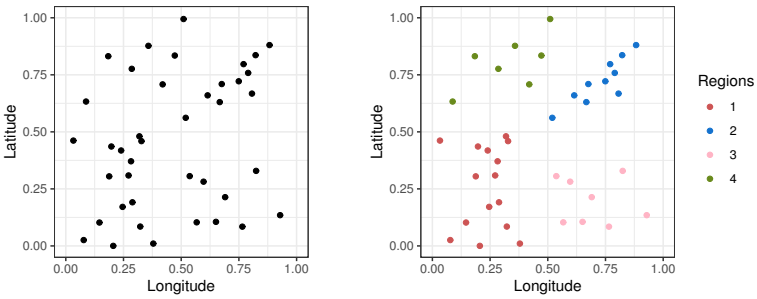

Fig. 2 The left panel shows the coordinates of the 40 locations considered in case 1, while the right panel displays the spatial structures of the assumed regions.

as described in Section 5. Denoted by M_1 , the model represents a mixture of the Schlather model with a powered exponential correlation function $\rho(h) = \exp(-(\|h\|/\phi_1)^{\alpha_1})$, and the Brown-Resnick model with a semivariogram $\gamma(h) = (\|h\|/\phi_2)^{\alpha_2}$. The parameters of the Schlather model are set to $\phi_1 = 0.2$ and $\alpha_1 = 1.5$. The parameters of the Brown-Resnick model are set to $\phi_2 = 0.8$ and $\alpha_2 = 0.5$. The spatially varying proportion was set as follows:

$$\pi(s) = \begin{cases} 0.2, & \text{if } s \in \text{region 1} \\ 0.4, & \text{if } s \in \text{region 2} \\ 0.6, & \text{if } s \in \text{region 3} \\ 0.8, & \text{if } s \in \text{region 4} \end{cases} \tag{22}$$

Afterward, we applied the adapted spectral clustering (ASC) to this process with the correct number of clusters k equals 4. The ASC outputs are displayed in the left panel of Figure 3. For comparison purposes, we applied the partitioning around medoids (PAM) for maxima introduced by Bernard et al (2013). It can done simply using Algorithm 2, and considering the pairwise F-madogram matrix as a distance matrix. The right panel of Figure 3 displays the PAM outputs.

To measure the ability of ASC and PAM to correctly group the simulated locations for each region into the same cluster, we used the clustering purity measure (Schütze et al (2008)). It calculates the ability of a clustering method to recover known groups. The clustering purity measure can be applied even when the number of clusters k is not equal to the number of known groups. It works by assigning a label to each cluster based on the most common group, then summing the number of correct group labels in each cluster and dividing it by the total number of data points. The clustering purity measure is a value in the interval [0,1], where the higher value

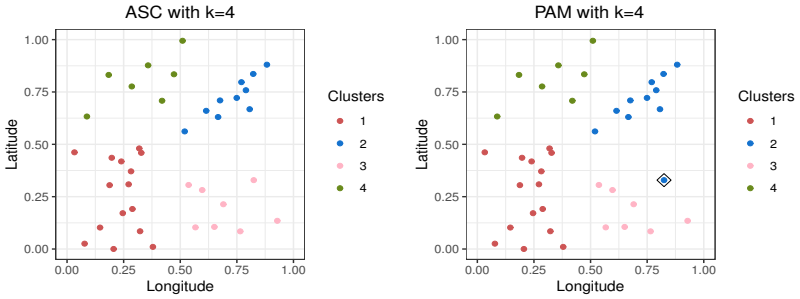

Fig. 3 The plot displays the clustering outputs from the adapted spectral clustering algorithm (left panel) and partitioning around medoids clustering algorithm (right panel). The black diamonds refer to the locations that are incorrectly clustered.

indicates better clustering performance, meaning that the algorithm can accurately identify clusters that correspond to the correct groups of data points. Depending on the specific clustering issue at hand, the formula for the purity measure is:

$$\text{purity}(\mathcal{C}, \mathcal{R}) = \frac{1}{m} \sum_{\ell=1}^{k} \max_{i=1,\cdots,g} |C_\ell \cap R_i| \tag{23}$$

Where $\mathcal{C} = \{C_1, \cdots, C_k\}$ is the set of identified clusters by ASC or PAM, $\mathcal{R} = \{R_1, \cdots, R_g\}$ is the set of simulated regions, $|C_\ell \cap R_i|$ is the number of locations in cluster ℓ being in region i and m is the total number of locations. Using Equation 23 we calculated the purity of ASC and PAM in clustering the process M_1, which is equal to 1 and 0.975, respectively.

To assess whether the number of observations n affects the accuracy of ASC and PAM algorithms, we implement the two algorithms with $n = 500$, 300, 100, and 50, respectively. For each number of observations, we simulated ten processes on different coordinates with the same parameters of M_1 as mentioned above. Table 1 shows the mean of purity for the ten simulations.

From Table 1, it is clear that ASC is more accurate than PAM, as its mean

Table 1 The mean of purity for the ten simulations of the process M_1 with the number of observations set to $n = 500$, 300, 100, and 50.

	$m = 40$							
	$n = 500$		$n = 300$		$n = 100$		$n = 50$	
no. of clusters	ASC	PAM	ASC	PAM	ASC	PAM	ASC	PAM
$k = 2$	0.975	0.860	0.945	0.925	0.925	0.903	0.848	0.763
$k = 3$	0.910	0.915	0.958	0.905	0.850	0.853	0.833	0.773
$k = 4$	0.903	0.938	0.943	0.928	0.800	0.785	0.738	0.705

of purity is higher in most cases. Also, the performance of the two algorithms decreases as the number of observations decreases. This may be due to the fact that ASC and PAM depends on empirical extremal dependence measures, which need enough observations to be estimated well.

We assess the ability of the proposed composite pairwise likelihood, as defined in Section 6, Equation 21, to estimate the parameters of model M_1 for modeling the whole region based on clusters. We simulate $N = 100$ from the non-stationary max-stable mixture processes (model M_1) on the same coordinate shown in the left panel of Figure 2, with the same conditions as explained previously, and the number of observations equals 300. Then, we use Equation 21 to obtain the estimator $\hat{\Psi}$. After that, we create boxplots of the errors of the estimated parameters $\hat{\Psi} - \Psi$ and calculate the mean estimate and the root mean square error (RMSE) for each estimated parameter. Let $\hat{\Psi}_i$ denoted the ith estimation, then

$$\text{RMSE} = \left[\frac{1}{N}\sum_{i=1}^{N}(\hat{\Psi}_i - \Psi)^2\right]^{1/2} \tag{24}$$

The boxplots of the errors in the estimated parameters are displayed in Figure 4, and the performance metrics values (mean estimates and RMSE) are shown in Table 2. From Figure 4, it is clear that the proposed com-

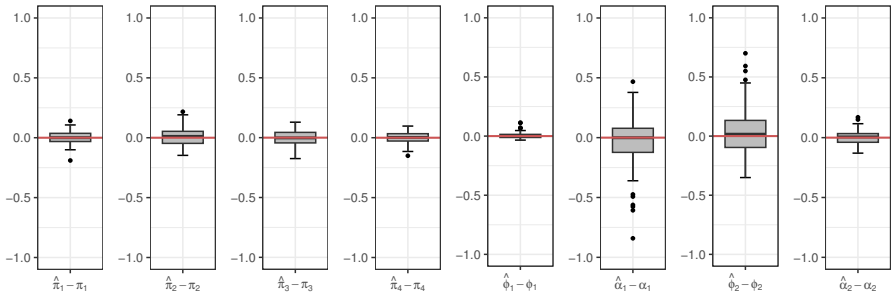

Fig. 4 Boxplots displaying $(\hat{\Psi} - \Psi)$ of the estimated parameters using composite pairwise likelihood (Equation 21) based on 100 simulations of model M_1 with parameters $\pi_1 = 0.2$, $\pi_2 = 0.4$, $\pi_3 = 0.6$, $\pi_4 = 0.8$, $\phi_1 = 0.2$, $\alpha_1 = 1.5$, $\phi_2 = 0.8$ and $\alpha_2 = 0.5$. The number of observations fixed at 300. The red horizontal line represents the zero value.

posite pairwise likelihood performs well when estimating the parameters for M_1 in the whole region depending on clustered regions. The variation is low except for α_1 and ϕ_2. Also, the mean estimates are nearly equal to the true parameter values, as explained in Table 2, indicating that the estimates are unbiased. Furthermore, the RMSE has reasonable values.

The main problem is to choose the correct number of clusters k. It is a hard issue, especially when the data points are not clearly separated. In general, the clustering validation metrics assess how well the clustering algorithm

Table 2 The performance metrics (mean estimates and RMSE) of the composite pairwise likelihood (Equation 21) based on 100 simulations of model M_1 with parameters $\pi_1 = 0.2$, $\pi_2 = 0.4$, $\pi_3 = 0.6$, $\pi_4 = 0.8$, $\phi_1 = 0.2$, $\alpha_1 = 1.5$, $\phi_2 = 0.8$ and $\alpha_2 = 0.5$.

	True parameters							
Performance metrics	π_1	π_2	π_3	π_4	ϕ_1	α_1	ϕ_2	α_2
Mean estimate	0.20	0.41	0.60	0.80	0.20	1.47	0.84	0.50
RMSE	0.05	0.08	0.06	0.05	0.03	0.21	0.2	0.06

has partitioned the locations into meaningful clusters, and it does not consider the overall process fit. We proposed to fit the considered model with a different number of clusters k and choose k that corresponds to the best CLIC. Table 3 summarized the results of fitting model M_1 consedering different numbers of clusters k on a simulated process from M_1 with k equals 4 as described at the beginning of case1. The lowest value of CLIC in Table 3 corresponds to k equals 4, which is the correct number of considered regions in the simulated process M_1.

Table 3 Summary of fitting M_1 model with different number of clusters k on the simulated process from M_1 with $k = 4$ which described in case 1. The bold number for CLIC indicates the lowest value

no. of clusters	$\hat{\pi}_1$	$\hat{\pi}_2$	$\hat{\pi}_3$	$\hat{\pi}_4$	$\hat{\phi}_1$	$\hat{\alpha}_1$	$\hat{\phi}_2$	$\hat{\alpha}_2$	CLIC
$k = 2$	0.17	0.58	-	-	0.20	1.78	0.72	0.45	1848583
$k = 3$	0.17	0.59	0.55	-	0.20	1.77	0.74	0.44	1848524
$k = 4$	0.17	0.42	0.57	0.78	0.21	1.68	0.86	0.42	**1846464**

- **Case 2:** In this case, we consider 80 locations with coordinates shown in the left panel of Figure 5. Here, we assumed the existence of three regions, each with different extremal dependence. So, we applied the PAM algorithm to the coordinates of the locations to construct the spatial structures of these regions, where they are shown in the right panel of Figure 5. We used these regions to simulate a non-stationary process. We mixed the Extremal-t model with an exponential correlation function $\rho(h) = \exp(-\|h\|/\phi_1)$, with a Brown-Resnick model with a semivariogram $\gamma(h) = (\|h\|/\phi_2)^{\alpha_2}$, as mentioned in Section 5, we denoted this model by M_2. In this model, the parameters of the Extremal-t are set to $\phi_1 = 0.3$ and $df_1 = 5$, where df_1 refers to the degree of freedom parameter. While the parameters of the Brown-Resnick model are set $\phi_2 = 0.9$ and $\alpha_2 = 1.5$. The spatially varying

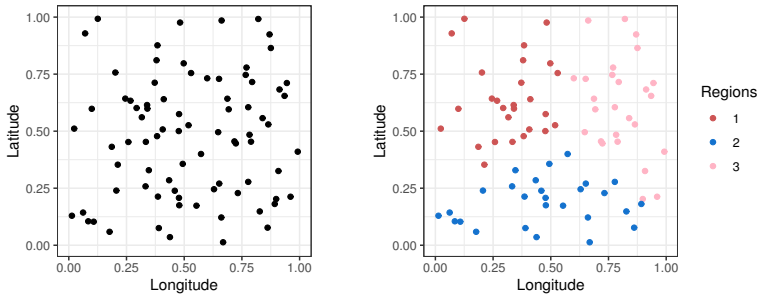

Fig. 5 The left panel shows the coordinates of the 80 locations considered in case 2, while the right panel displays the spatial structures of the assumed regions.

proportion was set as follows:

$$\pi(s) = \begin{cases} 0.1, & \text{if } s \in \text{region 1} \\ 0.4, & \text{if } s \in \text{region 2} \\ 0.7, & \text{if } s \in \text{region 3} \end{cases} \tag{25}$$

We applied the ASC and PAM (Bernard et al (2013)) algorithms to this process with the correct number of clusters k equals 3. The ASC and PAM outputs are displayed in the left and right panels of Figure 6, respectively. Figure 6 shows that ASC is more accurate than PAM in clustering the pro-

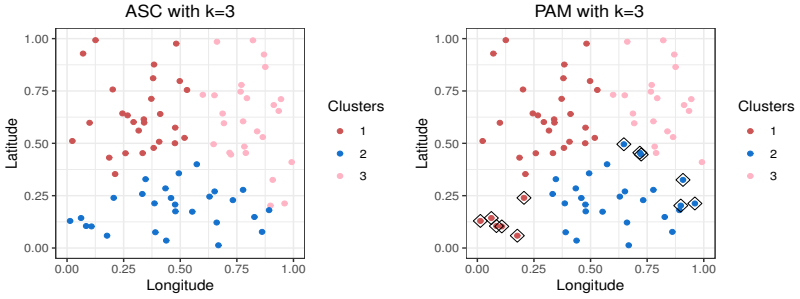

Fig. 6 The plot displays the clustering outputs from the adapted spectral clustering algorithm (left panel) and partitioning around medoids clustering algorithm (right panel). The black diamonds refer to the locations that are incorrectly clustered.

cess M_2. According to Equation 23, the purity of ASC equals one while it equals 0.85 for PAM. This conforms with the discussion of Saunders et al (2021) about the PAM algorithm (Bernard et al (2013)) as they demonstrated that PAM is sensitive to the spatial density of locations. The medoids of each cluster are biased toward regions of higher density of locations, which leads to the locations that are far away being clustered incorrectly.

We assess the accuracy of ASC and PAM algorithms considering different

numbers of observations, $n = 500$, 300, 100, and 50, respectively, in Table 4. For each number of observations, the value in Table 4 represents the mean of purity for ten simulations of the process M_2 with the same parameters as mentioned above, each simulation done on different coordinates. Compar-

Table 4 The mean of purity for the ten simulations of the process M_2 with the number of observations set to $n = 500$, 300, 100, and 50.

	$m = 80$							
	$n = 500$		$n = 300$		$n = 100$		$n = 50$	
no. of clusters	ASC	PAM	ASC	PAM	ASC	PAM	ASC	PAM
$k = 2$	0.916	0.688	0.900	0.691	0.816	0.714	0.786	0.693
$k = 3$	0.943	0.815	0.954	0.771	0.944	0.806	0.843	0.753
$k = 4$	0.931	0.804	0.926	0.765	0.856	0.741	0.864	0.731

ing the results in Table 4 and Table 1 confirms that the PAM is not accurate enough when clustering regions with highly dense locations. Also, it is clear from Table 4 that ASC outperforms PAM, especially with large n.

To assess the performance of the composite pairwise likelihood, we simulated $N = 100$ from the non-stationary process M_2 on the same coordinate shown in the left panel of Figure 5, with the same regions and parameters as stated previously. Also, the number of observations equals 300. Afterward, we estimate the parameters using Equation 21. Figure 7 displays the boxplots of the errors in the estimated parameters, while Table 5 displays the performance metrics values (mean estimates and RMSE).

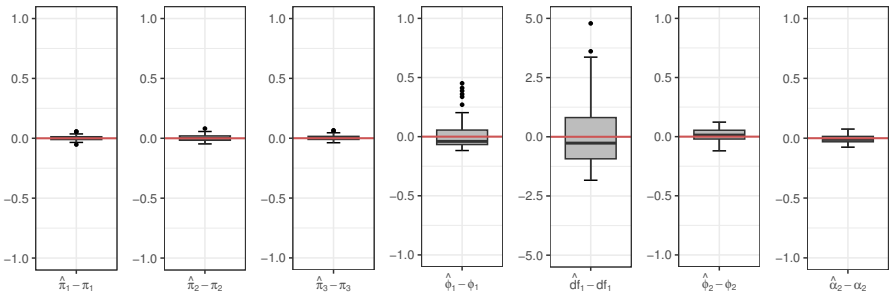

Fig. 7 Boxplots displaying $(\hat{\Psi} - \Psi)$ of the estimated parameters using composite pairwise likelihood (Equation 21) based on 100 simulations of model M_2 with parameters $\pi_1 = 0.1$, $\pi_2 = 0.4$, $\pi_3 = 0.7$, $\phi_1 = 0.3$, $df_1 = 5$, $\phi_2 = 0.9$ and $\alpha_2 = 1.5$. The number of observations fixed at 300. The red horizontal line represents the zero value.

Figure 7 indicates that the proposed composite pairwise likelihood works well in estimating the parameters for M_2, except for df_1 where we observe a

Table 5 The performance metrics (mean estimates and RMSE) of the composite pairwise likelihood (Equation 21) based on 100 simulations of model M_2 with parameters $\pi_1 = 0.1$, $\pi_2 = 0.4$, $\pi_3 = 0.7$, $\phi_1 = 0.3$, $df_1 = 5$, $\phi_2 = 0.9$ and $\alpha_2 = 1.5$.

	True parameters						
Performance metrics	π_1	π_2	π_3	ϕ_1	df_1	ϕ_2	α_2
Mean estimate	0.10	0.40	0.70	0.32	5.07	0.91	1.49
RMSE	0.02	0.02	0.02	0.12	1.29	0.05	0.03

huge variability. That is normal since the degree of freedom of the Extremal-t model is known to be difficult to estimate. Furthermore, the estimates are unbiased, as shown in Table 5, and the estimations are accurate as the RMSE values are low.

Finally, to confirm that we can use CLIC to choose the best number of clusters, we fit model M_2 considering different numbers of clusters k on a simulated process from M_2 with k equals 3 as mentioned in case 2 previously, where the results are summarized in Table 6. The lowest value of CLIC in Table 6 indicates that the best number of clusters that better represent the data is equal to three. That is equivalent to the correct number of clusters in the simulated process.

Table 6 Summary of fitting M_2 model with different number of clusters k on the simulated process from M_2 with $k = 3$ which described in case 2. The bold number for CLIC indicates the lowest value

no. of clusters	$\hat{\pi}_1$	$\hat{\pi}_2$	$\hat{\pi}_3$	$\hat{\pi}_4$	$\hat{\phi}_1$	$\hat{df}_1$	$\hat{\phi}_2$	$\hat{\alpha}_2$	CLIC
$k = 2$	0.24	0.70	-	-	0.32	5.22	0.70	1.62	7688422
$k = 3$	0.10	0.43	0.72	-	0.29	4.34	0.79	1.58	**7662620**
$k = 4$	0.10	0.45	0.71	0.57	0.26	3.99	0.78	1.56	7675077

In the two simulated cases, ASC works well provided that the simulated locations have a spatial structure relatively homogeneous. It was able to produce clusters that satisfy the local stationarity of the process.

To assess whether the ASC is sensitive to the choice of mixture components, we tested it in different scenarios. Considering scenarios where the mixing components represent different models with equal range parameters. Also, cases where the mixing components were from the same model with unequal range parameters. In all of these scenarios, the ASC still works well. The results are available upon request.

Since the simulation results appear satisfactory, we can use this technique for regionalization 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 and modeling technique, and then 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 Figure 8. 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.

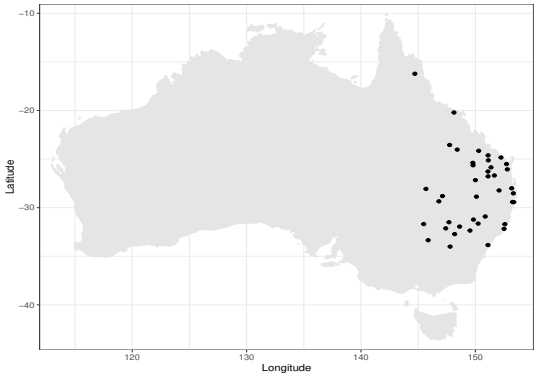

Fig. 8 Geographic locations of 40 stations on the east coast of Australia.

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 of the monthly maxima for the rainfall over east coast of Australia. We considered the number of clusters k equals 2,3 and 4, the outputs of ASC are shown in Figure 9. The clusters appear logical, as the spatially nearest locations were clustered together. The

resulting regions for each considered k are used in Section 8.1.3 to fit a non-stationary model. Accordingly, we can decide which k is better to represent the data depending on the overall model fit using CLIC.

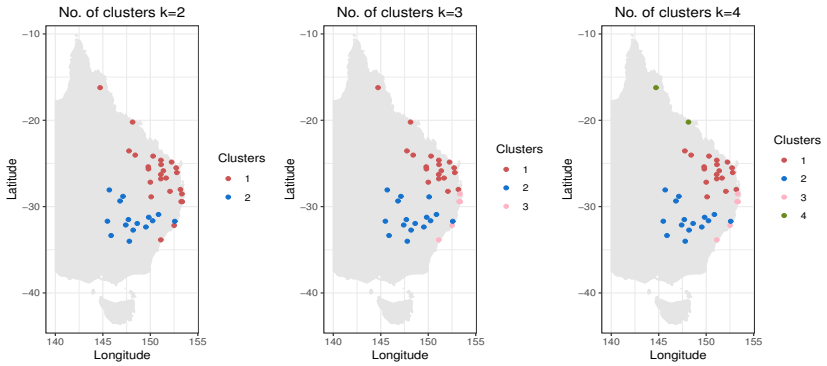

Fig. 9 represents the adapted spectral clustering (ASC) outputs for the monthly maxima of rainfall over the east coast of Australia during the period of 1972-2019.

8.1.3 Modelling the rainfall over east coast of Australia

For comparison purposes, we will proceed by fitting stationary models on the monthly maxima of rainfall over the east coast of Australia. We consider six arbitrary models that belong to two classes: max-stable models and stationary max-stable mixture models (constant mixing proportion). 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 $\Sigma = \phi\,\mathrm{Id}_2$

For all the fitted models, in this section and later in this paper, the univariate marginal distribution is fixed to a unit Fréchet. Therefore, we use

nonparametric transformation to unit Fréchet distribution defined as

$$X(s) \to \frac{-1}{\log\{F(X(s))\}}, \tag{26}$$

where $F(.)$ represents the empirical distribution function, i.e., $F(x) = \frac{1}{n+1}\sum_{i=1}^{n} \mathbb{1}_{X_i(s)<x}$. After that, we estimate the dependence parameters using the composite likelihood approach defined in Equation 16, assuming equal weights. For model selection, we calculate CLIC. The results are summarized in Table 7, where ϕ, α, df, and π represent the range, smoothness, degree of freedom, and spatial mixing proportion parameters, respectively. The subscript index 1 (2) refers to the first (second) mixing model. Regarding the

Table 7 Summary of fitting stationary models on the monthly maxima of the rainfall over the east coast of Australia. The bold number for CLIC indicates the lowest value.

Class I	$\hat{\phi}$	$\hat{\alpha}$	$\hat{df}$				CLIC
I1	169.11	1.06	-				1871755
I2	134.23	0.63	-				1866939
I3	661.62	-	3.23				1865962
Class II	$\hat{\phi}_1$	$\hat{\alpha}_1$	$\hat{df}_1$	$\hat{\pi}$	$\hat{\phi}_2$	$\hat{\alpha}_2$	CLIC
II1	380.17	1.37	-	0.25	115.73	0.61	1865515
II2	1716.34	-	4.66	0.54	48.66	0.54	**1865280**
II3	224.10	0.77	-	0.88	55.54	-	1866101

results in Table 7, the best fitting model for the data is II2. It refers to, the mixture models better representing the data than the max-stable models. To determine whether the clustering of extremal dependence is significant in modeling the monthly maxima of rainfall over the east coast of Australia, we fit non-stationary max-stable mixture models as defined in Equation 15, denoted by Class III.

Class III: consists models III1-III3 which are the non-stationary max-stable mixture models as defined in Equation 15 for models II1-II3, with k equals 2,3 and 4, and considering the regions which represented the outputs of ASC in Section 8.1.2.

The results of fitting Class III on the monthly maxima are summarized in Table 8. Looking at the CLIC values, the models belonging to model III2 better represent the data in general. Also, the CLIC values for $k = 3$ and $k = 4$ are very similar, so one can't say which is better. Furthermore, when $k = 4$, the values of the mixing proportions π_1 and π_4 are close. For that, a bootstrap study is done for model III2 with $k = 2, 3, 4$ to construct a confidence interval for the estimated parameters to evidence if an increase in the number of clusters is significant. Due to the computation complexity, we took 100 resamples in the bootstrap study and constructed a 90% confidence interval for the estimated parameters, reported in Table 8. There is a high intersection between

the confidence intervals of the mixing proportions π_1 and π_4 for the case when $k = 4$, which means their values can't considered different enough. There is no evidence for considering the case of $k = 4$. Accordingly, we choose the regionalization model III2 with $k = 3$ as the best model representing the data, as its mixing proportions can be considered different based on its confidence intervals. Furthermore, the three regions we obtained may be in concordance with the regions where the precipitations are similar in terms of extremes during El Niño/La niña.

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 regions. Figure 10 illustrates the geographic locations of the 80 stations. This data was studied by Bernard et al (2013) during the period 1993 - 2011.

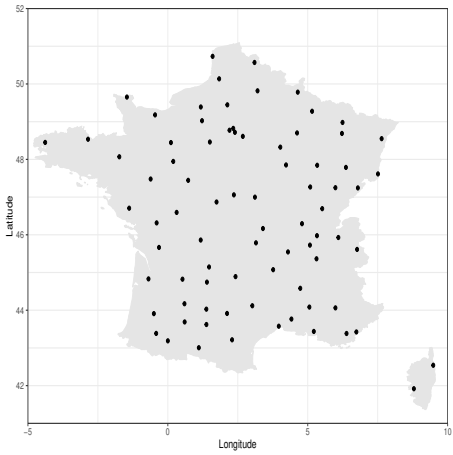

Fig. 10 Geographic locations of 80 stations over France region

8.2.2 Regionalizing the maxima for rainfall over France

We consider the weekly maxima of the rainfall over France. To construct a regionalization for the extremal dependence, we apply the ASC described in Section 4 Algorithm 1. The outputs of ASC considering different numbers of

Table 8 Summary of fitting Class III models on the monthly maxima of rainfall over the east coast of Australia. The 90% confidence intervals for the estimated parameters obtained from the bootstrap are reported between parentheses.

Model	k	$\hat{\pi}_1$	$\hat{\pi}_2$	$\hat{\pi}_3$	$\hat{\pi}_4$	$\hat{\phi}_1$	$\hat{\alpha}_1$	$\hat{df}_1$	$\hat{\phi}_2$	$\hat{\alpha}_2$	CLIC
III1	2	0.15	0.58	-	-	274.10	1.17	-	150.35	0.60	1865147
	3	0.24	0.45	0.02	-	390.66	1.26	-	129.54	0.60	1864448
	4	0.25	0.47	0.02	0.02	385.02	1.27	-	129.07	0.60	1864468
III2	2	0.41	0.88	-	-	604.85	-	2.44	125.29	0.47	1864845
		(0.30,0.54)	(0.77,0.95)	-	-	(509.92,737.95)	-	(1.97,2.81)	(88.31,156.39)	(0.40,0.55)	
	3	0.61	0.90	0.18	-	766.98	-	2.94	106.67	0.41	1863861
		(0.54,0.70)	(0.80,0.95)	(0.12,0.27)	-	(605.17,936.85)		(2.42,3.53)	(74.21,147.30)	(0.32,0.50)	
	4	0.62	0.91	0.18	0.57	766.76	-	2.94	104.09	0.40	1863859
		(0.54,0.69)	(0.81,0.96)	(0.12,0.27)	(0.47,0.66)	(604.35,1072.40)	-	(2.39,3.77)	(71.81,146.52)	(0.33,0.49)	
III3	2	0.89	0.87	-	-	224.78	0.77	-	53.26	-	1865982
	3	0.92	0.88	0.57	-	244.66	0.78	-	119.69	-	1865232
	4	0.90	0.87	0.57	0.99	247.86	0.79	-	122.95	-	1865217

clusters k, $k = 2, 3, 4$, are displayed in Figure 11. In Section 8.2.3, we use these regions to fit non-stationary models on the weekly maxima and choose which regionalization is better to represent the data depending on CLIC value.

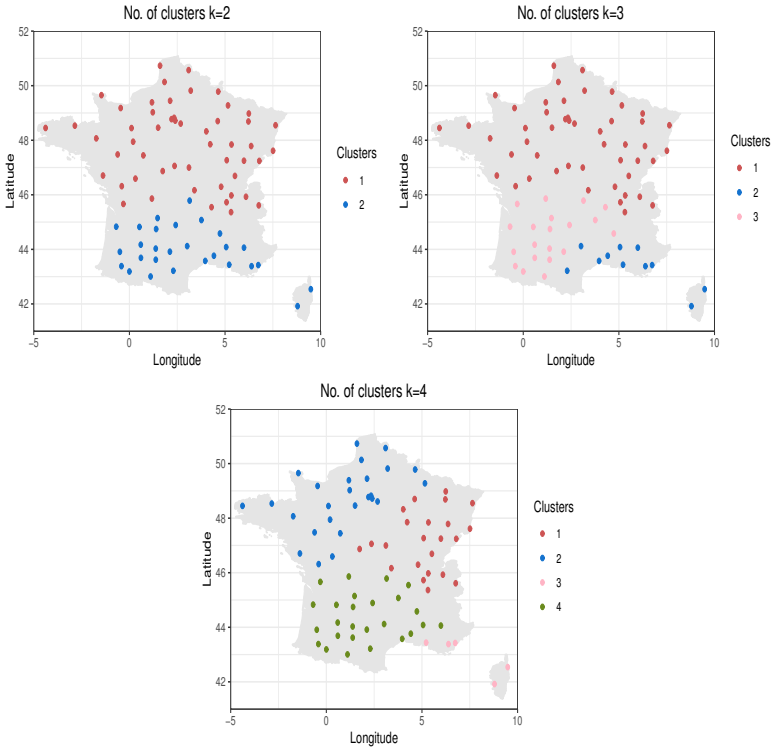

Fig. 11 represents the adapted spectral clustering (ASC) outputs for the weekly maxima of rainfall over France during the period of 1993-2021.

8.2.3 Modelling the rainfall over France

Again, we begin with fitting stationary models on the weekly maxima. We consider the same models presented in Section 8.1.3 and follow the same modeling steps. The summary of the results appears in Table 9. The lowest CLIC value indicates that the mixture model II2 is the best to represent the data. We fit non-stationary max-stable mixture models considering the regions obtained by ASC in Section 8.2.2. We fitted the same Class III models described in Section 8.1.3, and the results are summarized in Table 10. From Table 10, it is clear that the non-stationary models III2 are better than others to represent the data. For that, we construct 90% confidence intervals for their estimated parameters from bootstrap studies with 100 resamples. Confidence intervals are put between parentheses in Table 10. The lowest CLIC value indicates that the best-fitting model is III2 when k equals three. Although the value of the

Table 9 Summary of fitting stationary models on the weekly maxima of the rainfall over France. The bold number for CLIC indicates the lowest value.

Class I	$\hat{\phi}$	$\hat{\alpha}$	$\hat{df}$				CLIC
I1	164.89	0.96	-				9900596
I2	86.58	0.55	-				9918058
I3	344.08	-	1.88				9890874
Class II	$\hat{\phi}_1$	$\hat{\alpha}_1$	$\hat{df}_1$	$\hat{\pi}$	$\hat{\phi}_2$	$\hat{\alpha}_2$	CLIC
II1	408.39	1.05	-	0.48	56.90	0.55	9886465
II2	670.44	-	1.58	0.54	35.69	0.54	**9886158**
II3	323.26	0.89	-	0.73	289.11	-	9897158

mixing proportion π_1 and π_3 are close, we can consider it different, as their confidence intervals refer.

Furthermore, the regionalization with k equals three can interpreted easily. There is a specific rainfall pattern in the southeast of France with the Cévennes events. Also, the rainfall in the upper north part of France is affected by a huge influence of the Atlantic and a third region that mixes both patterns in the southwest.

9 Discussion and Conclusion

Modeling environmental extreme events requires an understanding of the extremal dependence structure. In many studies, the extremal dependence is assumed to be fixed. However, this assumption may be incorrect, especially in large regions or areas with complex geographical or climatic patterns. Therefore, finding a method to detect regions with homogeneous 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 the adapted 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. We found that the best-fitting mode for the extremal dependence of monthly maxima is a non-stationary max-stable mixture model considering three regional clusters obtained from the adapted spectral clustering. That supports the assumption that fixing the extreme dependence for each region is significant.

The second dataset consists of hourly precipitation data over France. The best regionalization obtained from the adapted spectral clustering divided France into three regional clusters: one in the north, one in the southeast, and the other in the southwest. Assuming a fixed extremal dependence for each region is significant when modeling the entire area.

Table 10 Summary of fitting Class III models on the weekly maxima of the rainfall over France. The 90% confidence intervals for the estimated parameters obtained from the bootstrap are reported between parentheses.

Model	k	$\hat{\pi}_1$	$\hat{\pi}_2$	$\hat{\pi}_3$	$\hat{\pi}_4$	$\hat{\phi}_1$	$\hat{\alpha}_1$	$\hat{df}_1$	$\hat{\phi}_2$	$\hat{\alpha}_2$	CLIC
III1	2	0.50	0.43	-	-	426.22	1.02	-	56.98	0.56	9885185
	3	0.52	0.35	0.46	-	431.25	0.98	-	58.66	0.57	9884714
	4	0.50	0.51	0.38	0.44	426.61	1.01	-	57.72	0.56	9885064
III2	2	0.59	0.49	-	-	840.59	-	1.9	30.39	0.55	9884561
		(0.54,0.62)	(0.44,0.53)	-	-	(698.23,1054.61)	-	(1.55,2.31)	(23.42,38.90)	(0.48,0.60)	
	3	0.60	0.41	0.53	-	855.29	-	1.92	29.77	0.54	9884021
		(0.55,0.63)	(0.34,0.47)	(0.48,0.56)	-	(698.16,1105.48)		(1.55,2.35)	(23.17,39.87)	(0.48,0.60)	
	4	0.59	0.59	0.43	0.51	839.66	-	1.89	30.38	0.55	9884443
		(0.53,0.62)	(0.53,0.64)	(0.31,0.50)	(0.47,0.54)	(699.09,1055.96)	-	(1.52,2.27)	(24.30,40.53)	(0.48,0.61)	
III3	2	0.76	0.66	-	-	331.47	0.87	-	316.07	-	9895379
	3	0.76	0.57	0.71	-	337.17	0.86	-	318.78	-	9894759
	4	0.77	0.75	0.56	0.69	332.40	0.86	-	301.42	-	9895094

Modeling these two datasets indicates that the best-fitted model is the max-stable mixture model with different spatial mixing proportions for each clustered region. That leads us to consider regionalizing extremal dependence to help model the non-stationary max-stable mixture process instead of relying on covariates, which are not always available to model spatial proportions. Finally, although the adapted clustering algorithm is simple, it is powerful. As a future direction for this study, one can study other variables like temperature. Another direction is to test the efficiency of the proposed technique 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.

Appendix A Parametric forms of the spatial max-stable models

In this appendix, we provided the parametric forms for some of the spatial max-stable models used in this paper. Let $\{X(s), s \in \mathcal{S}\}$ be a max-stable process with unit Fréchet margins, then depend on Equations (8) and (9), we can write the bivariate distribution function as

$$\mathbb{P}\{X(s) \leq x_1, X(s') \leq x_2\} = \exp\{-V_{\mathbf{S}}(x_1, x_2)\} \tag{A1}$$

here, $\mathbf{S}$ represent a set of two spatial locations : $\mathbf{S} = \{s, s'\} \subset \mathcal{S}$. The two locations s, s' are separated by spatial lag $h : h = s - s'$. Each max-stable model has its bivariate exponent function $V_{\mathbf{S}}(x_1, x_2)$, where its parametric form depends on the choice of the process $Y(s)$ in the spectral representation defined in Equation (6). we will show some of these models that are used in this paper.

1. **Smith model**
   Smith (1990) introduced this model, also called the Gaussian extreme value model. This model assumed $Y_i(s) = \phi_d(s - \mathcal{U}_i;\ \Sigma)$ in Equation (6), where $\{\mathcal{U}_i, i \geq 1\}$ represents a unit rate Poisson point process on $\mathcal{S} = \mathbb{R}^d$ and $\phi_d(.;\ \Sigma)$ represents the d-dimensional probability density function of a centered Gaussian distribution with covariance matrix Σ. In this model, the range of dependence and the degree of anisotropy are controlled by the matrix Σ. The bivariate exponent function of Smith model is defined as

$$V_{\mathbf{S}}(x_1, x_2) = \frac{1}{x_1}\Phi\left(\frac{\tau(h)}{2} + \frac{\log(\frac{x_2}{x_1})}{\tau(h)}\right) + \frac{1}{x_2}\Phi\left(\frac{\tau(h)}{2} + \frac{\log(\frac{x_1}{x_2})}{\tau(h)}\right) \tag{A2}$$

where $\tau(h) = \sqrt{h^\intercal \Sigma^{-1} h}$ represents the Mahalanobis distance between two locations s, s' and $\Phi(.)$ represents the standard Gaussian distribution. Depending on the relation between the exponent function and the extremal coefficient, the latter can written as

$$\theta(h) = 2\Phi\big(\frac{\tau(h)}{2}\big) \tag{A3}$$

2. **Schlather Model**
It is the so-called Extremal Gaussian Model introduced by Schlather (2002). Its spectral representation obtained by setting $\{Y(s), s \in \mathcal{S}\} \overset{\mathcal{D}}{=} \{\sqrt{2\pi}\max\{0, \varepsilon(s)\}, s \in \mathcal{S}\}$, where $\{\varepsilon(s), s \in \mathcal{S}\}$ is a standard Gaussian process with correlation function $\rho(.)$.
Its bivariate exponent function is defined as

$$V_{\mathbf{S}}(x_1, x_2) = \frac{1}{2}\big(\frac{1}{x_1} + \frac{1}{x_2}\big)\left[1 + \sqrt{1 - 2\frac{(\rho(h)+1)x_1 x_2}{(x_1+x_2)^2}}\right] \tag{A4}$$

So, the extremal coefficient for this model is defined as

$$\theta(h) = 1 + \left(\frac{1-\rho(h)}{2}\right)^{1/2} \tag{A5}$$

3. **Brown-Resnick model**
This model is presented by Kabluchko et al (2009). It is occasionally called the Geometric Gaussian model. The distribution assumption for $\{Y(s), s \in \mathcal{S}\}$ was set to be $\{Y(s), s \in \mathcal{S}\} \overset{\mathcal{D}}{=} \{\exp\{\epsilon(s) - \gamma(s)\}, s \in \mathcal{S}\}$, where $\{\epsilon(s), s \in \mathcal{S}\}$ represent a stationary centered Gaussian process with semivariogram $\gamma(.)$. The semivariogram is often considered as $\gamma(h) = (\|h\|/\phi)^\alpha$, where $\phi > 0$ and $0 < \alpha \leq 2$. Accordingly, the bivariate exponent function for this model is defined as

$$V_{\mathbf{S}}(x_1, x_2) = \frac{1}{x_1}\Phi\left(\frac{\sqrt{2\gamma(h)}}{2} + \frac{\log(\frac{x_2}{x_1})}{\sqrt{2\gamma(h)}}\right) + \frac{1}{x_2}\Phi\left(\frac{\sqrt{2\gamma(h)}}{2} + \frac{\log(\frac{x_1}{x_2})}{\sqrt{2\gamma(h)}}\right) \tag{A6}$$

As a result, its extremal coefficient is defined as

$$\theta(h) = 2\Phi\big(\frac{\sqrt{2\gamma(h)}}{2}\big) \tag{A7}$$

4. **Extremal-t model**
Opitz (2013) presented this model. It represents the generalization of the Schlather model, assuming that $\{Y(s), s \in \mathcal{S}\} \overset{\mathcal{D}}{=} \{c_{df}\max\{0, \varepsilon(s)\}^{df}, s \in \mathcal{S}\}$, $c_{df} = \sqrt{\pi}\, 2^{1-df/2}\big\{\Gamma\big(\frac{df+1}{2}\big)\big\}^{-1}$, $df \geq 1$, where $\{\varepsilon(s), s \in \mathcal{S}\}$ defined as in Schlather model, $\Gamma(.)$ represents the gamma function and df represents

the degree of freedom. For this model, the bivariate exponent function is defined as

$$V_{\mathbf{S}}(x_1, x_2) = \frac{1}{x_1} T_{df+1}\left[\frac{(x_2/x_1)^{1/df} - \rho(h)}{(df+1)^{-1/2}\left[1-\rho^2(h)\right]^{1/2}}\right] + \frac{1}{x_2} T_{df+1}\left[\frac{(x_1/x_2)^{1/df} - \rho(h)}{(df+1)^{-1/2}\left[1-\rho^2(h)\right]^{1/2}}\right], \tag{A8}$$

where $T_{df}(.)$ represents the Student t distribution function with degree of freedom df. So, the extremal coefficient for the Extremal-t model can be written as

$$\theta(h) = 2\, T_{df+1}\left[\frac{1-\rho(h)}{(df+1)^{-1/2}\left[1-\rho^2(h)\right]^{1/2}}\right] \tag{A9}$$

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

Castro-Camilo D, Huser R (2020) Local likelihood estimation of complex tail dependence structures, applied to us precipitation extremes. Journal of the American Statistical Association 115(531):1037–1054. https://doi.org/10.1080/01621459.2019.1647842

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. In: Dependence in Probability and Statistics. Springer, New York, NY, p 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

Dhanachandra N, Manglem K, Chanu YJ (2015) Image segmentation using k-means clustering algorithm and subtractive clustering algorithm. Procedia Computer Science 54:764–771

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

Gaetan C, Girardi P, Musau VM (2024) Spatial quantile clustering of climate data. Advances in Data Analysis and Classification pp 1–29

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

Hector EC, Reich BJ (2024) Distributed inference for spatial extremes modeling in high dimensions. Journal of the American Statistical Association 119(546):1297–1308

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

Kabluchko Z, Schlather M, De Haan L (2009) Stationary max-stable fields associated to negative definite functions. The Annals of Probability

37(5):2042–2065

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

Majumder R, Reich BJ (2023) A deep learning synthetic likelihood approximation of a non-stationary spatial model for extreme streamflow forecasting. Spatial Statistics 55:100,755

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. In: Nonlinear and Stochastic Climate Dynamics. Cambridge University Press, Cambridge, p 369–391, https://doi.org/10.1017/9781316339251.014

Ribatet M, Dombry C, Oesting M (2016) Spatial extremes and max-stable processes. In: Extreme value modeling and risk analysis: methods and applications. Chapman and Hall/CRC, New York, p 179–194, https://doi.org/10.1201/b19721

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

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

Schütze H, Manning CD, Raghavan P (2008) Introduction to information retrieval, vol 39. Cambridge University Press Cambridge

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

Talebi H, Peeters L, Mueller U, et al (2020) Towards geostatistical learning for the geosciences: A case study in improving the spatial awareness of spectral clustering. Mathematical Geosciences 52(8):1035–1048

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. Springer, Cham, Switzerland, https://doi.org/10.1007/978-3-319-69308-8

Wright E (2024) Accurately clustering biological sequences in linear time by relatedness sorting. Nature Communications 15(1):3047