



**HAL**  
open science

## Identifying regions of concomitant compound precipitation and wind speed extremes over Europe

Alexis Boulin, Elena Di Bernardino, Thomas Laloë, Gwladys Toulemonde

► **To cite this version:**

Alexis Boulin, Elena Di Bernardino, Thomas Laloë, Gwladys Toulemonde. Identifying regions of concomitant compound precipitation and wind speed extremes over Europe. 2023. hal-04293894

**HAL Id: hal-04293894**

**<https://hal.science/hal-04293894v1>**

Preprint submitted on 19 Nov 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Identifying regions of concomitant compound precipitation and wind speed extremes over Europe

Alexis Boulin<sup>\*,1,3</sup>, Elena Di Bernardino<sup>1</sup>, Thomas Laloë<sup>1</sup> and Gwladys Toulemonde<sup>2,3</sup>

<sup>1</sup>Université Côte d'Azur, CNRS, LJAD, France

<sup>2</sup>Univ Montpellier, CNRS, Montpellier, France

<sup>3</sup>Inria, Lemon

\*Corresponding author. Email: [aboulin@unice.fr](mailto:aboulin@unice.fr)

## Abstract.

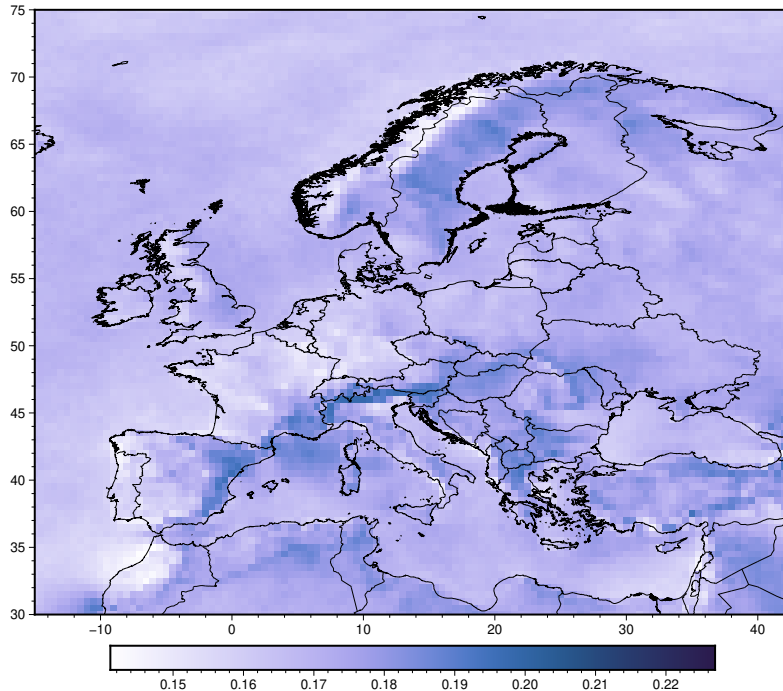
The task of simplifying the complex spatio-temporal variables associated with climate modeling is of utmost importance and comes with significant challenges. In this research, our primary objective is to tailor clustering techniques to handle compound extreme events within gridded climate data across Europe. Specifically, we intend to identify subregions that display asymptotic independence concerning compound precipitation and wind speed extremes. To achieve this, we utilise daily precipitation sums and daily maximum wind speed data derived from the ERA5 reanalysis dataset spanning from 1979 to 2022. Our approach hinges on a tuning parameter and the application of a divergence measure to spotlight disparities in extremal dependence structures without relying on specific parametric assumptions. We propose a data-driven approach to determine the tuning parameter. This enables us to generate clusters that are spatially concentrated, which can provide more insightful information about the regional distribution of compound precipitation and wind speed extremes. In the process, we aim to elucidate the respective roles of extreme precipitation and wind speed in the resulting clusters. The proposed method is able to extract valuable information about extreme compound events while also significantly reducing the size of the dataset within reasonable computational timeframes.

**Keywords:** Compound events, Extremes, High dimensional models, Spatial clustering, Variable clustering.

## Introduction

The occurrence of extreme weather events is often exacerbated by the convergence of distinct geographic factors and concurrent weather patterns, resulting in profound disruptions and extensive damage to society. Catastrophic climate phenomena such as floods, wildfires, and heatwaves frequently manifest due to the simultaneous intensification of multiple interacting processes. When these various processes coalesce to yield a substantial impact, it is referred to as a compound event. Among the primary manifestations of extreme weather, heavy precipitation and robust surface winds hold central positions, exerting adverse effects on both the natural world and human society. Extratropical cyclones, along with their associated wind patterns and storm surges, contribute significantly to economic and insured losses resulting from natural calamities in Europe. Furthermore, they disrupt transportation, trade, and energy supply systems, often leading to human casualties (refer to, for instance, Schwierz et al. 2010; Pinto et al. 2012). To mitigate these impacts, it is important to better understand the dependence structure of extreme weather events. However, modeling such complex scenarios, where multiple rare events occur simultaneously, can be incredibly challenging, especially with high-dimensional climate datasets that exhibit heavy tails.

**Considered ERA5 dataset** Large ensemble simulations present a unique opportunity for gaining deeper insights into the spatial regionalisation of compound precipitation and wind extremes. This is primarily because these simulations provide a more accurate representation of local-scale processes compared to global ensembles and do so without being hindered by data limitations. However, such simulations need to be interpreted with care as it is often largely unknown how well the employed



**Figure 1.** Proportion of the total precipitation or wind speed in the ERA5 dataset that exceed their respective 0.9th quantiles.

models represent observed compound events (Zscheischler et al. 2021), and differences might be large between models. In our endeavor to regionalize compound precipitation and wind speed, we turn to the ERA5 dataset (H. Hersbach et al. 2018). This dataset allows us to investigate the correlation between daily cumulative precipitation and daily peak wind speeds throughout the extended winter between season, spanning from November to March, across Europe, for the period between 1979 and 2022. ERA5 offers a comprehensive record of atmospheric conditions, land surface characteristics, and ocean wave patterns, spanning from 1950 to the present day. It is worth noting that ERA5 supersedes the previous ERA-interim reanalysis, which began in 1979 and was initiated in 2006.

ERA5 has benefited from significant advancements in model physics, core dynamics, and data assimilation techniques developed over the past decade. Produced using 4D-Var data assimilation technology within model cycle 41r2 (Cy41r2), ERA5 incorporates improved parameterization schemes (Hersbach et al. 2020). The dataset is available at a spatial resolution of  $0.25^\circ$  (approximately 27-28 km) on a regular grid. Our specific focus lies within the region defined by  $[-15^\circ E, 42.5^\circ E] \times [30^\circ N, 75^\circ N]$ , which encompasses Europe. We remap the original hourly data to a regularly spaced grid with a  $0.5^\circ$  spatial resolution, allowing us to compute daily precipitation totals and daily maximum wind speeds. We selected the  $0.5^\circ$  spatial resolution due to its ability to facilitate calculations within a reasonable timeframe while maintaining manageable storage requirements. The need for remapping can be circumvented with more extensive computing resources. The resulting dataset comprises 6655 daily precipitation totals and maximum wind speed measurements, covering  $91 \times 116$  grid cells with the chosen spatial resolution, totaling 10556 grid cells for clustering. To illustrate, Fig. 1 provides a visualization of the proportion of grid cells where either wind speed or precipitation exceed a significant threshold. As observed in both panels, there is a noticeable spatial variation in these proportions. In the following, we introduce some notations to describe the spatio-temporal process under consideration.

Consider a spatio-temporal random field denoted as  $(\mathbf{Z}_n^{(s)}, s \in D \subset \mathbb{R}^2, n \in \mathbb{N})$ . Here,  $\mathbf{Z}_n^{(s)} =$

$(Z_n^{(s,1)}, Z_n^{(s,2)})$  represents the vector of daily total precipitation and wind speed maxima at location  $s$  on day  $n$ . We assume that  $\mathbf{Z}_n^{(s)}$  is identically distributed over  $n$  for each location  $s$  in the domain  $D$ . Now, let's suppose that we have observations available at  $d$  spatial locations for each time  $n$ . We can represent these observations as  $\mathbf{Z}_n = (\mathbf{Z}_n^{(1)}, \dots, \mathbf{Z}_n^{(d)})$ , where  $\mathbf{Z}_n$  is random vector with stationary law  $\mathbf{Z} = (\mathbf{Z}^{(1)}, \dots, \mathbf{Z}^{(d)})$ . Each  $\mathbf{Z}^{(j)}$  is a random vector with two dimensions, corresponding to precipitation and wind speed. In the given dataset, each variable represents a location on a grid pixel scale and is characterised by multiple features that can exhibit extreme behavior when considered together. This implies that clustering methods designed to analyze the dependence structure should consider spatial extremal dependence across different locations. In the following discussion, we will review some studies that aim to understand the dependence structure of potentially high-dimensional random vectors.

**Related literature** In the field of high-dimensional extremes, researchers have made significant contributions to address the challenges associated with vectors that have univariate random margins, with a focus on unsupervised techniques such as support identification (see, e.g., Goix, Sabourin, and Cl  men  on 2017; Meyer and Wintenberger 2021), Principal Component Analysis of the angular measure of extremes (Cooley and Thibaud 2019; Drees and Sabourin 2021), graphical models for extremes (Hitz and Evans 2016; Engelke and Hitz 2020; Asenova, Mazo, and Segers 2021) and clustering methods (Jan  en and Wan 2020; Boulin et al. 2023). These methods can help identify hidden spatial patterns and sub-regions where variables are dependent on their extremes, which is crucial for regionalisation tasks.

Over the years, a number of clustering approaches have been suggested, with a focus on extremes, based on a comparison between univariate distributions. For instance, Bernard et al. 2013 analysed weekly maxima of precipitation in France and developed a clustering algorithm on a proper distance, the madogram, justified by Extreme Value Theory (EVT). The same approach was used in Bador et al. 2015 to evaluate the bias of climate model simulations of temperature maxima over Europe. In Durante, Pappad  , and Torelli 2015, a four-step clustering procedure was presented that considered a pairwise conditional Spearman's correlation coefficient, extracted from daily-log-returns of the adjusted stock price, as a measure of tail dependence. Pappad   et al. 2018 investigated spatial sub-regions (clusters) of flood risk behavior in the Po river basin in Italy using a copula-based Agglomerative Hierarchical Clustering. In the paper by Maume-Deschamps, Ribereau, and Zeidan 2023, they introduce a modified spectral clustering algorithm designed for analyzing spatial extreme events. This algorithm combines spectral clustering with the concept of extremal concurrence probability, as proposed by Dombry, Ribatet, and Stoev 2018. The goal of this approach is to determine whether a max-stable process exhibits a stationary dependence structure or not. Recently, Boulin et al. 2023 proposed a class of models, the Asymptotic Independent block (AI block) models, for variable clustering, which defines population-level clusters based on the independence of extremes between clusters. They exhibited an algorithm that compares the extremal dependence of univariate distributions at different locations and showed that it recovers the thinnest partition such that extremes between groups of random variables are mutually independent with high probability.

While regional analysis of univariate climate extreme events is a well-studied area of research, multivariate compound extreme events at larger scales have received less attention. Although the widely known Kullback-Leibler divergence has been adapted for use in the context of compound extreme events, it has been primarily employed to cluster data based on their bivariate extreme behavior (as demonstrated in Vignotto, Engelke, and Zscheischler 2021) and to analyze compound weather and climate events (as discussed in Zscheischler et al. 2021). However, this metric primarily summarises the differences in distribution between two sets of random variables when at least one

of their components is extreme, and it does not quantify deviations from asymptotic independence. recognising sub-regions characterised by concurrent extreme precipitation and wind speed events is essential for improving extreme event modeling. This is particularly evident in works like Chatelain, Fougères, and Nešlehová 2020 and Engelke and Hitz 2020, which rely on the assumption of asymptotic dependence in the data. Such insights are crucial for the development of strategies to mitigate the impacts of these extreme events.

**Proposed methodology.** In this paper, our objective is to expand upon the AI block model as introduced in Boulin et al. 2023 to tackle the challenges posed by this environmental dataset. We depart from the assumption that clusters of pixels are mutually independent univariate time series, with regard to their extremes. Instead, we shift to a framework where a collection of univariate time series is recorded for each pixels, with a particular emphasis on their extreme behavior. To tackle this intrinsic problem, we introduce the concept of *constrained* AI block model, compelling pixels represent a collection of univariate time series. This concept comes into play in our environmental problematic which concerns phenomena like precipitation and wind speed extremes recorded at a specific geographic locations represented as pixels within a large ensemble dataset.

Our objective is the following: cluster a number of  $d = 10556$  pixels across Europe based on their asymptotic independence on compound precipitation and wind speed extremes where data are relatively scarce, i.e., the sample size  $n = 6655$ . To efficiently implement a fast algorithm designed for this model-based approach in such a high-dimensional setting, we employ a divergence measure that highlights the differences in extremal dependence structures for asymptotically dependent and independent random vectors. Noticeably, this divergence measure adheres to several axioms that makes it a valid measure of dependence (De Keyser and Gijbels 2023a, 2023b) and also a coherent measure (Scarsini 1984). Furthermore, this divergence is linked to a well-known quantity in Extreme Value Theory and can be consistently estimated without the need of parametric assumptions. This consistent estimation is possible under the condition of weak mixing conditions to stay within the scope of our application where departures from the independence assumption are strongly suspected.

The algorithm requires the specification of a tuning parameter, and we suggest an approach based on data to determine its value. When applied to our environmental dataset, this clustering procedure is efficient and produces clusters that are spatially concentrated, which is a pattern commonly observed in spatial processes. Furthermore, we leverage the interpretability of classical AI block models to gain insights into the role of precipitation and wind speed extremes in the compound partition. In other words, we use this approach to understand how the clustering is influenced by both wind speed and precipitation. To further analyze the results, we make use of a straightforward modification of our dissimilarity measure. This modification allows us to comment on the different clusterings obtained through various algorithms and provide valuable insights into our proposed methodology.

**Structure of the paper.** The remainder of the paper is structured as follows. Section 1 of the paper provides an introduction to extreme value theory, followed by a detailed explanation of the method used to cluster compound extremes. The proposed clustering process relies an extremal dependence measure for random vectors, recently introduced in Boulin et al. 2023: the Sum of Extremal COefficient (SECO). In Section 2 we present the results obtained from applying the clustering algorithm for considered precipitation and wind speed extremes over Europe. Conclusions of our study and possible improvement of our method are explored in Section 3. In the external supplementary materials, we postponed some material about dependence measures for random vectors (see Section Appendix A). Technical arguments are given in Appendix B, Appendix C and

Appendix D. Finally additional figures are provided in Appendix F.

## 1. A clustering algorithm for compound extreme events

### 1.1 A measure for evaluating dependence between compound extremes

We consider a high-dimensional random vector  $\mathbf{Z} = (\mathbf{Z}^{(1)}, \dots, \mathbf{Z}^{(d)})$  with law  $F$  having  $d$  marginal random vectors  $\mathbf{Z}^{(j)} = (Z^{(j,1)}, \dots, Z^{(j,p_j)})$  for  $j = 1, \dots, d$ . To accommodate vectors of different sizes which will be useful later, we introduce a different notation from the one given in the introduction. Each  $\mathbf{Z}^{(j)}$  contains  $p_j$  marginal univariate random variables  $Z^{(j,\ell)}$  for  $\ell = 1, \dots, p_j$ . In this framework  $\mathbf{Z}$  has  $q = p_1 + \dots + p_d$  marginal univariate random variables.

We call for convenience a function  $u$  on  $\mathbb{R}$  a normalising function if  $u$  is non-decreasing, right continuous, and  $u(x) \rightarrow \pm\infty$  as  $x \rightarrow \pm\infty$ . For a stationary sequence  $(\mathbf{Z}_n, n \in \mathbb{N})$  of  $\mathbf{Z}$ , we say that the distribution  $F$  belongs to the max-domain of attraction of the Extreme Value Distribution (EVD)  $H$  if the following convergence result holds for properly normalised maxima:

$$\mathbb{P} \left\{ \bigvee_{i=1}^n \mathbf{Z}_i \leq \mathbf{u}_n(\mathbf{x}) \right\} \xrightarrow{n \rightarrow \infty} H(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^q, \quad (1)$$

where  $\mathbf{u}_n(\mathbf{x}) = (u_n^{(1,1)}(x^{(1,1)}), \dots, u_n^{(1,p_1)}(x^{(1,1)}), \dots, u_n^{(d,p_d)}(x^{(d,p_d)}))$  is a  $q$ -dimensional vector of normalising functions. The margins  $H^{(1,1)}, \dots, H^{(d,p_d)}$  of  $H$  must be univariate extreme value distributions and the dependence structure of  $H$  is determined by the relation

$$-\ln H(\mathbf{x}) = L \left( -\ln H^{(1,1)}(x^{(1,1)}), \dots, -\ln H^{(d,p_d)}(x^{(d,p_d)}) \right), \quad (2)$$

for all points  $\mathbf{x}$  such that  $H^{(j,\ell)}(x^{(j,\ell)}) > 0$  for all  $j = 1, \dots, d$ ,  $\ell = 1, \dots, p_j$ . The convergence result in (1) with the relation in (2) holds under mild assumptions on the dependence between  $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ , making the study of time series relevant within the framework of EVT. The stable tail dependence function  $L : [0, \infty)^q \rightarrow [0, \infty)$  can be retrieved from the distribution function  $F$  via

$$L(\mathbf{x}) = \lim_{t \rightarrow 0} t^{-1} \mathbb{P} \left\{ F^{(1,1)}(Z^{(1,1)}) > 1 - tx^{(1,1)} \text{ or } \dots \text{ or } F^{(d,p_d)}(Z^{(d,p_d)}) > 1 - tx^{(d,p_d)} \right\}. \quad (3)$$

We assume that the random vector  $\mathbf{Z}$  is in the max-domain of attraction of an EVD, denoted as  $H$ . Moreover, we aim that the extremal dependence can be modelled using an AI block model (see Boulin et al. 2023) where the definition is recalled below.

**Definition 1 (Asymptotic Independent block model).** Let  $(\mathbf{Z}_n, n \in \mathbb{N})$  be a  $q$ -variate stationary random sequence with law  $F$  in the max domain of attraction of  $H$ . The random sequence  $(\mathbf{Z}_n, n \in \mathbb{N})$  is said to follow an AI block model if there exists a partition  $O = \{O_g\}_{g=1}^G$  of  $\{1, \dots, q\}$  with  $|O_g| = d_g$  and marginal extreme value distributions  $H^{(O_g)} : \mathbb{R}^{d_g} \rightarrow [0, 1]$  such that  $H = \prod_{g=1}^G H^{(O_g)}$ .

The constrained AI block model requires improvements to the methods proposed in Boulin et al. 2023, which uses extremal correlation to detect asymptotic independence between random variables, to correctly identify the hidden partition. Asymptotic independence is a concept that describes the relationship between extremes of two random variables, denoted  $Z^{(a)}$  and  $Z^{(b)}$ . Each variable has its own cumulative distribution function, denoted  $F^{(a)}$  and  $F^{(b)}$ , respectively. The extremal correlation, denoted as  $\chi(a, b)$ , between these two random variables is formally stated by

$$\chi(a, b) = \lim_{t \rightarrow 0} \mathbb{P} \left\{ F^{(a)}(Z^{(a)}) > 1 - t | F^{(b)}(Z^{(b)}) > 1 - t \right\}.$$



The extremal correlation represents the probability of one variable being extreme given that the other is also extreme. If the extremal correlation coefficient  $\chi(a, b)$  is in the range of  $(0, 1]$ , then  $Z^{(a)}$  and  $Z^{(b)}$  are said to be asymptotically dependent. Otherwise, if  $\chi(a, b) = 0$ , the variables are asymptotically independent. For instance, the well-known bivariate Gaussian distribution with correlation coefficient  $\rho \in [-1, 1)$  satisfies  $\chi(a, b) = 0$ .

An extension beyond the bivariate case involves examining two groups of random variables. In Boulin et al. 2023, a new metric called Sum of Extremal COefficient (SECO) was introduced. To better understand the definition of the metric, the necessary notations for extremal coefficients of an extremal random vector with possibly different sizes are presented below:

$$\theta(1, \dots, d) = \lim_{t \rightarrow 0} t^{-1} \mathbb{P} \left\{ \max_{j=1, \dots, d} \max_{\ell=1, \dots, p_j} F^{(j, \ell)}(Z^{(j, \ell)}) > 1 - t \right\} \quad (4)$$

$$\theta(j) = \lim_{t \rightarrow 0} t^{-1} \mathbb{P} \left\{ \max_{\ell=1, \dots, p_j} F^{(j, \ell)}(Z^{(j, \ell)}) > 1 - t \right\}, \quad j = 1, \dots, d. \quad (5)$$

Then,  $\theta(j)$  corresponds to the extremal coefficient associated to the  $j$ th marginal random vector  $\mathbf{X}^{(j)}$ . The SECO metric is defined as the difference between the sum of the extremal coefficients of the  $d$  marginal random vectors  $\mathbf{Z}^{(j)}$  and the extremal coefficient of the entire vector, that is, formally stated

$$\text{SECO}(\mathbf{Z}^{(1)}, \dots, \mathbf{Z}^{(d)}) = \sum_{j=1}^d \theta(j) - \theta(1, \dots, d). \quad (6)$$

The SECO metric is always positive and quantifies the deviation from asymptotic independence between the  $d$  groups of variables. Indeed, Boulin et al. 2023, Proposition 8 (iii), Ferreira 2011 showed that the SECO metric is equal to zero if and only if the  $d$  groups of variables are independent extreme value random vectors. Moreover, the bivariate SECO between  $\mathbf{Z}^{(j)}$  and  $\mathbf{Z}^{(k)}$  simplifies to the extremal correlation when these are random variables. Recently, De Keyser and Gijbels 2023b developed an axiomatic framework to quantify dependence between multiple groups of random variables of possibly different sizes. For self-consistency, we recall them in Appendix A and we show that the SECO metric defined in (6) satisfies most of the stated axioms (see Lemma 1 in Appendix A). Also, we are interested in determining whether SECO remains coherent (as defined in Durante and Sempi 2015; Scarsini 1984) when comparing two random vectors with the same dimension, which occurs when  $p_a = p_b$ . In Appendix B of the supplementary materials, we examine the coherence of the SECO in nested extreme value copulae, which were introduced in Hofert, Huser, and Prasad 2018. A further objective is to investigate how SECO behaves in specific nested models.

## 1.2 Clustering for compound extremes

In this section, we introduce a modified version of the ECO algorithm as presented in Boulin et al. 2023, which is capable of clustering compound extremes.

To introduce flexibility in AI block models and bring notations into our application, we consider  $\mathbf{Z}_i^{(j)}$ ,  $i = 1, \dots, n$ ,  $j = 1, \dots, d$ , which are extracted from  $d$  pixels. We assume that each observation is distributed according to  $F$  in the max-domain of attraction an *constrained* AI block model as defined in 1.1. Any deviation from asymptotic independence between two pixels can then be measured by the empirical counterpart version of the SECO in (6). The empirical counterpart of the SECO between two pixels  $a$  and  $b$  is defined as:

$$\widehat{\text{SECO}}(\mathbf{Z}^{(a)}, \mathbf{Z}^{(b)}) = \hat{\theta}(a) + \hat{\theta}(b) - \hat{\theta}(a, b), \quad (7)$$

In this case study, we consider  $d = 10556$  pixels. For each pixel  $j = 1, \dots, d$ , we define a bivariate random vector  $\mathbf{Z}^{(j)} = (Z^{(j,1)}, Z^{(j,2)})$ , where, as a convention in this paper,  $Z^{(j,1)}$  and  $Z^{(j,2)}$  represent the stationary distributions of daily total precipitation and wind speed maxima at location  $j$ , respectively. Hence, to stick in this context,  $\hat{\theta}(a, b)$  is the empirical counterpart of the extremal coefficient for the joint vector  $(\mathbf{Z}^{(a)}, \mathbf{Z}^{(b)})$ , and  $\hat{\theta}(j)$  is the empirical counterpart of the extremal coefficient for the random vector  $\mathbf{Z}^{(j)}$ ,  $j \in \{a, b\}$ , i.e.,

$$\hat{\theta}(a, b) = \frac{1}{k} \sum_{i=1}^n \mathbb{1}_{\{R_i^{(a,1)} > n-k+0.5 \text{ or } R_i^{(a,2)} > n-k+0.5 \text{ or } R_i^{(b,1)} > n-k+0.5 \text{ or } R_i^{(b,2)} > n-k+0.5\}} \quad (8)$$

$$\hat{\theta}(j) = \frac{1}{k} \sum_{i=1}^n \mathbb{1}_{\{R_i^{(j,1)} > n-k+0.5 \text{ or } R_i^{(j,2)} > n-k+0.5\}}, \quad j = a, b, \quad (9)$$

where  $R_i^{(j,\ell)}$  denote the rank of  $Z_i^{(j,\ell)}$  among  $Z_1^{(j,\ell)}, \dots, Z_n^{(j,\ell)}$ ,  $j = a, b$ ,  $\ell = 1, 2$ . The notation described above can be easily extended to pixels of varying sizes. However, for the sake of clarity in notation and to maintain consistency with our application, we focus on the scenario where pixels consist of bivariate time series. The statistic in Equation (7) gauges the strength of dependence between two pixels by counting how often one of the pixels is extreme in terms of either precipitation or wind speed, while adjusted for the number of times a component is extreme in at least one of the pixels.

Classical non-parametric estimators of the extremal coefficient, as demonstrated in equation (8), is only relevant when the number of variables is less than the number of observations, meaning  $d \leq n$ . Indeed, in Appendix C (see Lemma 1 and Lemma 2), we exhibit bounds which show limitations that classical estimators of the extremal dependence structure face, preventing them from encompassing the full range of extremal dependence structures. In high-dimensional scenarios where the number of variables exceeds the number of observations ( $d > n$ ), classical estimators may fail to identify asymptotic independence in extremes. This is particularly relevant for estimators such as Equation (9). This phenomenon can also be observed in other estimators, such as the madogram, and other rank-based estimators of the tail dependence structure may suffer from the same issue. Therefore, caution must be exercised when dealing with high-dimensional data, particularly by taking a lower number of extremes (determined by the parameter  $k$ ), as advised by the upper bound which is equal to  $n/k$ . This approach allows for a wider range of values with the cost of a greater variance.

Nonetheless, while the dimension is arbitrary, the empirical estimator of the extremal coefficient in (9) is consistent and the asymptotic deviation is well-understood (see, for instance, Drees and Huang 1998; Einmahl, Krajina, and Segers 2012) in the standard assumption of independent observations. In Appendix D, Proposition 2, we present arguments regarding the consistence of the proposed estimator of the SECO which goes beyond this classical setup of independent observations.

If  $\mathbf{Z}^{(a)}$  and  $\mathbf{Z}^{(b)}$  are asymptotically independent, there is no guarantee that an extreme event in one vector will be accompanied by an extreme event in the other vector, then the statistic in (7) will converge in probability to zero (see Appendix D, Proposition 2). On the other hand, if  $\mathbf{Z}^{(a)}$  and  $\mathbf{Z}^{(b)}$  are asymptotically comonotone, then the SECO reduces to  $\hat{\theta}(a) = \hat{\theta}(b)$  almost surely, since an extreme event in one vector will always be accompanied by an extreme event in the other vector. Additionally, the lower and upper bounds of  $\widehat{\text{SECO}}(a, b)$  is given by:

$$0 \leq \widehat{\text{SECO}}(a, b) \leq \min\{\hat{\theta}(a), \hat{\theta}(b)\} \quad \text{a.s.}, \quad (10)$$

where the upper one is reached when  $\mathbf{Z}^{(a)}$  and  $\mathbf{Z}^{(b)}$  are asymptotically comonotone. The resulting matrix, denoted by  $\hat{\Theta}$ , is a  $d \times d$  matrix where each entry is given by

$$\hat{\Theta}(a, b) = \widehat{\text{SECO}}(a, b) / \min\{\hat{\theta}(a), \hat{\theta}(b)\}, \quad a, b \in \{1, \dots, d\}, \quad (11)$$



which is the normalised SECO metric.

---

**Algorithm (CAICE)** Clustering procedure for AI block models with compound extreme

---

```

1: procedure CAICE( $S, \tau, \hat{\Theta}$ )
2:   Initialise:  $S = \{1, \dots, d\}$ ,  $\hat{\Theta}(a, b)$  for  $a, b \in \{1, \dots, d\}$  and  $l = 0$ 
3:   while  $S \neq \emptyset$  do
4:      $l = l + 1$ 
5:     if  $|S| = 1$  then
6:        $\hat{O}_l = S$ 
7:     if  $|S| > 1$  then
8:        $(a_l, b_l) = \arg \max_{a, b \in S} \hat{\Theta}(a, b)$ 
9:       if  $\hat{\Theta}(a_l, b_l) \leq \tau$  then
10:         $\hat{O}_l = \{a_l\}$ 
11:       if  $\hat{\Theta}(a_l, b_l) > \tau$  then
12:         $\hat{O}_l = \{s \in S : \hat{\Theta}(a_l, s) \wedge \hat{\Theta}(b_l, s) \geq \tau\}$ 
13:        $S = S \setminus \hat{O}_l$ 
14:   return  $\hat{O} = (\hat{O}_l)_l$ 

```

---

The algorithm that we present in this section takes as input the matrix  $\hat{\Theta}$  in (11). This enables the division of  $d$  objects of interest into the thinnest partition possible such that mutual asymptotic independence holds between clusters. Algorithm (CAICE) summarises the procedure for clustering asymptotically independent compound extreme events. As detailed in Boulin et al. 2023 for the Algorithm ECO which is similar to the above algorithm ((CAICE)), the overall complexity of the estimation procedure is  $O(d^2(d \vee n \ln(n)))$ .

The  $\tau$  threshold is the only hyper-parameter in the (CAICE) Algorithm, and its selection is important in obtaining an effective partitioning. A useful tool for choosing an appropriate threshold is the SECO value for the resulting partition, which has been recommended in Boulin et al. 2023, Section 3.4. This metric measures the divergence between the sum of the extremal coefficients of each cluster and the extremal coefficient of the entire vector (see (6) with group of different sizes). An effective partitioning is achieved when this metric is minimised, ideally for a moderate value of the threshold  $\tau$ . By identifying the threshold that results in the lowest SECO value, one can establish a partition of pixels, ensuring that their clusters are asymptotically independent from each other.

## 2. Detecting concomitant extremes of compound precipitation and wind speed extremes

### 2.1 Non-serially independent

To statistically assert departures from serial independence of multivariate time series, we conducted a randomness test as proposed by Genest and Remillard 2004 and Ghoudi, Kulperger, and Remillard 2001. We use the methodology presented in Kojadinovic and Holmes 2009 and extended in Kojadinovic and Yan 2011 to detect serial dependence in continuous multivariate time series. Briefly, let  $\mathbf{Z}_1, \mathbf{Z}_2, \dots$  be  $d$ -dimensional random vectors. We chose an integer  $k > 1$ , and for  $\mathbf{u} \in [0, 1]^{dk}$ , the vector  $\mathbf{u}_{\{j\}}$  is defined as follows:

$$u_{\{j\}}^{(i)} = (u^{(i)} - 1) \mathbb{1}_{\{i \in \{(j-1)q+1, \dots, jq\}\}} + 1,$$

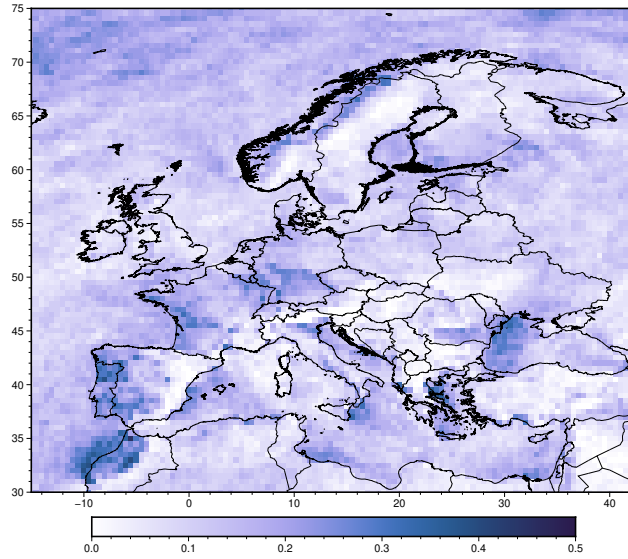
We form the  $dk$ -dimensional random vector  $\mathfrak{Z}_i = (\mathbf{Z}_i, \dots, \mathbf{Z}_{i+k-1})$ , with  $i = 1, \dots, n$ . The serial

independence empirical copula process in the multivariate setting thus write

$$\sqrt{n} \left( C_n^s(\mathbf{u}) - \prod_{j=1}^d C_n^s(\mathbf{u}_{\{j\}}) \right), \quad (12)$$

where  $C_n^s$  is the serial empirical copula process computed with  $\mathfrak{Z}_1, \dots, \mathfrak{Z}_n$ . Under the hypothesis of serial independence, one can establish the asymptotic behavior of (12) to a tight Gaussian process. To obtain potentially powerful test obtained above from the empirical process, Kojadinovic and Yan 2011 derived  $2^{k-1} - 1$  tests statistic based on a Möbius decomposition of the process in (12) in the continuous multivariate time series setting. Those tests are implemented in the copula R package (Yan 2007).

Due to computational limitations, we only considered three  $3 \times 3$  pixels, at different locations, covering the initial 304 days of the study, which represented the two first years of observation. We analysed precipitation and wind separately and the resulting dependogram is depicted in Fig. 11 in Appendix F of the supplementary materials. It is notable that most of the “subsets of lags” exhibited serial dependence. The function `multSerialIndepTest` computed three  $p$ -values, all of which provided robust evidence against serial independence. Furthermore, the decreasing trend of the computed statistic implied that the dependence was weakening for observations further apart. These results align with the mixing conditions stated in Appendix D.



**Figure 2.** Estimator of the extremal correlation within pixels, focusing on the daily sum of precipitation and wind speed maxima,  $\hat{\chi}$  in Equation (13), maps for the 100 largest values ( $k = 100$ ).

## 2.2 Exploratory analysis

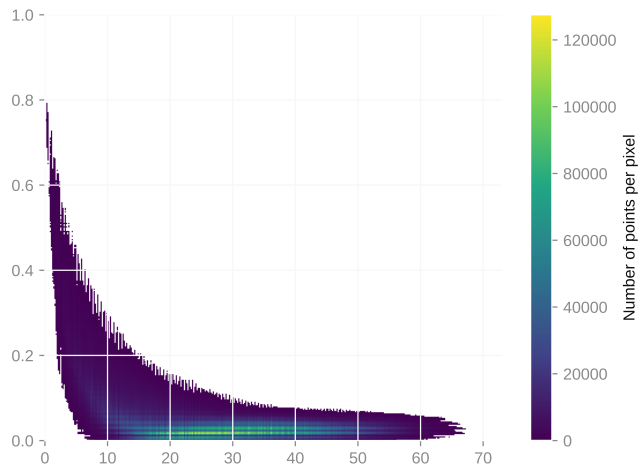
Extreme values of daily precipitation and wind speed maxima may occur together, and we aim to understand the spatial variability of this relationship. To identify regions for which extrema are non-concomitant between them, we consider the peak over threshold approach, which are values that lie above a certain value (see, for instance, Beirlant et al. 2006; De Haan and Ferreira 2006; Resnick 2007 for an overall introduction to classical statistical methods in EVT). We conduct an exploratory analysis of the extremal dependence structure between the two variables within pixels. The estimated  $\chi$  coefficient is defined by

$$\hat{\chi}(a) = \frac{1}{k} \sum_{i=1}^n \mathbb{1}_{\{R_i^{(a,1)} > n-k+0.5, R_i^{(a,2)} > n-k+0.5\}}, \quad (13)$$

where  $R_i^{(a,\ell)}$  denotes the rank of  $Z_i^{(a,\ell)}$  among  $Z_1^{(a,\ell)}, \dots, Z_n^{(a,\ell)}$ ,  $\ell = 1, 2$ . The estimated extremal coefficient between the variables reveals the highest co-occurrences in regions along the western coasts of Portugal, Spain, France, the UK, and Norway, as well as the northeastern coast of the Mediterranean (Fig. 2). Conversely, the smallest co-occurrences are observed on the eastern coasts of the UK, Sweden, and Spain, over the northwestern coast of the Mediterranean, and around the Carpathian and southeastern Norwegian mountain ranges. These results are consistent with prior research (see, e.g., Martius, Pfahl, and Chevalier 2016).

The low co-occurrence over eastern Norway and Spain may be attributed to the orographic enhancement of rain on the windward side of a mountain and the drying of the air as it reaches the lee (Martius, Pfahl, and Chevalier 2016). This could also explain the high co-occurrences over the eastern coasts of the Mediterranean, where cyclones from the Mediterranean storm track may arrive perpendicularly to the mountains on the western coast of Italy and the eastern coasts of the Adriatic Sea (Owen et al. 2021). Additionally, the Cierzo winds may be responsible for the low co-occurrence to the south of the Pyrenees (Martius, Pfahl, and Chevalier 2016).

Nevertheless, the aforementioned analysis only considers extreme behavior within individual pixels. To dig deeper, we take a further step by using the empirical SECO outlined in Equation (7) to explore interactions between pixels. Noteworthy, the empirical SECO does not inherently include spatial distances between pixels. As shown in Fig. 3, we observe strong or moderate dependence between extremes for locations that are close to each other, while locations that are far apart have a normalised SECO near zero, indicating that extremes are weakly dependent or independent. The number of points per pixel highlights that most pixels are widely separated, with Euclidean distances between 20 to 40, and have small SECO values, indicating weak dependence or independence between compound extremes of two pixels. Therefore, even though it is not explicitly designed for this purpose, in the case of this specific dataset, the empirical SECO captures significant spatial information. This information could be highly valuable for comprehending the spatial patterns of precipitation and wind speed in Europe.



**Figure 3.** Pairwise SECO as a function of distance between sites.

In this context, our aim is to gain a deeper understanding of the spatial patterns of extreme total precipitation and wind speed maxima across Europe. To broaden our analysis of extremal pattern within pixels, we propose employing the clustering algorithm described in (CAICE) to group pixels

while considering the extremal dependence of precipitation and wind speed between pixels.

### 2.3 Clustering with constrained AI block models

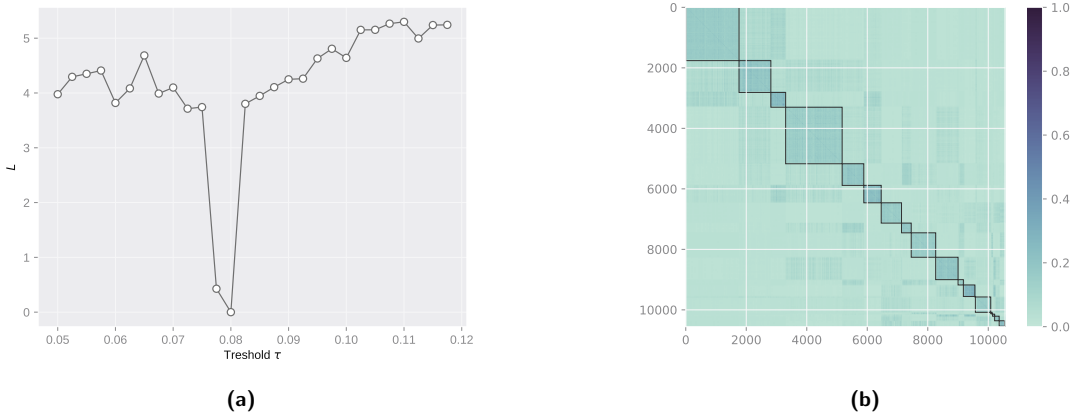
To delineate Europe into distinct regions that are mutually independent in compound weather extremes, we employ the clustering algorithm (CAICE). Additionally, we employ a data-driven approach for selecting a suitable threshold, as elaborated below. Our initial step involves working with the matrix  $\hat{\Theta}$  derived from Equation (11). The resultant clustering outcome is contingent on the value of  $\tau$ . Once Algorithm (CAICE) has been applied, we will denote the resulting partition as  $\hat{O}(\tau) = \{\hat{O}_g\}_{g=1}^G$  of the set  $\{1, \dots, d\}$ . Each cluster is characterised by a respective cardinality of  $d_g$ . As depicted in Figure 3, even a small  $\tau$  value within Algorithm (CAICE) can effectively partition the sub-regions of Europe. Nevertheless, a more rigorous approach to selecting the threshold value  $\tau$  need to be discussed.

Let us introduce the random vector  $\mathbf{Y}_\tau^{(g)}$ , composed of the variable index within  $\hat{O}_g$ . This means that  $\mathbf{Y}_\tau^{(g)}$  is a  $(2d_g)$ -dimensional random vector, considering two variables for each pixel in our case study. Moving forward, we can define the empirical SECO for groups of random vectors, which might have varying sizes within the given partition, using the following equation:

$$\widehat{\text{SECO}}(\mathbf{Y}_\tau^{(1)}, \dots, \mathbf{Y}_\tau^{(G)}) = \sum_{g=1}^G \hat{\theta}(g) - \hat{\theta}(1, \dots, d). \quad (14)$$

The estimator mentioned above varies with  $\tau$  due to its reliance on the partition  $\hat{O}(\tau)$ . As outlined in Section 3.4 of Boulin et al. 2023, the values of  $\tau$  that minimise SECO also ensure consistent recovery of our groups. For more details, we refer to Proposition 3 in Boulin et al. 2023. To address this, we construct a loss function, denoted as  $L$ , over a grid of  $\tau$  values denoted as  $\Delta$ , as follows:

$$L(\tau) = \ln \left( 1 + \left( \widehat{\text{SECO}}(\mathbf{Y}_\tau^{(1)}, \dots, \mathbf{Y}_\tau^{(G)}) - \min_{\tau \in \Delta} \widehat{\text{SECO}}(\mathbf{Y}_\tau^{(1)}, \dots, \mathbf{Y}_\tau^{(G)}) \right) \right), \quad \tau \in \Delta. \quad (15)$$



**Figure 4.** Value of the function  $L(\tau)$  in (15) for different values of  $\tau \in \Delta = \{0.05, 0.0525, \dots, 0.12\}$  in Panel a. Partitions of the SECO similarity matrix with threshold  $\tau = 0.08$  in Panel b. Squares represent the clusters of variables.

In our current high-dimensional context, where we have  $n = 6655$  daily observations and  $d = 10556$  pixels ( $n < d$ ), estimating the extremal coefficient for the entire vector  $\theta(1, \dots, d)$  encounters an upper bound that prevents it from reaching its theoretical maximum of  $d$ , as discussed

in Section 1.2. However, the loss function  $L$  in equation (15) remains unaffected by this bias since it is mitigated by the subtraction operation. Nonetheless, it is essential to be mindful of this bias when computing  $L$  for partitions with larger clusters, which typically occurs for smaller values of  $\tau$ . Therefore, we recommend reducing the number of extremes to  $k = 30$  when evaluating  $L$ . This adjustment widens the range of values used for estimating the extremal coefficient from higher values, but it comes at the expense of increased variance in the estimation process.

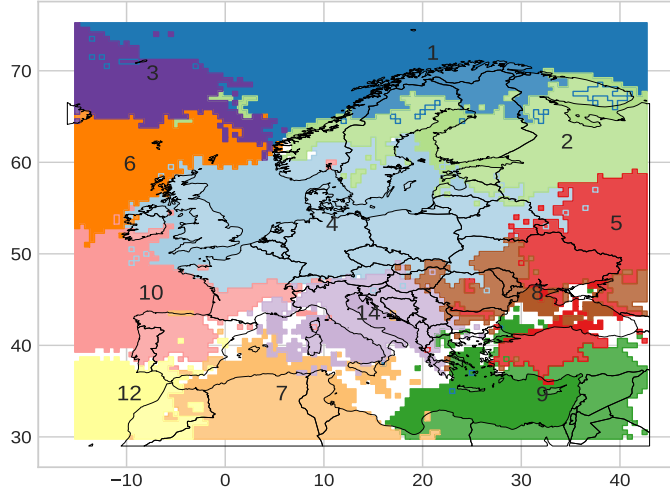
The value of  $L$  with different values of  $\tau$  with  $\Delta = \{0.05, 0.0525, \dots, 0.12\}$  suggests that the best partitioning is found for  $\tau = 0.08$  (Fig. 4, Panel a). For this threshold, we obtain 22 clusters with 3 of them having less than 10 entities. We report the clustered SECO matrix defined in (11) in Fig. 4, Panel b.

## 2.4 Results

In Fig. 5, we present the twelve largest clusters obtained with the partition setting  $\tau = 0.08$ . Our algorithm effectively identifies sub-regions with strong dependence within clusters, as well as *near-independence* or independence among compound extremes of daily precipitation and wind speed maxima in different clusters. The most prominent cluster, the fourth one, encompasses the North Sea and the Baltic Sea, which are connected basins. The North Sea is a marginal basin of the North Atlantic, and a shallow connection to the Baltic Sea exists through the Skagerrak and Kattegat regions. This area is particularly important for climate sciences and hydrology, and has inspired several works (see, for example, Andrée et al. 2022; Gröger et al. 2019; Wang et al. 2015). Another interesting area is the ninth cluster, consisting of the southwestern Black Sea, the Levantine Basin, and the southeast of the Anatolian peninsula. The Levantine Basin is known to be one of the windiest areas of the Mediterranean Sea, and the spatial distribution of the Levantine Basin in Figure 2 of Soukissian et al. 2018 outlines the geometry of this cluster. Additionally, extreme storm situations typically occur in December-January in the southwestern part of the Black Sea (Divinsky et al. 2020). To the best of our knowledge, there are no studies on the tail dependence structure of climate variables in the southwestern Black Sea and the Levantine Basin. Furthermore, our algorithm sometimes distinguishes between the extremal behavior of land and sea, as illustrated by the first and sixth clusters for the Norwegian Sea and the Atlantic Ocean, respectively.

Identifying regions with simultaneous extreme events can be valuable for statistical modeling, and the method proposed here can address this question effectively. However, it is important to understand whether wind or precipitation is more important for the resulting clustering. To investigate this, a similar clustering process was applied separately to daily total precipitation and wind speed maxima using the extremal correlation as a dissimilarity. The resulting partitions are denoted by  $\hat{O}^P$ , and  $\hat{O}^W$  with an adapted threshold  $\tau$ . We will denote by  $\hat{O}^{PW}$  the resulting partition for compound daily total precipitation and wind speed maxima. The figures depicting the results for the individual variables can be found in Fig. 12 in Appendix F.

The adapted threshold for the partition  $\hat{O}^P$  is  $\tau = 0.09$ , resulting in 70 medium-sized clusters. For daily wind speed maxima, the optimal clustering  $\hat{O}^W$  is obtained by setting  $\tau = 0.07$ , resulting in 24 clusters. Upon visual inspection, our algorithmic partitioning reveals mosaic block patterns along the diagonal, while no clear patterns could be discerned from the off-diagonal. Additionally, the off-diagonal showcases moderate asymptotic dependence between groups or asymptotic independence, indicating that the resulting clustering aligns with the purpose of AI block models. The clusters for daily wind speed maxima are larger than those for daily total precipitation, which supports previous studies indicating that heavy gusts have a larger spatial impact than precipitation events (see, for example, Pfahl and Wernli 2012; Raveh-Rubin and Wernli 2015). With regard to spatial precipitation, several studies have shown that dependence tends to weaken for the largest observations



**Figure 5.** Representation of the 12 largest clusters of the partition of the SECO matrix between pixels with Algorithm (CAICE) and threshold  $\tau = 0.08$ . Number of each cluster is depicted inside the cluster.

(Lalancette, Engelke, and Volgushev 2021; Le et al. 2018). This knowledge could explain why there are a large number of clusters with only a few entities for the clustering of daily total precipitation.

To compare different clustering methods, we use the Adjusted Rand Index (ARI), a popular measure used in clustering analysis (see, for instance, Hubert and Arabie 1985; Rand 1971). To summarise, the ARI gives a concordance score between two different partitions. It takes values between 0 and 1 and the closer to 1, the more similar the partitions. For more details about its computation, we refer the reader to Appendix E of the supplementary materials. Computing the ARI between  $\hat{O}^{PW}$  and  $\hat{O}^W$  (resp,  $\hat{O}^{PW}$  and  $\hat{O}^P$ ), we obtain

$$\text{ARI}(\hat{O}^{PW}, \hat{O}^W) = 0.5, \quad \text{ARI}(\hat{O}^{PW}, \hat{O}^P) = 0.3.$$

An ARI value of 0 indicates that the two sets are completely random and have nothing in common, while a value of 1 indicates a perfect match between the two partitions. In this case, the ARI value of 0.5 between the clustering of compound daily total precipitation and wind speed maxima and the clustering of the sole wind speed suggests that there is moderate similarity between the two sets, implying that there are fewer matching data points or clusters between them. In light of these results, we can conclude that the clustering of compound extreme is induced by both variables with a little more emphasis driven by wind speed maxima.

### 2.5 Alternative clustering method using SECO

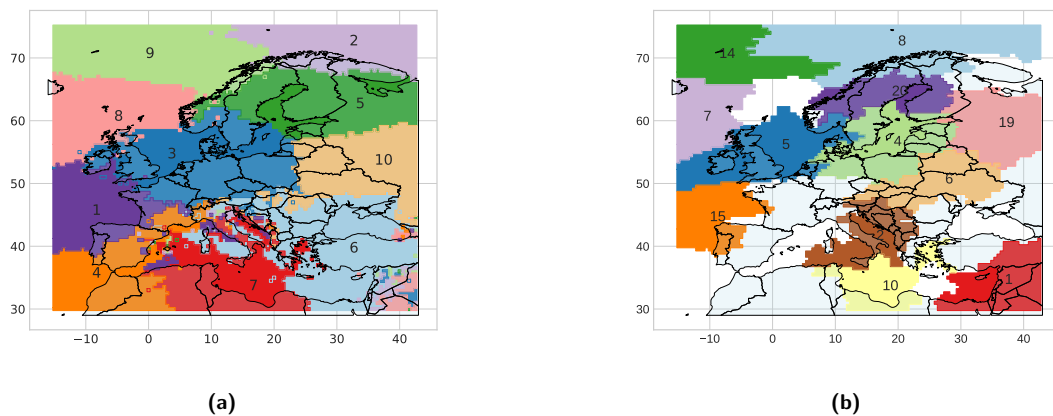
In equation (11),  $\hat{\Theta}$  can be envisioned as a similarity matrix, and  $1 - \hat{\Theta}$  as a dissimilarity matrix. This dissimilarity metric is smaller (or larger) for compound extremes that are dependent (or independent) between two pixels. The upper bound of the dissimilarity metric is 1, which is reached when the compound extremes are independent. Thus, it is possible to perform classical method of clustering using this dissimilarity matrix. In particular, we have chosen to explore two different methods: a quantization-based approach and a Hierarchical clustering. These two approach requiring a specification of the number of clusters (unknown in practice), we use the ‘‘silhouette coefficient’’ developed by, Rousseeuw 1987 and which compares the tightness of clusters with their dissociation.



In practice, the number of clusters, denoted as  $K$ , is determined by selecting the maximum average silhouette coefficient.

Quantization, also known as lossy data compression in information theory, involves the task of substituting data with an efficient and compact representation that allows for the reconstruction of the original observations with a certain degree of accuracy. A clustering problem can be viewed as an optimal quantization process aimed at minimising a specific loss function (for example, see Banerjee et al. 2005 or Linder 2002, p. 15). To generate clusters through the optimal quantization process, we utilised the algorithm described in references Laloë 2010 and Laloë 2021, setting arbitrary the number of clusters to  $K = 10$  as the silhouette coefficient did not provide a suitable number of clusters.

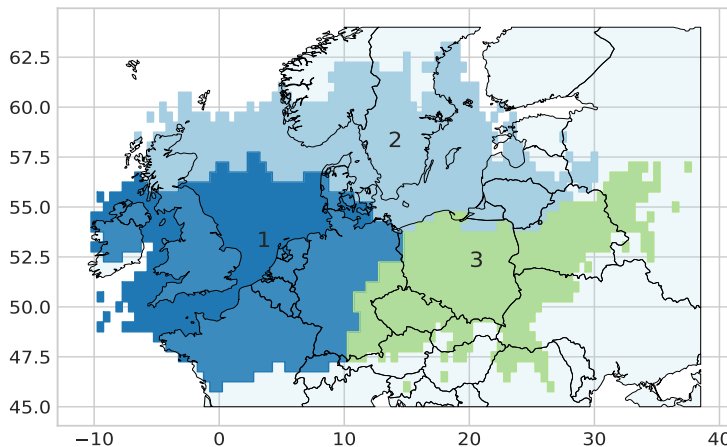
The process of hierarchical clustering begins with a basic partition, initially comprising of  $d$  individual data points where each pixel stands alone as its own cluster. Subsequently, the data is grouped together incrementally, with clusters gradually merging until a single cluster encompassing all variables is achieved. At each step, the algorithm combines the two closest cluster centers according to a specific definition (for more details, refer to Giraud 2021, Chapter 12), while keeping other clusters unchanged. For this approach, the silhouette coefficient leads us to set  $K = 25$ .



**Figure 6.** In Panel **a** is depicted the representation of the partition obtained through quantization-based approach with  $K = 10$ . In Panel **b** a similar representation for the 12 largest clusters is proposed through hierarchical clustering approach with  $K = 25$ . These clusters were generated based on data related to daily total precipitation and wind speed maxima extremes from 10556 pixels. Number of each cluster is depicted inside the cluster.

The resulting partitions for this two methods are depicted in Figure 6, and the clustered matrices can be found in Fig. 13 and Fig. 14 for further visualizations, in Appendix F. Notably, the clusters obtained through the quantization-based approach in the northern regions bear a resemblance to those obtained through Algorithm (CAICE). For instance, clusters 3, 5, and 8 in Figure 6, Panel a, are closely similar to clusters 2, 4, and 6 in Figure 5 as identified through the (CAICE) Algorithm. Furthermore, as observed in the (CAICE) Algorithm, the quantization-based approach effectively distinguishes extreme behaviors between land and sea. This distinction is apparent in clusters 2 and 8 in Figure 6, Panel a. These findings suggest the uniqueness of the SECO similarity measure in extracting valuable spatial information. One notable difference between the hierarchical clustering and Algorithm (CAICE) is that the clusters obtained through hierarchical clustering are smaller and depict stronger cross-dependencies. This phenomenon can be explained by the fact that some clusters which are separate in the hierarchical clustering partition are combined in the output of Algorithm (CAICE). For instance, in Fig. 6, Panel b, most of pixels of clusters 4 and 5 in the

hierarchical clustering are combined into a single cluster in Algorithm (CAICE), the fourth one in Fig. 5. This suggests that compound extremes of those clusters are dependent, but smaller groups in the northern and Baltic Sea areas observe more concomitant extremes. Below, we investigate a hierarchical clustering of the fourth cluster given by Algorithm (CAICE) to inquiry whether or not we obtain a similar partition given by cluster 4 and 5 obtained in the hierarchical clustering.



**Figure 7.** Representation of the 3 clusters of the partition of the 1868 pixels of the fourth cluster of the partition given by Algorithm (CAICE) using extremes of daily total precipitation and wind speed maxima. Number of each cluster is depicted inside the cluster.

Fig. 7 displays the results of hierarchical clustering analysis on the most prominent cluster, i.e., cluster 4 from the output of Algorithm (CAICE). The analysis is performed on a reduced dataset, and the number of clusters is calibrated using the silhouette coefficient Rousseeuw 1987, with the maximum value obtained for  $K = 3$  (see Fig. 15). The resulting matrix,  $\hat{\Theta}$ , reveals a strong dependence among compound extremes, indicating the presence of a whole asymptotic dependent vector with three distinct blocks with more concomitant compound extremes.

These results reveal intriguing patterns of co-occurring wind and precipitation extremes. The third cluster in Fig. 7, which represents central-eastern Europe, is spatially coherent and having no access to the sea. The first cluster is connected to the North Sea and consists of western European countries (except Scotland). The second cluster comprises Scandinavian countries and the Baltic Sea, with a shallow connection to the North Sea. Notably, the North Sea’s section belonging to this area corresponds to the Fladen Ground, the deepest part of the North Sea in the Scottish sector.

We also observe a similar partition given by the hierarchical clustering among all locations, where the North Sea is separated from the Baltic Sea. However, there is a difference in the Fladen Ground’s classification, which is in different clusters for the hierarchical clustering among all pixels belonging to the North Sea cluster, see cluster 5 in Fig. 6b, and to the Baltic Sea cluster for the hierarchical clustering applied to the reduced dataset, see cluster 2 in Fig. 7.

### 3. Conclusion and perspectives

Clustering spatial pixels is highly significant as it allows for a better comprehension of the inherent spatial pattern of a relevant physical phenomenon, enhances the accuracy of statistical procedures in situations where data is limited, and helps to identify regions where joint preventive measures can be taken to mitigate the impact of weather-related risks.

Traditionally, climate science research has focused on analysing single drivers or univariate dangers, which simplifies the complex dynamics of climate and its consequences. However, in reality, climate hazards often interact, leading to compound extremes. Therefore, statistical analyses that consider multiple hazards simultaneously are needed. In this paper, our goal is to identify subregions in Europe that demonstrate asymptotic independence concerning compound precipitation and wind speed extremes. In simpler terms, we want to find areas where two distinct subregions cannot experience concurrent compound extremes. To achieve this, we introduce a multivariate extreme value measure known as the SECO metric. This metric helps us quantify the extent to which random vectors of varying sizes deviate from asymptotic independence. We not only introduce this metric but also demonstrate the reliability of a non-parametric estimator for it, even in scenarios that go beyond the typical setup of independent observations. Building on this, we propose an algorithm specifically tailored for *constrained* AI block model. This model ensures that pixels represent collections of univariate time series. Our algorithm allows us to pinpoint the largest partition where compound extremes exhibit independence over Europe. Interestingly, we uncover specific geographical patterns without relying on positional information, and it is worth noting that we do not need to pre-determine the number of clusters.

The proposed methodology can be extended to the case where the dataset has more than two variables, denoted as  $p > 2$ . Furthermore, the methodology can also be extended to pixels with varying lengths of recorded time series, all the while preserving the concept of asymptotic independence. However, caution should be taken when interpreting the results in terms of coherence, or it may not be appropriate at all.

A major issue highlighted in this paper is how to estimate the dependence structure in high-dimensional datasets, where the number of variables is greater than the number of observations. As explained in [Appendix C](#), traditional estimators are not able to accurately recover the dependence structure of an extreme value random vector if its margins are not sufficiently dependent, quantified by the condition  $L(\mathbf{x}) \geq n/k$  for  $\mathbf{x} \in \mathbb{R}^q$  where  $n$  is the number of observations and  $k$  the number of considered extremes. However, extreme value data are often scarce, making the high-dimensional setting common, except in certain cases.

### Acknowledgments

This work has been supported by the project ANR McLaren (ANR-20-CE23-0011). This work has been partially supported by the French government, through the 3IA Côte d’Azur Investments in the Future project managed by the National Research Agency (ANR) with the reference number ANR-19-P3IA-0002. This work was also supported by the project ANR GAMBAS (ANR-18-CE02-0025) and by the french national programme LEFE/INSU.

### Data Availability Statement

Our code, SECO matrix and extremal correlation matrices are available from the following GitHub repository: [https://github.com/Aleboul/compound\\_extreme](https://github.com/Aleboul/compound_extreme).

### References

- Andrée, Elin, Martin Drews, Jian Su, Morten Andreas Dahl Larsen, Nils Drønen, and Kristine Skovgaard Madsen. 2022. Simulating wind-driven extreme sea levels: sensitivity to wind speed and direction. *Weather and Climate Extremes* 36:100422.
- Asenova, Stefka, Gildas Mazo, and Johan Segers. 2021. Inference on extremal dependence in the domain of attraction of a structured Hüsler–Reiss distribution motivated by a Markov tree with latent variables. *Extremes* 24:461–500.

- Bador, Margot, Philippe Naveau, Eric Gilleland, Mercè Castellà, and Tatiana Arivelo. 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.
- Banerjee, Arindam, Srujana Merugu, Inderjit S Dhillon, Joydeep Ghosh, and John Lafferty. 2005. Clustering with Bregman divergences. *Journal of machine learning research* 6 (10).
- Beirlant, Jan, Yuri Goegebeur, Johan Segers, and Jozef L Teugels. 2006. *Statistics of extremes: theory and applications*. John Wiley & Sons.
- Berghaus, Betina, Axel Bücher, and Stanislav Volgushev. 2017. Weak convergence of the empirical copula process with respect to weighted metrics. *Bernoulli* 23 (1): 743–772. <https://doi.org/10.3150/15-BEJ751>. <https://doi.org/10.3150/15-BEJ751>.
- Bernard, Elsa, Philippe Naveau, Mathieu Vrac, and Olivier Mestre. 2013. Clustering of maxima: spatial dependencies among heavy rainfall in France. *Journal of climate* 26 (20): 7929–7937.
- Boulin, Alexis, Elena Di Bernardino, Thomas Laloë, and Gwladys Toulemonde. 2023. High-dimensional variable clustering based on sub-asymptotic maxima of a weakly dependent random process. *arXiv preprint arXiv:2302.00934*.
- Bücher, Axel, and Johan Segers. 2014. Extreme value copula estimation based on block maxima of a multivariate stationary time series. *Extremes* 17:495–528.
- Chatelain, Simon, Anne-Laure Fougères, and Johanna G. Nešlehová. 2020. Inference for Archimax copulas. *The Annals of Statistics* 48 (2): 1025–1051.
- Cooley, Daniel, and Emeric Thibaud. 2019. Decompositions of dependence for high-dimensional extremes. *Biometrika* 106 (3): 587–604.
- Csörgő, Miklós. 1983. *Quantile processes with statistical applications*. SIAM.
- De Haan, Laurens, and Ana Ferreira. 2006. *Extreme value theory: an introduction*. Vol. 3. Springer.
- De Keyser, Steven, and Irène Gijbels. 2023a. Copula-based divergence measures for Dependence Between Random Vectors. In *Building Bridges between Soft and Statistical Methodologies for Data Science*, edited by Luis A. García-Escudero, Alfonso Gordaliza, Agustín Mayo, María Asunción Lubiano Gomez, Maria Angeles Gil, Przemyslaw Grzegorzewski, and Olgierd Hryniewicz, 104–111. Cham: Springer International Publishing. ISBN: 978-3-031-15509-3.
- . 2023b. Parametric dependence between random vectors via copula-based divergence measures. *arXiv preprint arXiv:2302.13611*.
- Divinsky, Boris V, Vladimir V Fomin, Ruben D Kosyan, and Yuri D Ratner. 2020. Extreme wind waves in the Black Sea. *Oceanologia* 62 (1): 23–30.
- Dombry, Clément, Mathieu Ribatet, and Stilian Stoev. 2018. Probabilities of concurrent extremes. *Journal of the American Statistical Association* 113 (524): 1565–1582.
- Doukhan, Paul, Pascal Massart, and E. Rio. 1995. Invariance principles for absolutely regular empirical processes. *Annales de l'I.H.P. Probabilités et statistiques* 31:393–427.
- Drees, Holger, and Xin Huang. 1998. Best attainable rates of convergence for estimators of the stable tail dependence function. *Journal of Multivariate Analysis* 64 (1): 25–46.
- Drees, Holger, and Anne Sabourin. 2021. Principal component analysis for multivariate extremes. *Electronic Journal of Statistics* 15 (1): 908–943.
- Durante, Fabrizio, Roberta Pappadà, and Nicola Torelli. 2015. Clustering of time series via non-parametric tail dependence estimation. *Statistical Papers* 56:701–721.
- Durante, Fabrizio, and Carlo Sempi. 2015. *Principles of copula theory*. CRC press.
- Einmahl, John H. J., Andrea Krajina, and Johan Segers. 2012. An M-estimator for tail dependence in arbitrary dimensions. *The Annals of Statistics* 40 (3): 1764–1793.
- Engelke, Sebastian, and Adrien S Hitz. 2020. Graphical models for extremes. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 82 (4): 871–932.

- Ferreira, H. 2011. Dependence between two multivariate extremes. *Statistics & Probability Letters* 81 (5): 586–591. ISSN: 0167-7152. <https://doi.org/https://doi.org/10.1016/j.spl.2011.01.014>. <https://www.sciencedirect.com/science/article/pii/S0167715211000216>.
- Genest, Christian, and Bruno Remillard. 2004. Test of independence and randomness based on the empirical copula process. *TEST: An Official Journal of the Spanish Society of Statistics and Operations Research* 13 (February): 335–369.
- Ghoudi, Kilani, Reg J Kulperger, and Bruno Rémillard. 2001. A nonparametric test of serial independence for time series and residuals. *Journal of Multivariate Analysis* 79 (2): 191–218.
- Giraud, Christophe. 2021. *Introduction to high-dimensional statistics*. CRC Press.
- Goix, Nicolas, Anne Sabourin, and Stephan Cléménçon. 2017. Sparse representation of multivariate extremes with applications to anomaly detection. *Journal of Multivariate Analysis* 161:12–31.
- Gröger, M, Lars Arneborg, Christian Dieterich, Anders Höglund, and HEM Meier. 2019. Summer hydrographic changes in the Baltic Sea, Kattegat and Skagerrak projected in an ensemble of climate scenarios downscaled with a coupled regional ocean–sea ice–atmosphere model. *Climate Dynamics* 53:5945–5966.
- Gumbel, Emil J. 1960. Bivariate exponential distributions. *Journal of the American Statistical Association* 55 (292): 698–707.
- Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, et al. 2020. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* 146 (730): 1999–2049.
- Hersbach, H., B. Bell, P. Berrisford, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, et al. 2018. *ERA5 hourly data on single levels from 1959 to present* [in eng].
- Hitz, Adrien, and Robin Evans. 2016. One-component regular variation and graphical modeling of extremes. *Journal of Applied Probability* 53 (3): 733–746.
- Hofert, Marius, Raphaël Huser, and Avinash Prasad. 2018. Hierarchical archimax copulas. *Journal of Multivariate Analysis* 167:195–211.
- Hubert, Lawrence, and Phipps Arabie. 1985. Comparing partitions. *Journal of classification* 2:193–218.
- Hüsler, Jürg, and Rolf-Dieter Reiss. 1989. Maxima of normal random vectors: between independence and complete dependence. *Statistics & Probability Letters* 7 (4): 283–286.
- Janßen, Anja, and Phyllis Wan. 2020. *k*-means clustering of extremes. *Electronic Journal of Statistics* 14 (1): 1211–1233.
- Kojadinovic, Ivan, and Mark Holmes. 2009. Tests of independence among continuous random vectors based on cramér–von mises functionals of the empirical copula process. *Journal of Multivariate Analysis* 100 (6): 1137–1154.
- Kojadinovic, Ivan, and Jun Yan. 2011. Tests of serial independence for continuous multivariate time series based on a möbius decomposition of the independence empirical copula process. *Annals of the Institute of Statistical Mathematics* 63:347–373.
- Lalancette, Michaël, Sebastian Engelke, and Stanislav Volgushev. 2021. Rank-based estimation under asymptotic dependence and independence, with applications to spatial extremes. *The Annals of Statistics* 49 (5): 2552–2576.
- Laloë, Thomas. 2010. L 1-quantization and clustering in banach spaces. *Mathematical Methods of Statistics* 19:136–150.
- . 2021. Quantization based clustering: an iterative approach. *Pattern Recognition Letters* 142:51–57.
- Le, Phuong Dong, Anthony C Davison, Sebastian Engelke, Michael Leonard, and Seth Westra. 2018. Dependence properties of spatial rainfall extremes and areal reduction factors. *Journal of Hydrology* 565:711–719.
- Linder, Tamás. 2002. Learning-theoretic methods in vector quantization. In *Principles of nonparametric learning*, 163–210. Springer.
- Martius, Olivia, Stephan Pfahl, and Clément Chevalier. 2016. A global quantification of compound precipitation and wind extremes. *Geophysical Research Letters* 43 (14): 7709–7717.

- Maume-Deschamps, Véronique, Pierre Ribereau, and Manal Zeidan. 2023. Detecting the stationarity of spatial dependence structure using spectral clustering. <https://hal.science/hal-03918937>, April. <https://hal.science/hal-03918937>.
- Meyer, Nicolas, and Olivier Wintenberger. 2021. Sparse regular variation. *Advances in Applied Probability* 53 (4): 1115–1148.
- Owen, Laura E, Jennifer L Catto, David B Stephenson, and Nick J Dunstone. 2021. Compound precipitation and wind extremes over europe and their relationship to extratropical cyclones. *Weather and Climate Extremes* 33:100342.
- Pappadà, R, F Durante, G Salvadori, and C De Michele. 2018. Clustering of concurrent flood risks via hazard scenarios. *Spatial Statistics* 23:124–142.
- Pfahl, Stephan, and Heini Wernli. 2012. Spatial coherency of extreme weather events in germany and switzerland. *International journal of climatology* 32 (12): 1863–1874.
- Pinto, Joaquim G, Melanie K Karremann, Kai Born, Paul M Della-Marta, and Matthias Klawa. 2012. Loss potentials associated with european windstorms under future climate conditions. *Climate Research* 54 (1): 1–20.
- Rand, William M. 1971. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association* 66 (336): 846–850.
- Raveh-Rubin, Shira, and Heini Wernli. 2015. Large-scale wind and precipitation extremes in the mediterranean: a climatological analysis for 1979–2012. *Quarterly Journal of the Royal Meteorological Society* 141 (691): 2404–2417.
- Resnick, Sidney I. 2007. *Heavy-tail phenomena: probabilistic and statistical modeling*. Springer Science & Business Media.
- Rousseeuw, Peter J. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics* 20:53–65.
- Scarsini, Marco. 1984. On measures of concordance. *Stochastica* 8 (3): 201–218.
- Schwierz, Cornelia, Pamela Köllner-Heck, Evelyn Zenklusen Mutter, David N Bresch, Pier-Luigi Vidale, Martin Wild, and Christoph Schär. 2010. Modelling european winter wind storm losses in current and future climate. *Climatic change* 101:485–514.
- Shorack, Galen R, and Jon A Wellner. 2009. *Empirical processes with applications to statistics*. SIAM.
- Soukissian, Takvor, Flora Karathanasi, Panagiotis Axaopoulos, Evangelos Voukouvalas, and Vassiliki Kotroni. 2018. Offshore wind climate analysis and variability in the Mediterranean Sea. *International Journal of Climatology* 38 (1): 384–402.
- Vignotto, Edoardo, Sebastian Engelke, and Jakob Zscheischler. 2021. Clustering bivariate dependencies of compound precipitation and wind extremes over Great Britain and Ireland. *Weather and climate extremes* 32:100318.
- Wang, Shiyu, Christian Dieterich, Ralf Döscher, Anders Höglund, Robinson Hordoier, HE Markus Meier, Patrick Samuelsson, and Semjon Schimanke. 2015. Development and evaluation of a new regional coupled atmosphere–ocean model in the North Sea and Baltic Sea. *Tellus A: Dynamic Meteorology and Oceanography* 67 (1): 24284.
- Yan, Jun. 2007. Enjoy the joy of copulas: with a package copula. *Journal of Statistical Software* 21:1–21.
- Zscheischler, Jakob, Philippe Naveau, Olivia Martius, Sebastian Engelke, and Christoph C Raible. 2021. Evaluating the dependence structure of compound precipitation and wind speed extremes. *Earth system dynamics* 12 (1): 1–16.



### Appendix A. Axioms for a valid dependence measure

In this section, we recall the axiomatic framework to quantify dependence between multiple groups of random variables of possibly different sizes De Keyser and Gijbels 2023b. We recall that we consider a random vector  $\mathbf{Z}$  with distribution  $F$  that is in the max domain of attraction of and EVD  $H$ . Plausible axioms, for a valid dependence measure of a random vector  $\mathbf{X}$ , denoted as  $\mathcal{D}(\mathbf{X})$ , are as follows

- (A1) For every permutation  $\pi$  of  $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(d)}$ :  $\mathcal{D}(\mathbf{X}) = \mathcal{D}(\pi(\mathbf{X}))$ ; and for every permutation  $\pi_j$  of  $X^{(j,1)}, \dots, X^{(j,p_j)}$ , for  $j \in \{1, \dots, d\}$ , it holds:

$$\mathcal{D}(\mathbf{X}) = \mathcal{D}(\mathbf{X}^{(1)}, \dots, \pi^{(j)}(\mathbf{X}^{(j)}), \dots, \mathbf{X}^{(d)}).$$

- (A2)  $0 \leq \mathcal{D}(\mathbf{X}) \leq 1$ .

- (A3)  $\mathcal{D}(\mathbf{X}) = 0$  if and only if  $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(d)}$  are mutually independent.

- (A4)  $\mathcal{D}(\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(d)}, \mathbf{X}^{(d+1)})$  with equality if and only if  $\mathbf{X}^{(d+1)}$  is independent of  $(\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(d)})$ .

- (A5)  $\mathcal{D}(\mathbf{X})$  is well defined for any  $q$ -dimensional random vector  $\mathbf{X}$  and is a functional of solely the copula  $C$  of  $\mathbf{X}$ .

- (A6) Let  $T^{(j,\ell)}$  for  $j = 1, \dots, d$  and  $\ell = 1, \dots, p_j$  be strictly increasing, continuous transformations. Then

$$\mathcal{D}(\mathbf{T}^{(1)}(\mathbf{X}^{(1)}), \dots, \mathbf{T}^{(d)}(\mathbf{X}^{(d)})) = \mathcal{D}(\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(d)}),$$

where  $\mathbf{T}^{(j)} = (T^{(j,1)}(X^{(j,1)}), \dots, T^{(j,p_j)}(X^{(j,p_j)}))$  for  $j = 1, \dots, d$ .

- (A7) Let  $T^{(j,\ell)}$  be a strictly decreasing, continuous transformation for a fixed  $j \in \{1, \dots, d\}$  and a fixed  $\ell \in \{1, \dots, p_j\}$ . Then

$$\mathcal{D}(\mathbf{X}^{(1)}, \dots, T^{(j,\ell)}(\mathbf{X}^{(j)}), \dots, \mathbf{X}^{(d)}) = \mathcal{D}(\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(d)}),$$

where  $T^{(j,\ell)}(\mathbf{X}^{(j)}) = (X^{(j,1)}, \dots, T^{(j,\ell)}(X^{(j,\ell)}), \dots, X^{(j,p_j)})$ .

- (A8) Let  $(\mathbf{X}_n)_{n \in \mathbb{N}}$  be a sequence of  $q$ -dimensional reduction random vectors having copulas  $(C_n)_{n \in \mathbb{N}}$ , then

$$\lim_{n \rightarrow \infty} \mathcal{D}(\mathbf{X}_n) = \mathcal{D}(\mathbf{X})$$

if  $C_n \rightarrow C$  uniformly, where  $C$  denotes the copula of  $\mathbf{X}$ .

Having those necessary materials, we now detail below which axioms hold for the SECO metric to measure the dependence among extreme of random vectors.

**Proposition 1.** *Let SECO be the metric defined in (6). Then, it satisfies the system of axioms (A1), (A2) with the following bounds*

$$0 \leq \text{SECO}(\mathbf{Z}^{(1)}, \dots, \mathbf{Z}^{(d)}) \leq \min_{j=1, \dots, d} \left\{ \sum_{k \neq j} \theta(k) \right\},$$

(A3), (A4), (A5), (A6) and (A8) stated in De Keyser and Gijbels 2023b.

**Proof** The reason why Property (A1) holds is due to the fact that the set union and addition of numbers are commutative. Proposition 7 and 8 from Boulin et al. 2023 imply the results stated about (A2) for the lower bound and (A3), where the latter is related to asymptotic independence. To obtain the upper bound, it is observed that

$$0 \leq \theta(1, \dots, d) \leq \min\{\theta(1), \dots, \theta(d)\}.$$

For property (A4), the inequality

$$\text{SECO}(\mathbf{Z}^{(1)}, \dots, \mathbf{Z}^{(d)}, \mathbf{Z}^{(d+1)}) \geq \text{SECO}(\mathbf{Z}^{(1)}, \dots, \mathbf{Z}^{(d)})$$

can be rewritten as

$$\theta(1, \dots, d) + \theta(d+1) \geq \theta(1, \dots, d, d+1),$$

which holds true by Proposition 7 of Boulin et al. 2023. Moreover, this inequality holds as an equality if and only if  $\mathbf{Z}^{(d+1)}$  is asymptotically independent of  $(\mathbf{Z}^{(1)}, \dots, \mathbf{Z}^{(d)})$ , according to Proposition 8 of Boulin et al. 2023.

We have for every  $u \in (0, 1)$ ,

$$\theta(1, \dots, d) = \ln(C(u, \dots, u)) / \ln(u),$$

where  $C$  is the extreme value copula in the domain of attraction of  $\mathbf{Z}$ , i.e.,

$$C(u^{(1,1)}, \dots, u^{(d,pd)}) = \exp \left\{ -L \left( -\ln u^{(1,1)}, \dots, -\ln u^{(d,pd)} \right) \right\}.$$

Hence (A5) and (A6) are fulfilled. Let  $(\mathbf{Z}_n)_{n \in \mathbf{N}}$  be a sequence of random vector, for each  $n$ , suppose that  $\mathbf{Z}_n$  is in the max-domain of attraction of a random vector  $H_n$  an extreme value distribution with extreme value copula  $C_n$ . Suppose that the sequence of extreme value copulae  $(C_n)_{n \in \mathbf{N}}$  converges uniformly to  $C$ , where  $C$  is the extreme value copula of  $H$ . Then, for every  $\mathbf{u} \in (0, 1)$ , we have the point wise convergence:

$$C_n(\mathbf{u}) \rightarrow C(\mathbf{u}), \quad n \rightarrow \infty.$$

Thus, as  $\ln : (0, 1) \rightarrow \mathbb{R}_-$  is continuous, we have

$$\ln(C_n(\mathbf{u})) \rightarrow \ln(C(\mathbf{u})), \quad n \rightarrow \infty.$$

Then

$$\text{SECO}(\mathbf{Z}_n^{(1)}, \dots, \mathbf{Z}_n^{(d)}) \rightarrow \text{SECO}(\mathbf{Z}^{(1)}, \dots, \mathbf{Z}^{(d)}), \quad n \rightarrow \infty.$$

□

## Appendix B. A coherent measure for extreme value random vectors

In this section, we borrow the notation previously introduced in the specific case where  $d = 2$  and  $p_a = p_b = p$ . Let  $\mathbf{Z}$  and  $\mathbf{W}$  be two random vectors with marginal random vectors  $\mathbf{Z}^{(a)}$  and  $\mathbf{Z}^{(b)}$  (resp.  $\mathbf{W}^{(a)}$  and  $\mathbf{W}^{(b)}$ ), each having  $p$  components. We set  $q = 2p$ . Assume that  $\mathbf{Z}$  and  $\mathbf{W}$  are in the max-domain of attraction of two extreme value distributions with stable tail dependence function  $L_{\mathbf{Z}}$  and  $L_{\mathbf{W}}$ . Here,  $L_{\mathbf{Z}}$  and  $L_{\mathbf{W}}$  are nested, and for  $\mathbf{x} \in [0, \infty)^q$ ,  $\mathbf{x}^{(a)} = (x^{(a,1)}, \dots, x^{(a,p)})$ ,  $\mathbf{x}^{(b)} = (x^{(b,1)}, \dots, x^{(b,p)})$ , we have

$$L_{\mathbf{Z}}(\mathbf{x}) = L_{\mathbf{Z}}^{(0)} \left( L_{\mathbf{Z}}^{(a)}(\mathbf{x}^{(a)}), L_{\mathbf{Z}}^{(b)}(\mathbf{x}^{(b)}) \right), \quad L_{\mathbf{W}}(\mathbf{x}) = L_{\mathbf{W}}^{(0)} \left( L_{\mathbf{W}}^{(a)}(\mathbf{x}^{(a)}), L_{\mathbf{W}}^{(b)}(\mathbf{x}^{(b)}) \right),$$

where  $L_{\mathbf{Z}}^{(0)}$  and  $L_{\mathbf{Z}}^{(j)}$  are the “mother” stable tail dependence function and the stable tail dependence functions associated with the extreme value random vectors  $\mathbf{X}^{(j)}$ , respectively. Similarly,  $L_{\mathbf{W}}^{(0)}$  and  $L_{\mathbf{W}}^{(j)}$  represent the stable tail dependence functions associated with the extreme value random vectors  $\mathbf{Y}^{(j)}$  for  $j = a, b$ . The SECO for these two models thus reduces to

$$\begin{aligned} \text{SECO}(\mathbf{Z}^{(a)}, \mathbf{Z}^{(b)}) &= L_{\mathbf{Z}}^{(a)}(\mathbf{1}^{(a)}) + L_{\mathbf{Z}}^{(b)}(\mathbf{1}^{(b)}) - L_{\mathbf{Z}}^{(0)} \left( L_{\mathbf{Z}}^{(a)}(\mathbf{1}^{(a)}), L_{\mathbf{Z}}^{(b)}(\mathbf{1}^{(b)}) \right), \\ \text{SECO}(\mathbf{W}^{(a)}, \mathbf{W}^{(b)}) &= L_{\mathbf{W}}^{(a)}(\mathbf{1}^{(a)}) + L_{\mathbf{W}}^{(b)}(\mathbf{1}^{(b)}) - L_{\mathbf{W}}^{(0)} \left( L_{\mathbf{W}}^{(a)}(\mathbf{1}^{(a)}), L_{\mathbf{W}}^{(b)}(\mathbf{1}^{(b)}) \right) \end{aligned}$$

Taking advantage of the definition introduced by Scarsini 1984, we say that  $\mathbf{Z}$  is more concordant than  $\mathbf{W}$  in the max-domain of attraction if for every possible value of  $\mathbf{x} \in \mathbb{R}^q$ ,  $L_{\mathbf{Z}}^{(0)}(\mathbf{x}) \leq L_{\mathbf{W}}^{(0)}(\mathbf{x})$ . Now, let's suppose that  $\mathbf{Z}$  and  $\mathbf{W}$  have the same marginal extremal dependence structure, which means that their stable tail dependence functions are the same for both components, i.e.,

$$L_{\mathbf{Z}}^{(a)}(\mathbf{x}^{(a)}) = L_{\mathbf{W}}^{(a)}(\mathbf{x}^{(a)}), \quad L_{\mathbf{Z}}^{(b)}(\mathbf{x}^{(b)}) = L_{\mathbf{W}}^{(b)}(\mathbf{x}^{(b)}), \quad \mathbf{x} \in \mathbb{R}^q.$$

Thus, we have

$$\begin{aligned} \text{SECO}(\mathbf{Z}^{(a)}, \mathbf{Z}^{(b)}) - \text{SECO}(\mathbf{W}^{(a)}, \mathbf{W}^{(b)}) = \\ L_{\mathbf{W}}^{(0)}\left(L_{\mathbf{W}}^{(a)}(\mathbf{1}^{(a)}), L_{\mathbf{W}}^{(b)}(\mathbf{1}^{(b)})\right) - L_{\mathbf{Z}}^{(0)}\left(L_{\mathbf{W}}^{(a)}(\mathbf{1}^{(a)}), L_{\mathbf{W}}^{(b)}(\mathbf{1}^{(b)})\right) \end{aligned}$$

which is positive. Thus the SECO is a coherent measure given the marginal random vector's dependence structure is the same.

To better understand the behaviour of the SECO, we will consider two specific nested models: the nested Gumbel and Hüsler-Reiss models, for which  $p = 2$  and  $d = 2$ . For these models, we will analyse the corresponding stable tail dependence functions.

$$\begin{aligned} L^{\text{Gu}}(x_1, x_2) &= \left(x_1^{1/\alpha} + x_2^{1/\alpha}\right)^\alpha, \\ L^{\text{HR}}(x_1, x_2) &= \Phi\left(\frac{\lambda}{2} + \frac{x_2 - x_1}{\lambda}\right)x_1 + \Phi\left(\frac{\lambda}{2} + \frac{x_1 - x_2}{\lambda}\right)x_2, \end{aligned}$$

with  $\alpha \in [0, 1]$ ,  $\lambda \in [0, \infty)$  and  $\Phi$  denotes the cumulative distribution function of a standard normal random variable. The first stable tail dependence function is known as the Gumbel (or Logistic) distribution introduced by Gumbel 1960, the parameter  $\alpha$  represent the strength of dependence: if  $\alpha \rightarrow 0$ , then the two random variables are comonotone while if  $\alpha = 1$  it reduces to independence. The second stable tail dependence function is the Hüsler Reiss distribution (Hüsler and Reiss 1989), both asymptotic comonotony and independence are depicted by the respective limits  $\lambda \rightarrow 0$  and  $\lambda \rightarrow \infty$ .

For the nested Gumbel model, the SECO is equal to

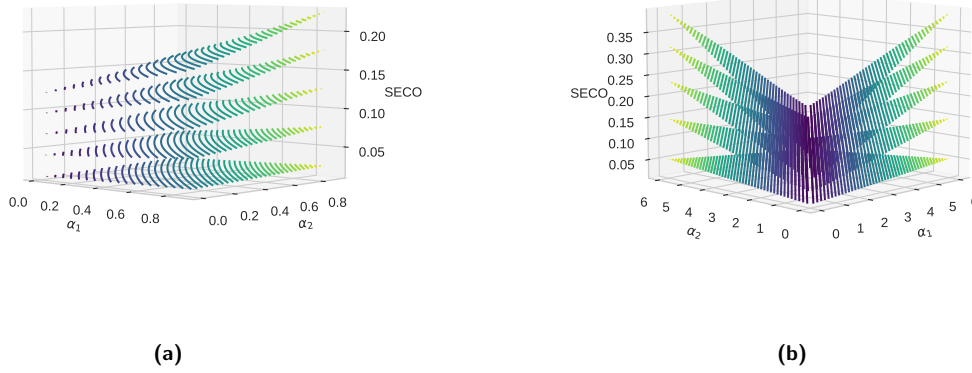
$$2^{\alpha_a} + 2^{\alpha_b} - \left(2^{\alpha_a/\alpha_0} + 2^{\alpha_b/\alpha_0}\right)^{\alpha_0}. \quad (16)$$

This equation involves the parameters  $\alpha_0, \alpha_a$ , and  $\alpha_b$ , which correspond to the mother, the first, and the second random vector margins, respectively. In Fig. 8 are depicted level sets of (16) and the normalised version, i.e., (16) divided by  $\min\{2^{\alpha_a}, 2^{\alpha_b}\}$ , according to  $\alpha_0$  for different dependence structure between the marginal random vectors. Additionally, we can also calculate the SECO for the Hüsler-Reiss nested model which equals to

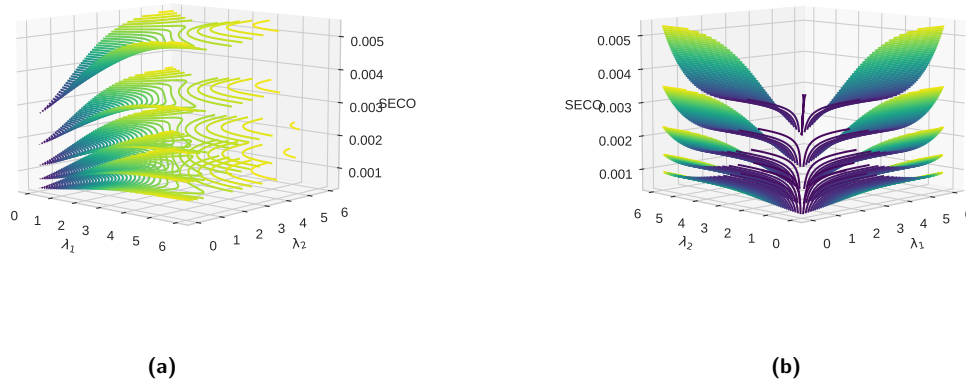
$$\theta^{(a)} + \theta^{(b)} - \left[ \Phi\left(\frac{\lambda_0}{2} + \frac{\theta^{(b)} - \theta^{(a)}}{\lambda_0}\right)\theta^{(a)} + \Phi\left(\frac{\lambda_0}{2} + \frac{\theta^{(a)} - \theta^{(b)}}{\lambda_0}\right)\theta^{(b)} \right], \quad (17)$$

where  $\theta^{(j)} = 2\Phi(\lambda_j/2)$ ,  $j = a, b$  are the extremal coefficients and includes several parameters such as  $\lambda_a, \lambda_b$ , and  $\lambda_0$  which correspond to the first, the second random vector margins and the mother respectively. In Fig. 9, we represent level sets of (17) and the normalised version of (17), that is divided by  $\min\{\theta^{(a)}, \theta^{(b)}\}$ , the bound found in Proposition 1.

In both Fig. 8 and Fig. 9, the behaviour of SECO exhibits similarities, particularly in relation to the level sets. These level sets are characterised by their monotonicity, whereby the level set's



**Figure 8.** Level sets of the SECO for the nested Logistic model (see (16)) in Panel 8a and the normalised SECO in Panel 8b for  $\alpha_0 \in \{0.91, 0.93, 0.95, 0.97, 0.99\}$  and  $\alpha_1, \alpha_2$  range in  $\{0.01, 0.02, \dots, 0.9\}$ . The coherence property ensures that the level sets are arranged in ascending order based on the values of  $\alpha_0$ .



**Figure 9.** Level sets of the SECO for the nested Hüsler-Reiss model (see (17)) in Panel 9a and the normalised SECO in Panel 9b for  $\lambda_0 \in \{6.0, 6.25, 6.5, 6.75, 7.0\}$  and  $\lambda_1, \lambda_2$  range in  $\{0.01, 0.02, \dots, 6.0\}$ . The coherence property ensures that the level sets are arranged in ascending order based on the values of  $\lambda_0$ .

height increases with the degree of dependence among the marginal random vectors. In other words, the greater the dependence between the marginal random vectors, the higher the level set will be. Additionally, the SECO demonstrates monotonicity through  $\alpha_1, \alpha_2$  (or alternatively,  $\lambda_1, \lambda_2$ ), whereby the value of the SECO increases as the random variables in the marginal random vectors become more concordant. In simpler terms, the more closely related the random variables are, the higher the SECO will be. Moving on to the normalised SECO, we observe that the highest values are attained when the random variables within the marginal random vectors display asymmetric behaviour, meaning that one variable is comonotonic while the other is independent. Conversely, the lowest values are obtained when both variables share the same dependence structure.

### Appendix C. Incompleteness tail dependence structure estimation in high dimension

Throughout the section, assume that we have  $\mathbf{Z}_1, \dots, \mathbf{Z}_n$  identically distributed observations of the  $d$ -dimensional random vector  $\mathbf{Z}$ , which is in the max-domain of attraction of  $H$ , an EVD where each of the components of  $\mathbf{Z}$  are asymptotically independent. Let  $R_{n,i}^{(j)}$  denote the rank of  $Z_i^{(j)}$  among  $Z_1^{(j)}, \dots, Z_n^{(j)}$ ,  $i = 1, \dots, n$ ,  $j = 1, \dots, d$ . For  $k \in \{1, \dots, n\}$ , define a nonparametric estimator of  $\theta$ , the extremal coefficient by

$$\hat{\theta}_{n,k}^{\text{EKS}} := \frac{1}{k} \sum_{i=1}^n \mathbb{1}_{\{R_{n,i}^{(1)} > n+0.5-k \text{ or } \dots \text{ or } R_{n,i}^{(d)} > n+0.5-k\}},$$

see Einmahl, Krajina, and Segers 2012.

**Lemma 1.** For  $k \in \{1, \dots, n\}$  we have

$$\hat{\theta}_{n,k}^{\text{EKS}} \leq \frac{n}{k}$$

**Proof** The upper bound is trivial since

$$\sum_{i=1}^n \mathbb{1}_{\{R_{n,i}^{(1)} > n+0.5-k \text{ or } \dots \text{ or } R_{n,i}^{(d)} > n+0.5-k\}} \leq n.$$

□

Now, let us divide the sample of size  $n$  of  $\mathbf{Z}$  into  $k$  blocks of length  $m$ , so that  $k = n/m$  (where we suppose, without loss of generality that  $m$  divide  $n$ ). For the  $i$ th block, the maximum value in the  $j$ -component is denoted by

$$M_{m,i}^{(j)} = \max\{Z_t^{(j)} : t \in (im - m, im]\}.$$

Let  $R_{n,m,i}^{(j)}$  denote the rank of  $M_{m,i}^{(j)}$  among  $M_{m,1}^{(j)}, \dots, M_{m,j}^{(j)}$ ,  $i = 1, \dots, n$ ,  $j = 1, \dots, d$ . Define a non parametric estimator of the multivariate madogram

$$\hat{v}_{n,m} := \frac{1}{k} \sum_{i=1}^k \left[ \bigvee_{j=1}^d \frac{R_{n,m,i}^{(j)}}{k+1} - \frac{1}{d} \sum_{j=1}^d \frac{R_{n,m,i}^{(j)}}{(k+1)} \right]$$

**Lemma 2.** For  $m \in \{1, \dots, n\}$ , we have

$$\hat{v}_{n,m} \leq \frac{k}{k+1} - \frac{1}{2}.$$

Consequently,

$$\hat{\theta}_{n,m}^{\text{MAD}} \leq \frac{n}{m}.$$

**Proof** One can easily deduce the following upper bound

$$\bigvee_{j=1}^d \frac{R_{n,m,i}^{(j)}}{k+1} \leq \frac{k}{k+1}.$$

Thus we obtain that

$$\hat{v}_{n,m} \leq \frac{k}{k+1} - \frac{1}{k} \frac{1}{d} \sum_{j=1}^d \sum_{i=1}^k \frac{R_{n,m,i}^{(j)}}{k+1}.$$

The right hand side of the equation is equal to

$$\frac{1}{k} \frac{1}{d} \sum_{j=1}^d \sum_{i=1}^k \frac{R_{n,m,i}^{(j)}}{k+1} = \frac{1}{k} \frac{1}{d} \sum_{j=1}^d \frac{k(k+1)}{2(k+1)} = \frac{1}{2}.$$

Let us consider the following function

$$f: \left[ 0, \frac{k}{k+1} - \frac{1}{2} \right] \rightarrow \mathbb{R}$$

$$x \mapsto \frac{0.5 + x}{0.5 - x}.$$

Since it is an nondecreasing function, we must have

$$\hat{\theta}_{n,m}^{\text{MAD}} \leq f\left(\frac{k}{k+1} - \frac{1}{2}\right) = \frac{n}{m}$$

□

The consequence of both lemmas is that the nonparametric and the madogram-based estimator of the extremal coefficient can only recover values in the following range:

$$1 \leq \hat{\theta}_{n,k}^{\text{EKS}} \leq \frac{n}{k}, \quad 1 \leq \hat{\theta}_{n,m}^{\text{MAD}} \leq \frac{n}{m}.$$

If we suppose that  $d > n/k$  or  $d > n/m$ , then both estimators cannot expect to retrieve dependencies above the thresholds stated by our two lemmas. In particular, in high dimension, i.e., when  $d > n$ , these estimators are unable to detect asymptotic independence. Indeed, in cases where the  $d$  variables are asymptotically independent, extremes occur in one variable without influencing extremes in the others. Thus, when  $d > n$ , it is highly probable to observe a rank that is greater than  $n - k + 0.5$  for at least one variable and this occurs for every observation  $i = 1, \dots, n$ . Since this happens for every observation  $i = 1, \dots, n$ , the characteristic function is (with high probability) always equal to one, resulting in an overall extremal coefficient equal to  $n/k$  when taking the sum. However, in high dimensions, this cannot be equal to  $d$ , the value taken in asymptotic independence.

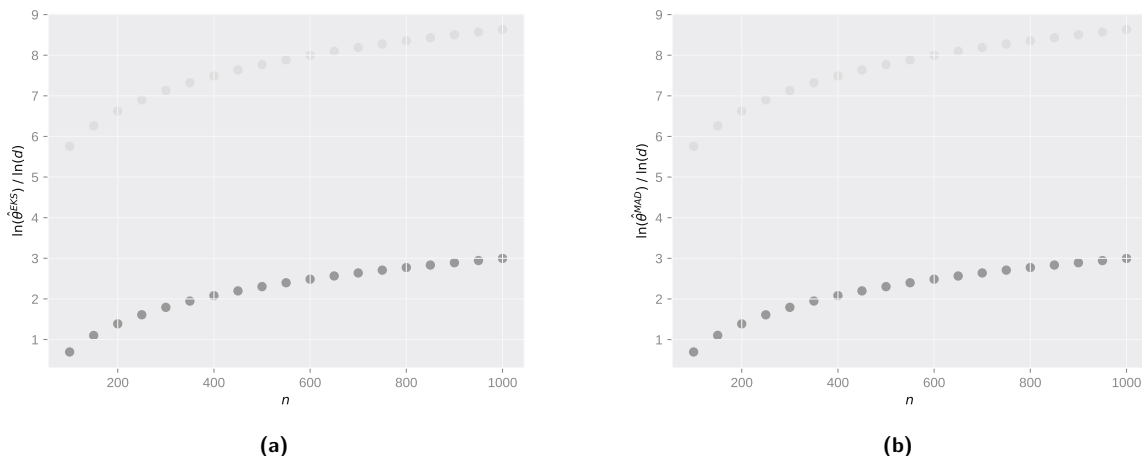
To illustrate these findings, we consider the following numerical setup. Consider as the sample size  $n \in \{100, 150, \dots, 1000\}$  and the high dimensional setting given by  $d = n^{1.25}$  where we want to estimate the extremal coefficient of the random vector of  $\mathbf{Z}$  where its components are asymptotically independent. In this setup, we know that the theoretical value of  $\theta$  is given by  $d$ . When studying the dependence structure of extreme events, we face the "curse of dimensionality" in two ways. Firstly, traditional estimators do not cover the full spectrum of possible values. Secondly, in order to expand the range of values, one may need to reduce the number of extremes considered, denoted by  $k$ , or decrease the size of block maxima, denoted by  $m$ . However, this may lead to an increase in variance or a decrease in bias, depending on the estimator used.

In Fig. 10, the bias of the estimator is clearly depicted. For each value of  $n$ , both estimators reach their upper bounds, that is  $\hat{\theta}_{n,k}^{\text{EKS}} = n/k$  and  $\hat{\theta}_{n,m}^{\text{MAD}} = n/m$ .

#### Appendix D. Consistent estimation of SECO

Consider a  $q$ -dimensional random vector  $\mathbf{Z} = (Z^{(1,1)}, \dots, Z^{(d,p_d)})$ , where  $p_1 + \dots + p_d = q$ . This vector has a joint cumulative distribution function (c.d.f.) denoted as  $F$ , and each of its components has continuous marginal c.d.f.s  $F^{(1,1)}, \dots, F^{(d,p_d)}$ . The copula, denoted as  $C$ , associated with  $F$





**Figure 10.** Estimator  $\hat{\theta}_{n,k}^{\text{EKS}}$  (Panel 10a, black dots) and  $\hat{\theta}_{n,m}^{\text{MAD}}$  (Panel 10b, black dots) according to different values of  $n \in \{100, 150, \dots, 1000\}$  and  $d = n^{1.25}$ . The theoretical value of the extremal coefficient is depicted in grey dots. For both estimators, we took  $k = m = 50$ .

(or equivalently, with  $\mathbf{Z}$ ) is defined as the c.d.f. of another random vector  $\mathbf{U} = (U^{(1,1)}, \dots, U^{(d,p_d)})$ . These random variables  $\mathbf{U}$  are obtained through the marginal application of the probability integral transform, meaning that  $U^{(j,\ell)} = F^{(j,\ell)}(Z^{(j,\ell)})$  for  $j = 1, \dots, d$  and  $\ell = p_1, \dots, p_d$ . Notably, the marginal c.d.f.s of the copula  $C$  are uniformly distributed on the interval  $[0, 1]$ . According to Sklar's theorem, the copula  $C$  is a unique function that satisfies the following relationship for all  $\mathbf{x} = (x^{(1,1)}, \dots, x^{(d,p_d)}) \in \mathbb{R}^q$ :

$$F(x^{(1,1)}, \dots, x^{(d,p_d)}) = C \left\{ F^{(1,1)}(x^{(1,1)}), \dots, F^{(d,p_d)}(x^{(d,p_d)}) \right\}.$$

In simpler terms, this equation describes how the joint distribution  $F$  can be represented in terms of the copula  $C$  and the marginal distributions  $F^{(j,\ell)}$ .

Now, let us consider a sequence of observed data,  $\mathbf{Z}_i$  for  $i = 1, \dots, n$ , which represents a stationary time series. Importantly, each  $\mathbf{Z}_i$  follows the same distribution as  $\mathbf{Z}$ . Set  $\mathbf{U}_i = (U_i^{(1,1)}, \dots, U_i^{(d,p_d)}) \sim C$  with  $U_i^{(j,\ell)} = F^{(j,\ell)}(Z_i^{(j,\ell)})$ . Define

$$\alpha_n^{(j)}(\mathbf{u}) = \sqrt{n} \left( G_n^{(j)}(u) - u \right), \quad G_n^{(j)}(u) = n^{-1} \sum_{i=1}^n \mathbf{1}_{\{U_i^{(j)} \leq u\}},$$

denote the (unobservable) empirical processes based on  $\mathbf{U}_1, \dots, \mathbf{U}_n$ .

For any sequence  $(\mathbf{Z}_n, n \in \mathbb{N})$ , let

$$\mathcal{F}_k = \sigma(\mathbf{Z}_n, n \leq k), \quad \text{and} \quad \mathcal{G}_k = \sigma(\mathbf{Z}_n, n \geq k),$$

be the natural filtration and "reverse" filtration of the sequence  $(\mathbf{Z}_n, n \in \mathbb{N})$ . Define

$$\beta(\mathcal{A}_1, \mathcal{A}_2) = \sup \frac{1}{2} \sum_{i,j \in I \times J} |\mathbb{P}(A_i \cap B_j) - \mathbb{P}(A_i)\mathbb{P}(B_j)|,$$

where the sup is taken over all finite  $(A_i)_{i \in I}$  and  $(B_j)_{j \in J}$  of  $\Omega$  with the sets  $A_i$  (resp.  $B_j$ ) in the sigma field  $\mathcal{A}_1$  (resp.  $\mathcal{A}_2$ ). The  $\beta$ -mixing (or completely regular) coefficient is defined as

$$\beta(\ell) = \sup_{n \in \mathbb{N}} \beta(\mathcal{F}_n, \mathcal{G}_{n+\ell}). \quad (18)$$

For the formulation of the consistency result for our estimator of the SECO, we need a couple of conditions over the regularity of the sequence  $(\mathbf{Z}_n, n \in \mathbb{N})$  which are the following:

**Condition A.** There exists an intermediary sequence  $m = m_n$  such that  $m_n = o(n)$  and  $\beta(m_n) \rightarrow 0$  as  $n \rightarrow \infty$ , where  $\beta$  is defined in (18).

**Condition B.** There exists some  $\theta_1 \in (0, 1/2]$  such that, for all  $\mu \in (0, \theta_1]$  and all sequences  $\delta_n \rightarrow 0$ , we have

$$M_n(\delta_n, \mu) := \sup_{|u-v| \leq \delta_n} \frac{|\alpha_n^{(j)}(u) - \alpha_n^{(j)}(v)|}{\max\{|u-v|^\mu, n^{-\mu}\}} = o_{\mathbb{P}}(1), \quad j = 1, \dots, d$$

Condition B can, for instance, be verified in the i.i.d. case with  $\theta_1 = 1/2$  or for  $\beta$ -mixing sequence with  $\beta(n) = o(n^{-a})$  as  $n \rightarrow \infty$  for some  $a \in (0, 1)$ , see Proposition 4.4 of Berghaus, Bücher, and Volgushev 2017. Here we state the proposition that the actual appendix is devoted to prove.

**Proposition 2.** *Let  $k = k_n$  be an intermediary sequence. Provided that  $k \rightarrow \infty$ ,  $k/n \rightarrow 0$  as  $n \rightarrow \infty$ , under the regularity condition over the sequence  $(\mathbf{Z}_n, n \in \mathbb{N})$  stated by Condition A and Condition B and  $\mathbf{Z}$  is in the max-domain of attraction of  $H$ , then*

$$\widehat{\text{SECO}}(\mathbf{Z}^{(a)}, \mathbf{Z}^{(b)}) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \text{SECO}(\mathbf{Z}^{(a)}, \mathbf{Z}^{(b)}).$$

**Proof** Since convergence in probability remains stable through addition and subtraction, our task is to establish the consistency of each component:  $\hat{\theta}(a)$ ,  $\hat{\theta}(b)$ , and  $\hat{\theta}(a, b)$ . We address this concern in Lemma 3 below, where, without loss of generality, we focus on demonstrating the consistency of  $\hat{\theta}(a)$ . Also, without loss of generality, we study the following estimator

$$\hat{\theta}(a) = \frac{1}{k} \sum_{i=1}^n \mathbb{1}_{\{\hat{F}_n^{(a,1)}(Z_i^{(a,1)}) > 1-k/n \text{ or } \dots \text{ or } \hat{F}_n^{(a,p)}(Z_i^{(a,p)}) > 1-k/n\}}, \quad (19)$$

where  $\hat{F}_n^{(a,\ell)}$  denotes the empirical distribution function of  $Z_1^{(a,\ell)}, \dots, Z_n^{(a,\ell)}$  for  $\ell = 1, \dots, p_a$ . This estimator mirrors the one presented in (8), with the exception that we employ uniform margins instead of ranks, and we omit the constant factor of  $1/2$ , which, crucially, does not alter the estimator's asymptotic behavior.

**Lemma 3.** *Under the conditions of Proposition 2, we have*

$$\hat{\theta}(a) = \theta_q(a) + o_{\mathbb{P}}(1),$$

with

$$\theta_t(a) = \mathbb{P} \left\{ F^{(a,1)}(Z^{(a,1)}) > 1-t \text{ or } \dots \text{ or } F^{(a,p)}(Z^{(a,p)}) > 1-t \right\} / t,$$

and  $\lim_{t \rightarrow 0} \theta_t(a) = \theta(a)$ .

**Proof** Without confusions, we set in this proof  $p_a = p$  and that  $\mathbf{Z} = (Z^{(1)}, \dots, Z^{(p)}) := (Z^{(a,1)}, \dots, Z^{(a,p)})$  is a  $p$ -dimensional random vector. We begin by introducing some useful notations. In the same spirit, define the random variable  $U_i^{(j)} = F^{(j)}(Z_i^{(j)})$  (here  $Z_i^{(j)}$  denote the  $j$ th entry of the vector  $\mathbf{Z}_i$ ) and the vectors  $\mathbf{U}_i := (U_i^{(1)}, \dots, U_i^{(p)})$  with stationary

distribution  $C$ . Denote by  $\hat{G}_n^{(j)}$  the empirical distribution of  $U_1^{(j)}, \dots, U_n^{(j)}$ . Define the vector  $\hat{G}_n^{\leftarrow}(\mathbf{x}) = ((\hat{G}_n^{(1)})^{\leftarrow}(x^{(1)}), \dots, (\hat{G}_n^{(p)})^{\leftarrow}(x^{(p)}))$ , the function

$$C_n^o(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{U_i^{(1)} \leq x^{(1)}, \dots, U_i^{(p)} \leq x^{(p)}\}}$$

and

$$\hat{C}_n \left( 1 - \frac{k\mathbf{x}}{n} \right) := C_n^o \left( \hat{G}_n^{\leftarrow} \left( 1 - \frac{k\mathbf{x}}{n} \right) \right)$$

Note that the estimator  $\hat{\theta}$  depends only on the marginal ranks of  $Z_i^{(j)}$  with  $j = 1, \dots, p$ ; thus we have almost surely

$$|\hat{\theta}(a) - \theta_t(a)| = \left| \frac{n}{k} \hat{C}_n(1 - k/n, \dots, 1 - k/n) - t^{-1} C(1 - t, \dots, 1 - t) \right|$$

Standard arguments gives that under  $k \rightarrow \infty$  and  $k/n \rightarrow t \in (0, 1)$  and  $n \rightarrow \infty$ , the right hand side of the latter equation is equal to

$$\left| \frac{n}{k} \left( \hat{C}_n \left( 1 - \frac{k}{n}, \dots, 1 - \frac{k}{n} \right) - C \left( 1 - \frac{k}{n}, \dots, 1 - \frac{k}{n} \right) \right) \right| + o_{\mathbb{P}}(1).$$

Now, we can bound the first term by

$$\frac{n}{k} \left| C_n^o(\hat{G}_n^{\leftarrow}(1 - k\mathbf{1}/n)) - C(\hat{G}_n^{\leftarrow}(1 - k\mathbf{1}/n)) \right| + \frac{n}{k} \left| C(\hat{G}_n^{\leftarrow}(1 - k\mathbf{1}/n)) - C(1 - k\mathbf{1}/n) \right|.$$

Using Lipschitz continuity of  $C$ , we obtain the following upper bound:

$$\frac{n}{k} \|C_n^o - C\|_{\infty} + \frac{n}{k} \sum_{j=1}^d \|u_n^{(j)}\|_{\infty} \quad (20)$$

where  $\|\cdot\|_{\infty}$  is the uniform norm and  $u_n^{(j)}(u) = (\hat{G}_n^{(j)})^{\leftarrow}(u) - u$ ,  $u \in [0, 1]$ .

By Berbee's coupling Lemma (Doukhan, Massart, and Rio 1995; Bücher and Segers 2014), one can construct inductively a sequence  $(\bar{\mathbf{Z}}_{im+1}, \dots, \bar{\mathbf{Z}}_{im+m})_{i \geq 0}$  such that the following three properties hold:

- (i)  $(\bar{\mathbf{Z}}_{im+1}, \dots, \bar{\mathbf{Z}}_{im+m}) \stackrel{d}{=} (\mathbf{Z}_{im+1}, \dots, \mathbf{Z}_{im+m})$  for any  $i \geq 0$ ;
- (ii) both  $(\bar{\mathbf{Z}}_{2im+1}, \dots, \bar{\mathbf{Z}}_{2im+m})_{i \geq 0}$  and  $(\bar{\mathbf{Z}}_{(2i+1)m+1}, \dots, \bar{\mathbf{Z}}_{(2i+1)m+m})_{i \geq 0}$  sequences are independent and identically distributed;
- (iii)  $\mathbb{P}\{(\bar{\mathbf{Z}}_{im+1}, \dots, \bar{\mathbf{Z}}_{im+m}) \neq (\mathbf{Z}_{im+1}, \dots, \mathbf{Z}_{im+m})\} \leq \beta(m)$ .

Let  $\bar{C}_n^o$  be defined analogously to  $C_n^o$  but with  $\mathbf{Z}_1, \dots, \mathbf{Z}_n$  replaced by  $\bar{\mathbf{Z}}_1, \dots, \bar{\mathbf{Z}}_n$ . Now write

$$C_n^o(\mathbf{u}) - C(\mathbf{u}) = \{C_n^o(\mathbf{u}) - \bar{C}_n^o(\mathbf{u})\} + o_{\mathbb{P}}(1). \quad (21)$$

The term in brackets in the right hand side is  $o_{\mathbb{P}}(1)$  uniformly in  $\mathbf{u}$ , since

$$|C_n^o(\mathbf{u}) - \bar{C}_n^o(\mathbf{u})| \leq \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{\bar{\mathbf{Z}}_i \neq \mathbf{Z}_i\}}.$$

Hence by Markov's inequality, for any  $\epsilon > 0$

$$\mathbb{P} \left\{ \sup_{\mathbf{u} \in [0,1]^q} |C_n^o(\mathbf{u}) - \bar{C}_n(\mathbf{u})| > \epsilon \right\} \leq \frac{\beta(m)}{\epsilon}.$$

By Condition  $\mathcal{A}$ , we obtain that the first summand in brackets in (21) is  $o_{\mathbb{P}}(1)$  as  $n \rightarrow \infty$ , uniformly in  $\mathbf{u} \in [0,1]^q$ . We obtain that

$$\|C_n^o - C\|_{\infty} = o_{\mathbb{P}}(1).$$

We also have

$$\sup_{u \in [0,1]} |u_n^{(j)}(u)| \leq \sup_{u \in [0,1]} |\alpha_n^{(j)}(u)| + \frac{n \sup_{u \in [0,1]} |\hat{G}_n^{(j)}(u) - \hat{G}_n^{(j)}(u-)| - 1}{n}.$$

To understand this, we start by recognizing that the maximum of either  $\alpha_n^{(j)}(\cdot)$  or  $-\alpha_n^{(j)}(\cdot)$ , and consequently,  $|\alpha_n^{(j)}(\cdot)|$ , must occur at one of the discontinuities in  $\hat{G}_n^{(j)}$ . These discontinuities correspond to the values  $\{U_{i:n}^{(j)}, 1 \leq i \leq n\}$ , where  $U_{1:n}^{(j)} \leq \dots \leq U_{n:n}^{(j)}$  represent the order statistics. Hence, the quantity

$$n \times (u_n^{(j)}(i/n) + \alpha_n^{(j)}(U_{i:n}^{(j)})) \quad (22)$$

is equal to the highest count of  $U_i^{(j)}$  that are equal to  $U_{i:n}^{(j)}$  minus 1. Assuming there are no ties among  $U_1^{(j)}, \dots, U_n^{(j)}$  (which, for example, happens in the i.i.d. case), this expression equals 1. Consequently, we derive the classical identity for the uniform quantile process, as outlined in Csörgő 1983, Section 1.4 or Shorack and Wellner 2009, Chapter 3.

$$\sup_{0 \leq u \leq 1} |\alpha_n^{(j)}(u)| = \sup_{0 \leq u \leq 1} |u_n^{(j)}(u)|.$$

In the general case, Equation (22) is limited above by the maximum count of  $U_i^{(j)}$  which are equal minus 1. It is worth noting that this maximum count can be expressed as

$$n \times \sup_{u \in [0,1]} |\hat{G}_n^{(j)}(u) - \hat{G}_n^{(j)}(u-)|.$$

We have, following the proof of lemma 4.6 in Berghaus, Bücher, and Volgushev 2017

$$\begin{aligned} \sup_{u \in [0,1]} |\hat{G}_n^{(j)}(u) - \hat{G}_n^{(j)}(u-)| &\leq \sup_{u,v \in [0,1], |u-v| \leq 1/n} |\hat{G}_n^{(j)}(u) - \hat{G}_n^{(j)}(v)| \\ &\leq \sup_{u,v \in [0,1], |u-v| \leq 1/n} |\hat{G}_n^{(j)}(u) - \hat{G}_n^{(j)}(v) - (u-v)| + \frac{1}{n} \\ &\leq \frac{1}{\sqrt{n}} \sup_{u,v \in [0,1], |u-v| \leq 1/n} |\alpha_n^{(j)}(u) - \alpha_n^{(j)}(v)| + \frac{1}{n}. \end{aligned}$$

Using Condition  $\mathcal{B}$ , the above term is  $o_{\mathbb{P}}(n^{-1/2-\mu})$  for  $\mu \in (0, \theta_1)$  and  $\theta_1 \in [0, 1/2]$ . Additionally, using  $\|C_n^o - C\|_{\infty} = o_{\mathbb{P}}(1)$ , we obtain

$$\|u_n^{(j)}\|_{\infty} = o_{\mathbb{P}}(1).$$

Thus,  $\forall t \in (0, 1)$

$$\hat{\theta}_n(a) = \theta_t(a) + o_{\mathbb{P}}(1).$$

Since  $F$  is in the max-domain of attraction of  $H$ , we have

$$\theta(a) = \lim_{t \rightarrow 0} \theta_t(a).$$

Hence the result of the lemma.  $\square$

Since convergence in probability is reliably preserved under continuous transformations, we attain the outcomes outlined in Proposition 2.  $\square$

### Appendix E. Definition of the Adjusted Rand Index (ARI)

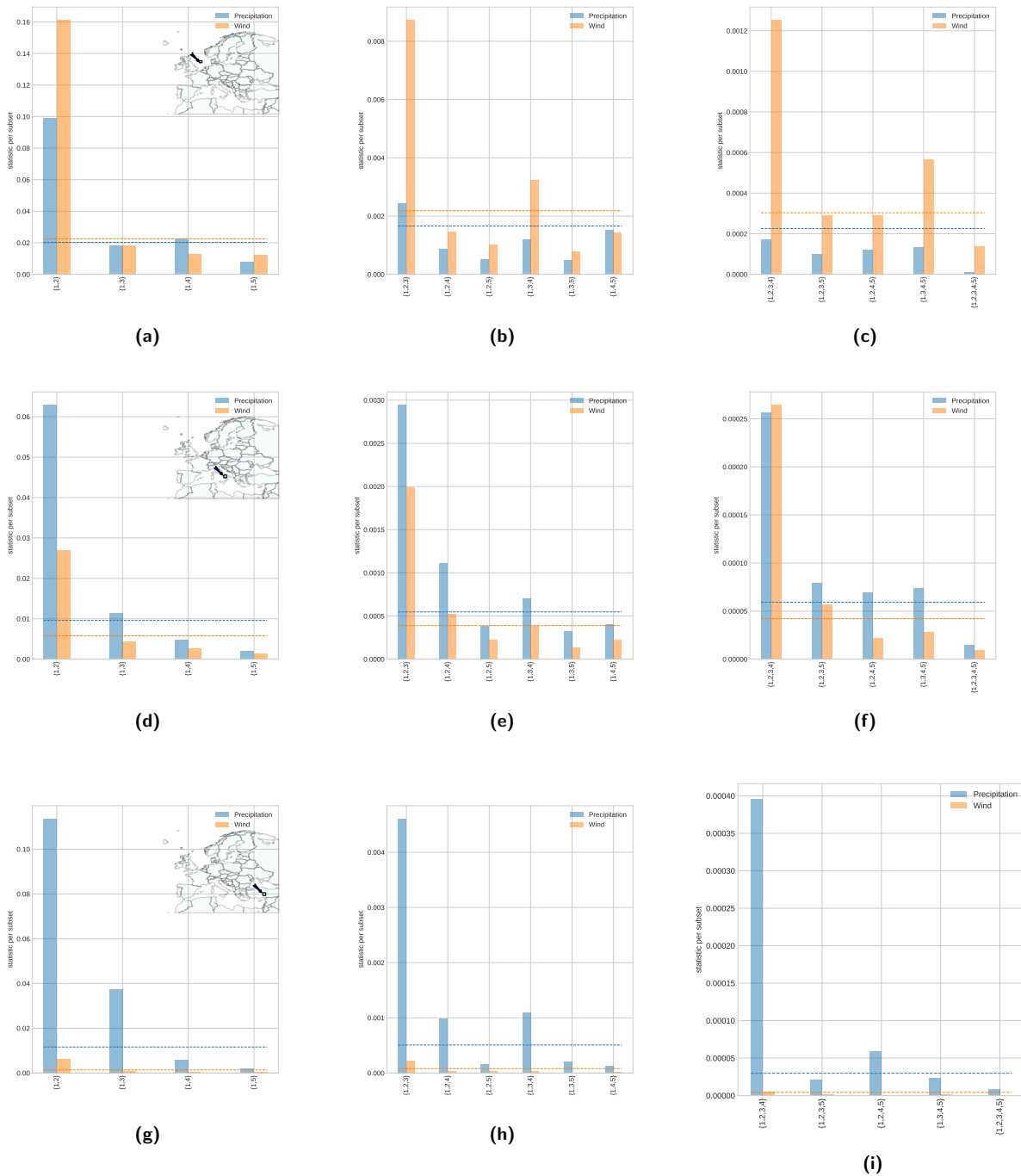
The ARI is computed as follows: Let  $O = \{O_g\}_{g=1,\dots,G}$  and  $S = \{S_h\}_{h=1,\dots,H}$  be two partitions with  $d$  entities, and let  $d_{gh}$  be the number of entities in cluster  $O_g$  in partition  $O$  and in cluster  $S_h$  in partition  $S$ . Denote by  $d_{g\cdot}$  (resp.  $d_{\cdot h}$ ) the number of entities in cluster  $O_g$  (resp.  $S_h$ ) in partition  $O$  (resp.  $S$ ). The ARI is evaluated using the following expressions:

$$r_0 = \sum_{g=1}^G \sum_{h=1}^H \binom{d_{gh}}{2}, \quad r_1 = \sum_{g=1}^G \binom{d_{g\cdot}}{2}, \quad r_2 = \sum_{h=1}^H \binom{d_{\cdot h}}{2}, \quad r_3 = \frac{2r_1 r_2}{d(d-1)},$$

$$\text{ARI}(O, S) = \frac{r_0 - r_3}{0.5(r_1 + r_2) - r_3},$$

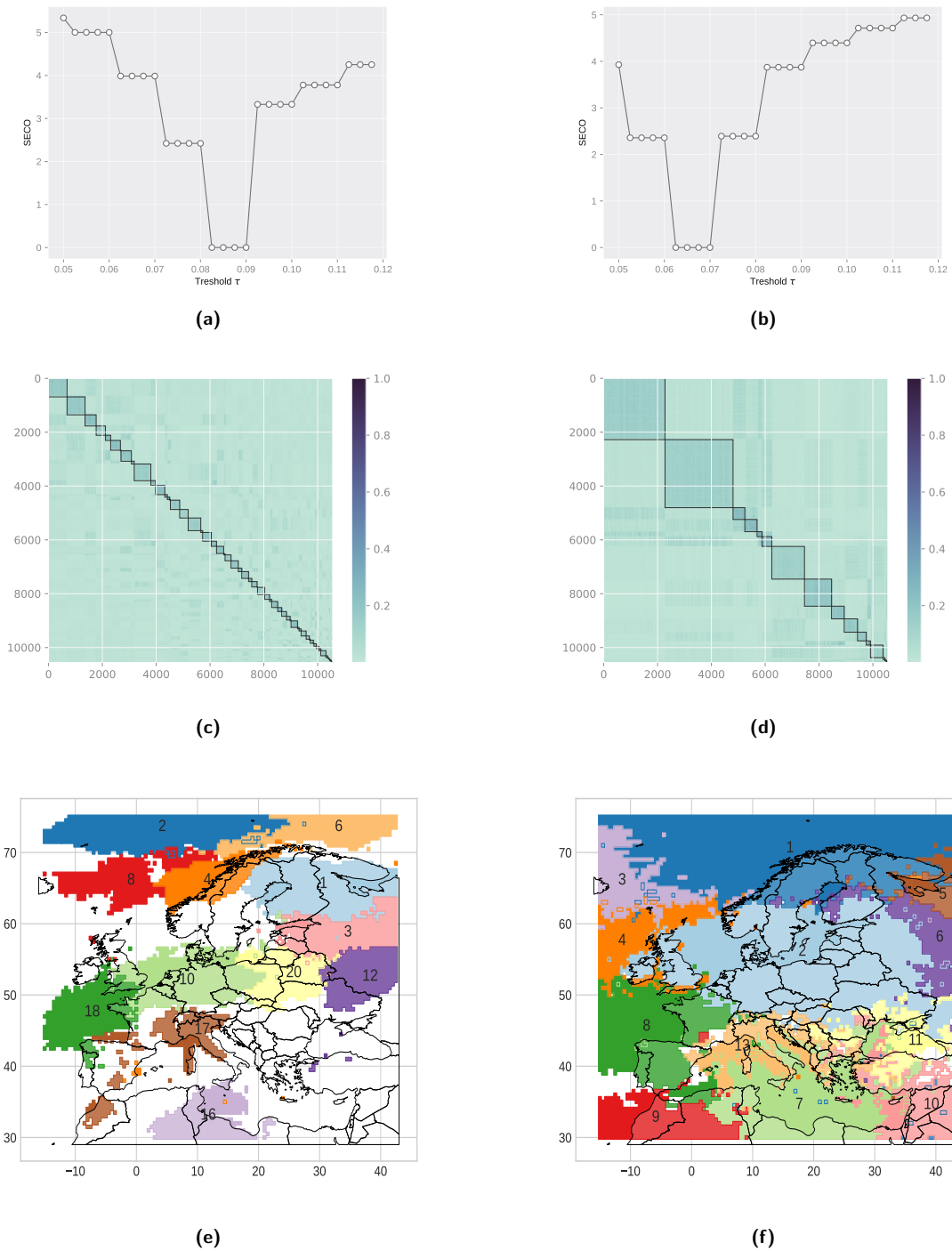
where  $\binom{n}{k}$  is the binomial coefficient.

### Appendix F. Supplementary Figures

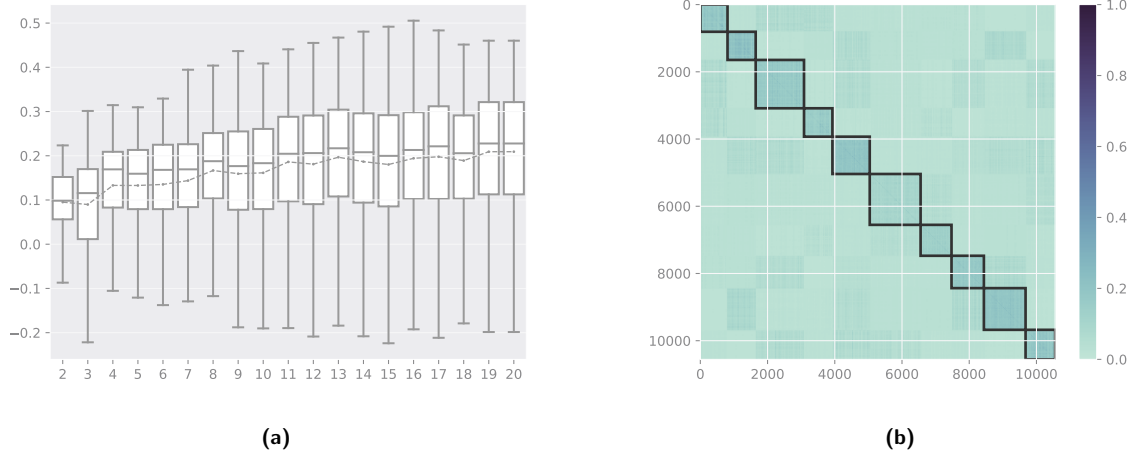


**Figure 11.** The dependogram provides a concise summary of randomness test results conducted on daily total precipitation and wind speed maxima in the ERA5 dataset. This study covers three European regions, examining 304 days of observations across nine distinct pixels. The first column displays a map with each row representing an area and the nine pixels marked by red squares. Test statistics are represented by bars, and critical values are depicted by dotted horizontal lines. The dependogram columns focus on pairwise, three-wise, and four-wise randomness tests with a lag of 4.

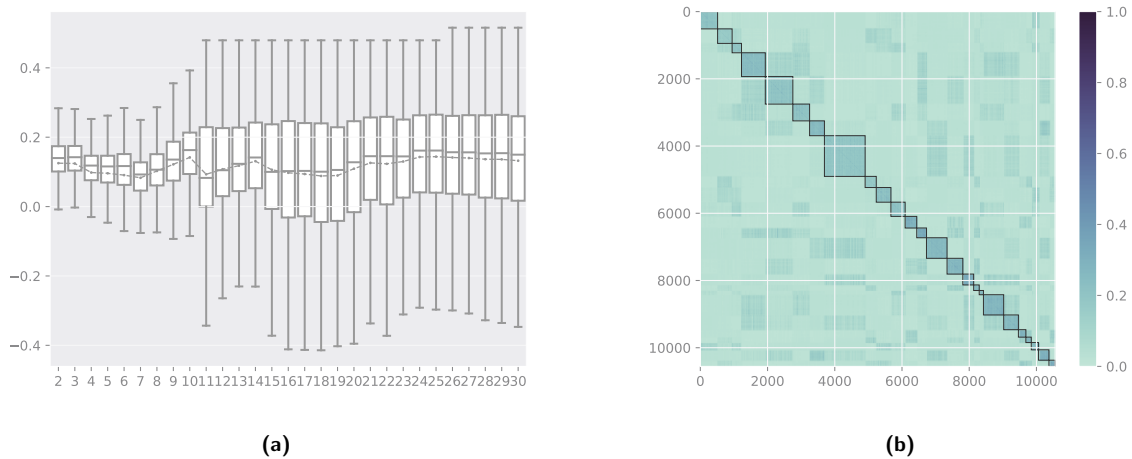




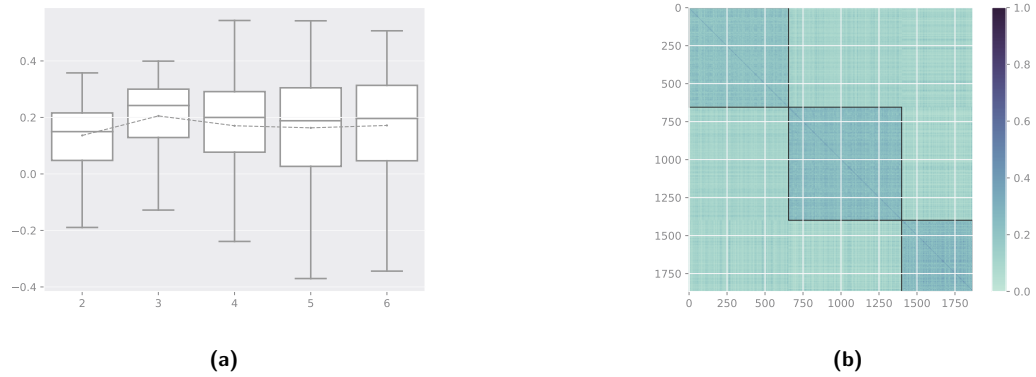
**Figure 12.** Value of the function  $L$  for different values of  $\tau \in \Delta = \{0.05, 0.0525, \dots, 0.12\}$  in Panels 12a (Precipitation) and 12b (Wind). Partitions of the extremal correlation similarity matrix with threshold  $\tau = 0.09$  for Panel 12c (Precipitation) and  $\tau = 0.07$  12d (Wind). Squares represent the clusters of variables. Representation of the 12 largest clusters (in decreasing order) of the partition of the extremal correlation matrix of total precipitation and wind speed maxima with threshold  $\tau = 0.09$  and  $\tau = 0.07$ , respectively in Panels 12e and 12f. Number of each cluster is depicted at the top-left corner of the corresponding panel.



**Figure 13.** Boxplots of the silhouette coefficients for different values of  $K$  using the quantization-based algorithm. Thick lines indicate the median, boxes the interquartile range and whiskers the full range of the distribution. Partitions of the SECO similarity matrix with  $K = 10$  for the quantization-based approach. Squares represent the clusters of variables.



**Figure 14.** Boxplots of the silhouette coefficients for different values of  $K$  using the hierarchical clustering algorithm. Thick lines indicate the median, boxes the interquartile range and whiskers the full range of the distribution. The average silhouette is depicted by the dotted line. Partitions of the SECO similarity matrix with  $K = 25$ . Squares represent the clusters of variables.



**Figure 15.** Boxplots of the silhouette coefficients for different values of K performed by a hierarchical clustering algorithm on the fourth partition given by Algorithm (CAICE). Thick lines indicate the median, boxes the interquartile range and whiskers the full range of the distribution. The average silhouette is depicted by the dotted line. The clustered matrix SECO is represent in Panel 15b where squares represent the clusters.