Clustering bivariate dependences in the extremes of climate variables

Edoardo Vignotto, Sebastian Engelke, Jakob Zscheischler
{edoardo.vignotto, sebastian.engelke}@unige.ch
zscheischler@climate.unibe.ch

UNIVERSITÉ DE GENÈVE

UNIVERSITÄT BERN

OESCHGER CENTRE CLIMATE CHANGE RESEARCH

Introduction

Abstract

Identifying hidden spatial patterns that define sub-regions characterized by a similar behaviour is a central topic in statistical climatology. This task, often called regionalization, is helpful for recognizing areas in which the variables under consideration have a similar stochastic distribution and thus, potentially, for reducing the dimensionality of the data. Many examples for regionalization are available, spanning from hydrology to weather and climate science. However, the majority of regionalization techniques focuses on the spatial clustering of a single variable of interest and are often not tailored to extremes. Climate extremes often have severe impacts, which can be amplified when co-occurring with extremes in other variables. Given the importance of characterizing climate extremes at the regional scale, here we develop an algorithm that identifies homogeneous spatial sub-regions that are characterized by a common bivariate dependence structure in the tails of a bivariate distribution. In particular, we use a novel non-parametric distance able to capture the similarities and differences in the tail behaviour of bivariate distributions as the core of our clustering procedure. We apply the approach to identify homogeneous regions with varying coherence in the co-occurrence of extremely low sea level pressure and heavy precipitation extremes in Great Britain and Ireland.

Multivariate extreme value theory

- **Asymptotic dependence**

$$\chi = \lim_{u \to 1} \mathbb{P}(F_1(X_1) > u | F_2(X_2) > u) > 0$$

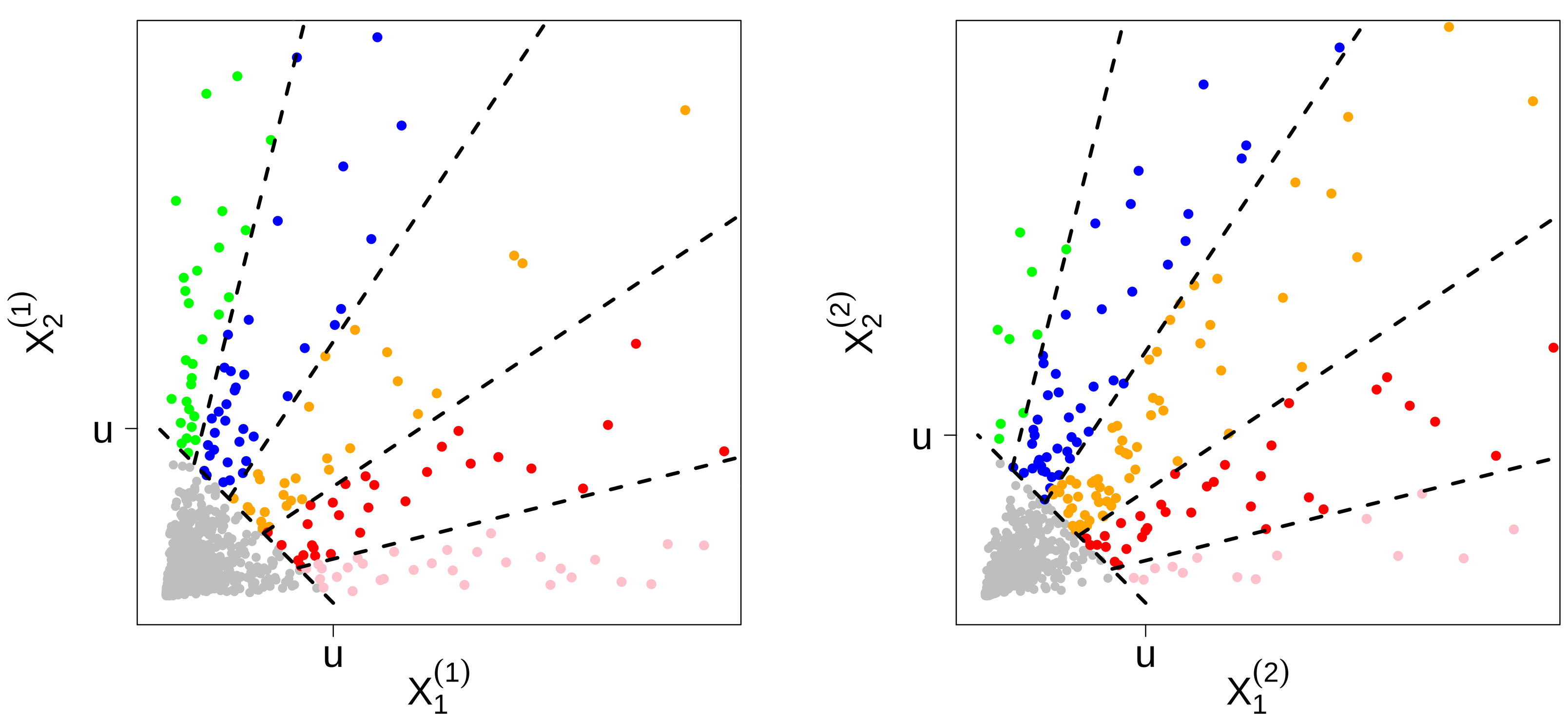
- **Asymptotic independence**

$$\overline{\chi} = \lim_{u \to \infty} \frac{2 \log \mathbb{P}(F_1(X_1) > u)}{\log \mathbb{P}(F_1(X_1) > u, F_2(X_2) > u)} - 1 < 1$$

- **Multivariate Pareto distribution**

$$\mathbb{P}(\boldsymbol{Y} \leq z) = \lim_{u \to \infty} \mathbb{P}(\boldsymbol{X}/u \leq z | \, \|\boldsymbol{X}\|_\infty > u) = \frac{\Lambda(z \wedge 1) - \Lambda(z)}{\Lambda(1)}$$

A distance measure between tail dependencies



$$\widehat{p}_w^{(z)} = \frac{\#\{X_i^{(z)} \in A_w^{(z)}\}}{\#\{r(X_i^{(z)}) > u\}}, \quad w = 1, \ldots, W$$

$$d_{12} = d(X^{(1)}, X^{(2)}) = \sum_{w=1}^{W} (\widehat{p}_w^{(1)} - p_w^{(2)}) \log(p_w^{(1)}/p_w^{(2)})$$

Partitioning Around Medoids (PAM) algorithm

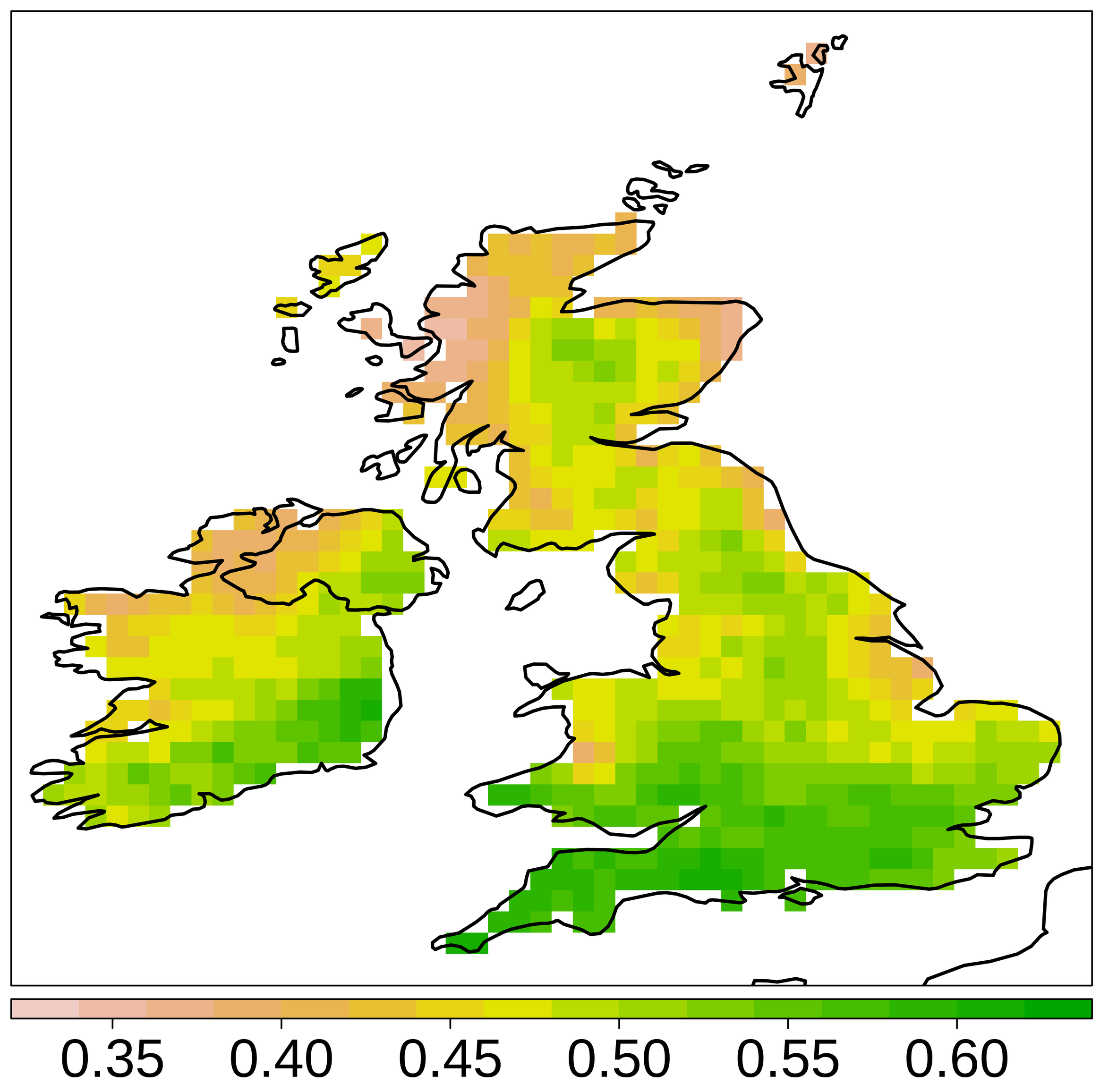
Fix a number K of clusters, and for each cluster C_k, $k = 1, \ldots, K$, randomly choose a point x_l as initial center, called medoid.

1. Form K clusters by assigning every point $x_1, \ldots, x_n$ to its closest (measured with the distances d_{ij}) medoid.
2. For each cluster $C_k = 1, \ldots, K$, find the new medoid $x_i \in C_k$ for which the total intracluster distance is minimized.
3. If at least one medoid has changed, then go back to point 1, otherwise end the algorithm.

Results

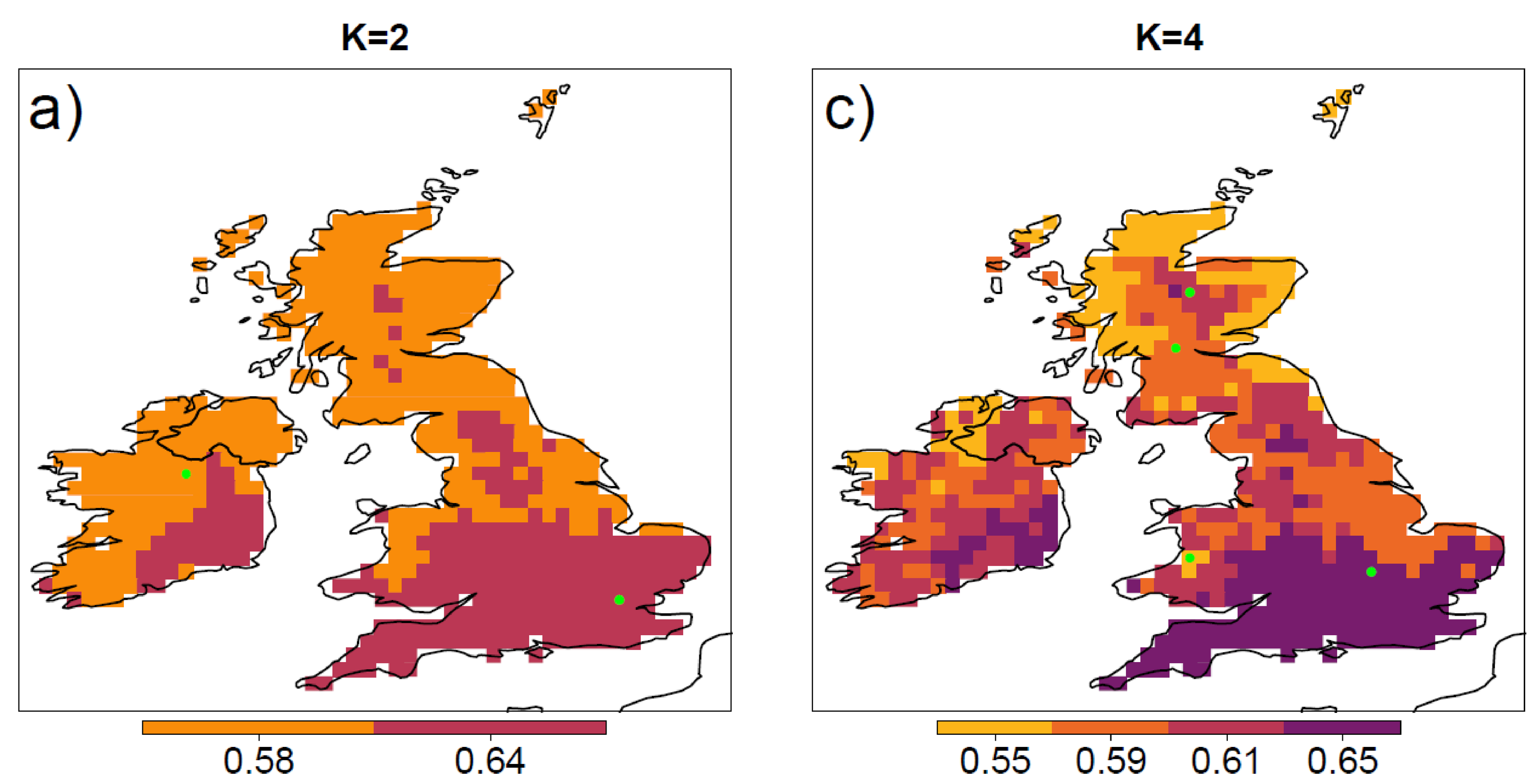
We analyze E-OBS average weekly temperature av over Great Britain and Ireland from 1950 to 2018. In particular, we focus on the tail dependence structure between temperature extremes at each location and temperature extremes in its surrounding.

Exploratory analysis



Grid-based tail dependence coefficient χ between weekly averages of precipitation and inverted sea level pressure in winter ($q = 0.9$).

Clustering results



Partitions obtained with different values of K. Locations that are medoids for the respective cluster are highlighted with green dots. Colors denote the averages of the estimated tail dependence coefficients χ ($q = 0.9$) of each grid point for each cluster.