Regionalization of Europe based on K-Mean Clustering Analysis of the climate change of Temperatures and Precipitation

M. J. Carvalho, P. M. Melo-Gonçalves, J. Teixeira and A. Rocha

CESAM – Centro de Estudos do Ambiente e do Mar
Physics Department University of Aveiro
Correspond to: mariajcarvalho@ua.pt

2014 EcoHCC Tomar–Portugal

CliM@UA® Grupo de Meteorologia e Climatologia da Universidade de Aveiro

CESAM
Centre for Environmental and Marine Studies
www.cesam.ua.pt

The problems of Regionalization

- **Geographic** → coherence between grid points not guaranteed;

- **Reference Period** → may be different in projected future;

Solution ?

Using the **differences** between projected future and the reference period.

Objective

Definition of regions of coherent climate change patterns in Europe

How ?

Methodology

1) Determination of the daily climatology for each grid point (for each of the variables under study);

3) Difference between reference period and long-term future climatologies;

2) Number of clusters - *k*? (Mathematical determination + sensitivity to *k*)

4) K-means cluster analysis → Each grid point is assigned to 1 cluster.

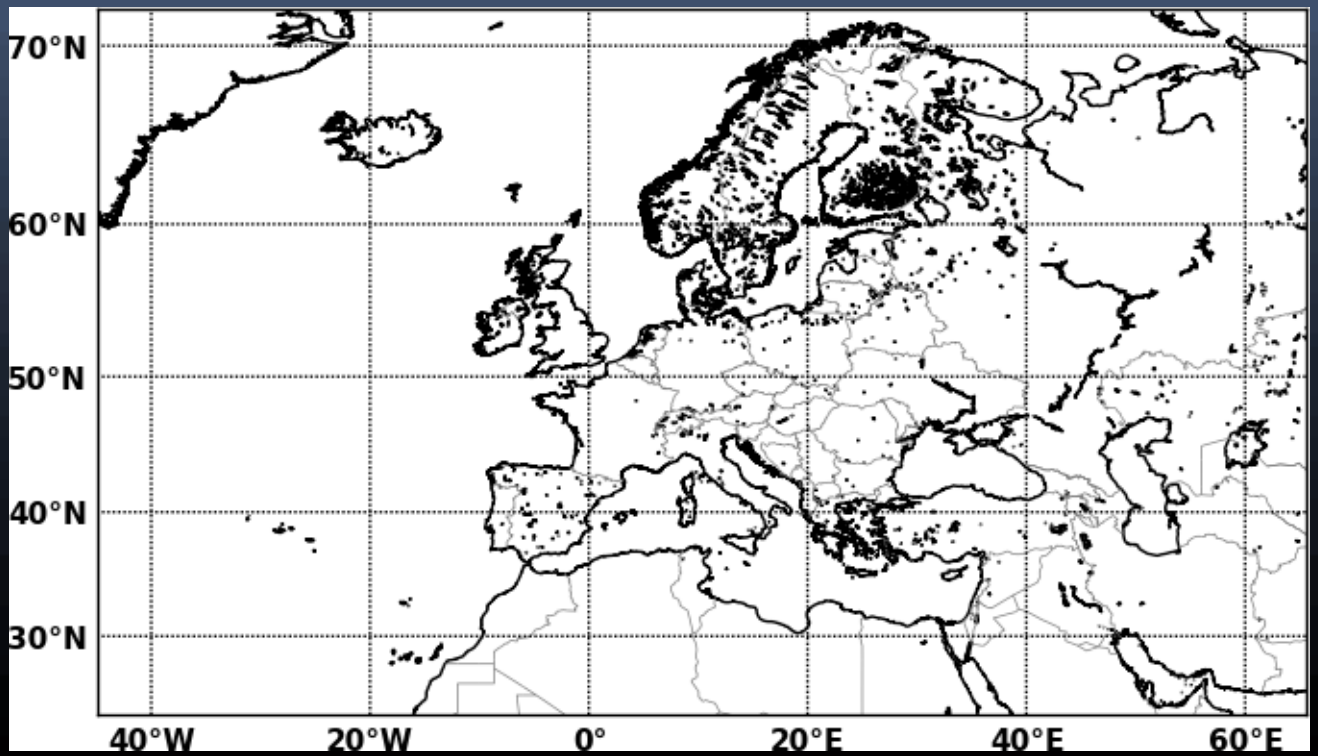
- Univariate
- Multivariate

(Sensitivity to the number of clusters)

Data

Daily data from **MPI-ESM-LR** r1i1p1 (CMIP5 project) simulations with 1.9º horizontal resolution for:

- Recent-past: 1986 - 2005
- RCP8.5 Long-term future: 2081 – 2100



Variables:

- tasmax
- tasmin
- pr

Why K-means?

- Non-hierarchical method;

  (vectors can be reassigned)

- Minimizes the variance between cluster members, maximizing variance between clusters

Determination of *k* in K-means:

- Gap Statistic $\rightarrow$ $k = 6$

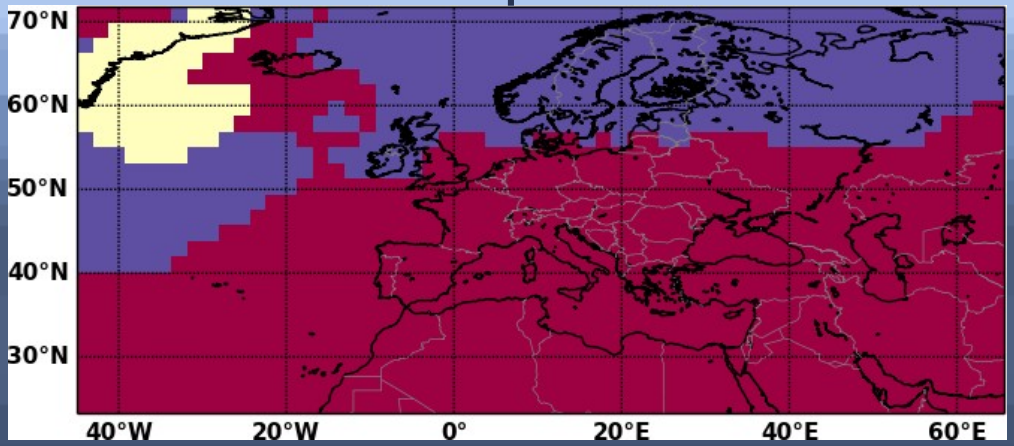
- $k = 3$
- $k = 10$
- $k = 13$

} Evaluation of the validity of the mathematically determined *k*.

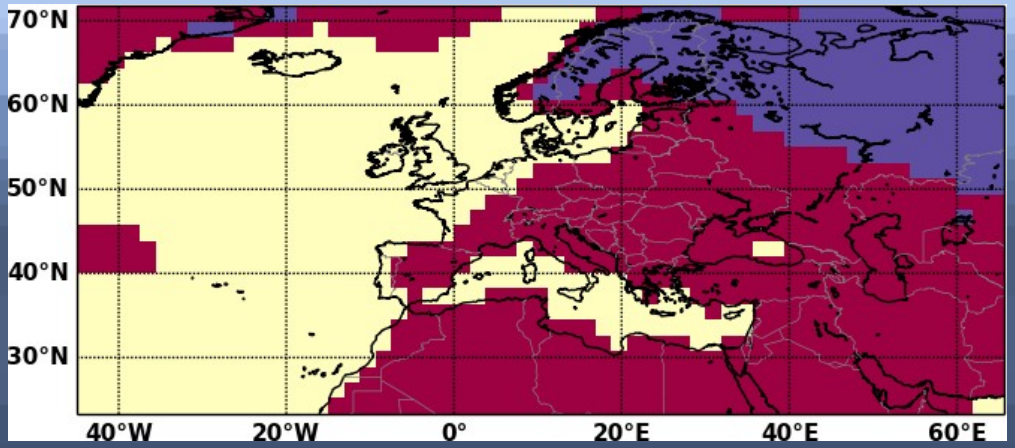
Results

K = 3

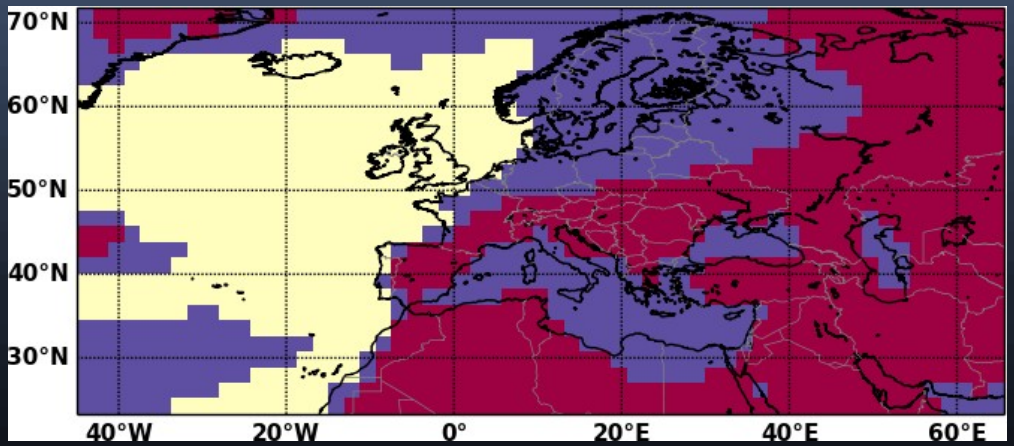
pr



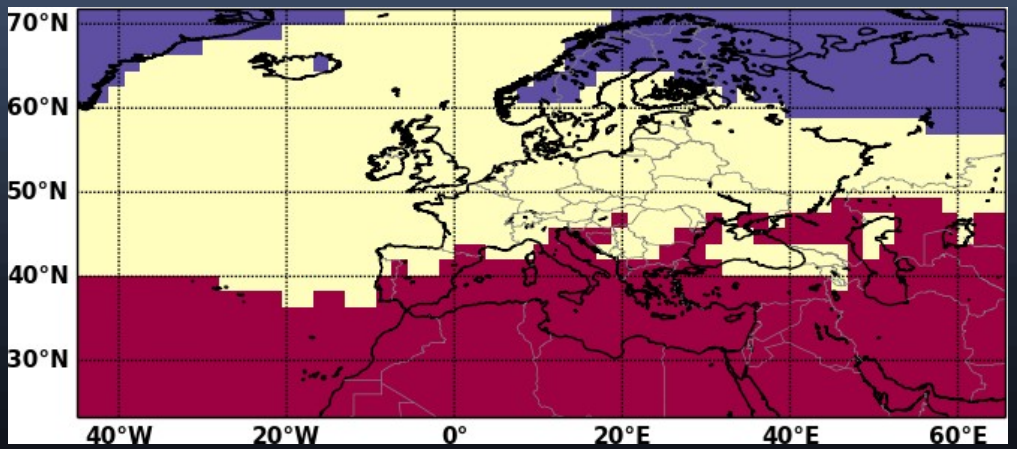
tasmin



tasmax



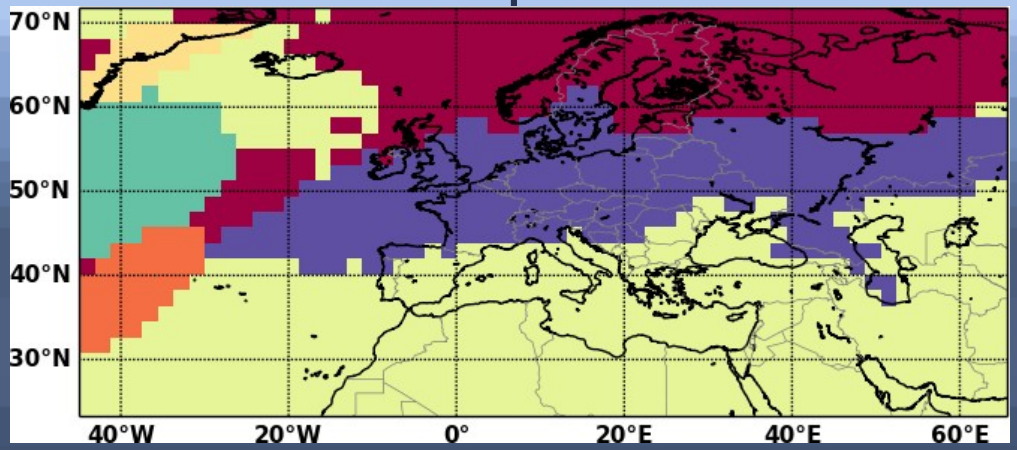
3-var



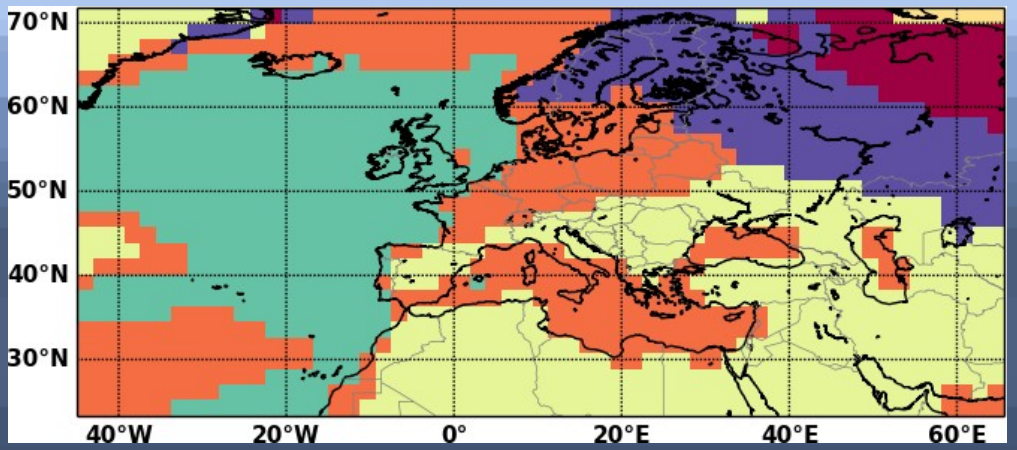
Results

K = 6

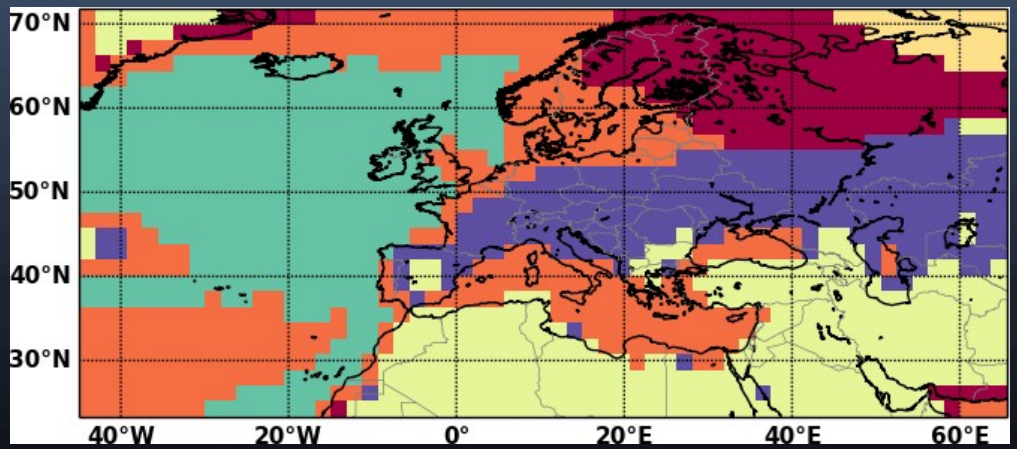
pr



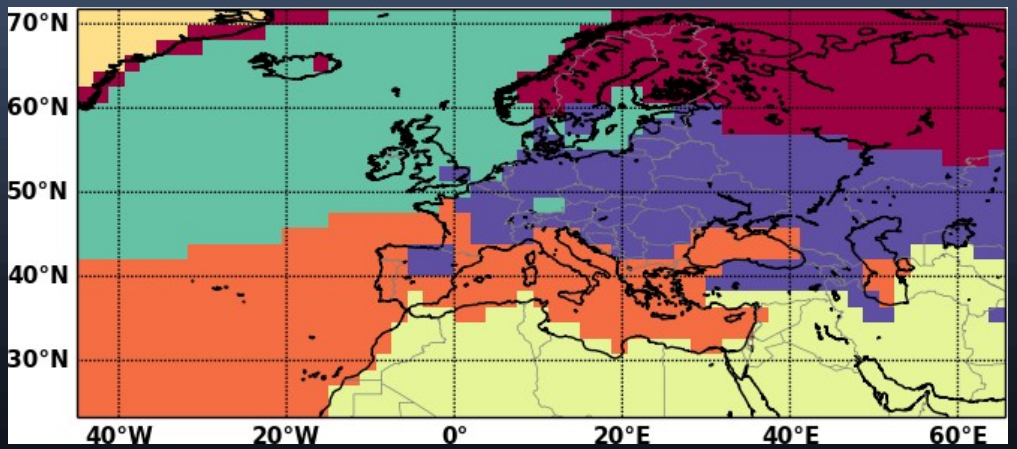
tasmin



tasmax



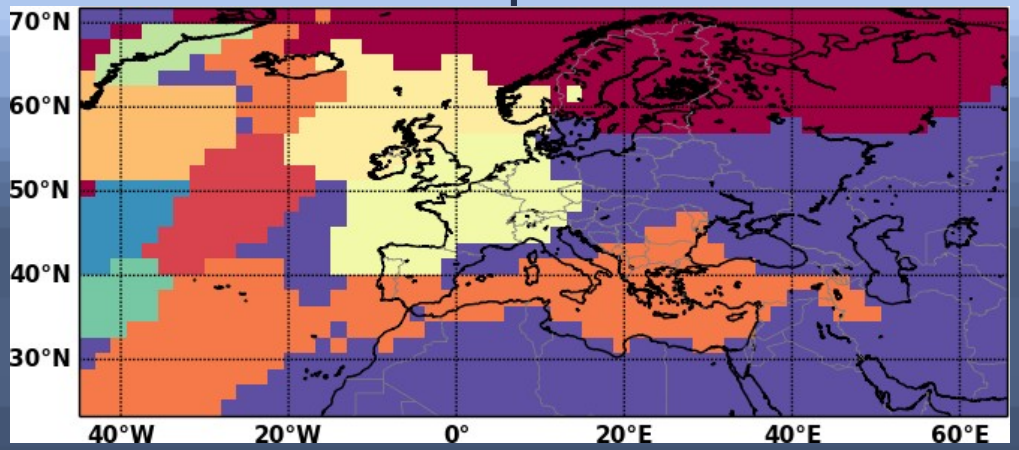
3-var



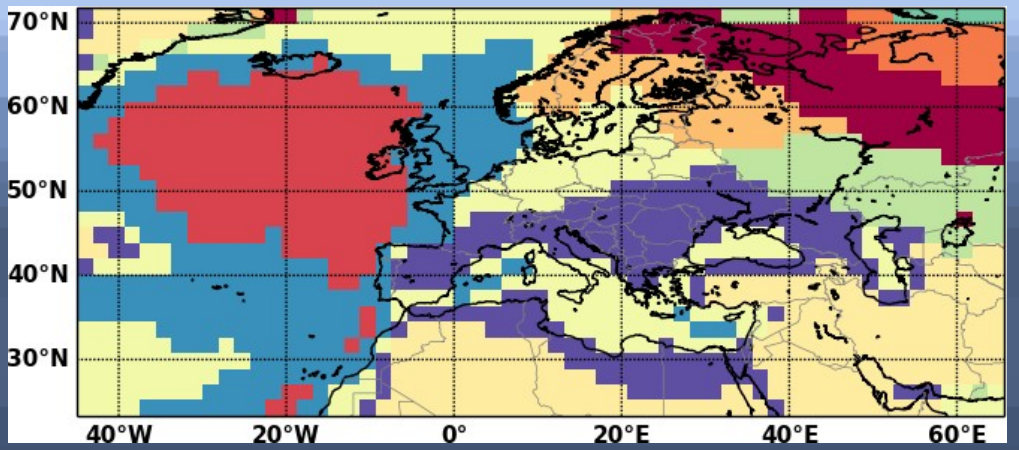
Results

K = 10

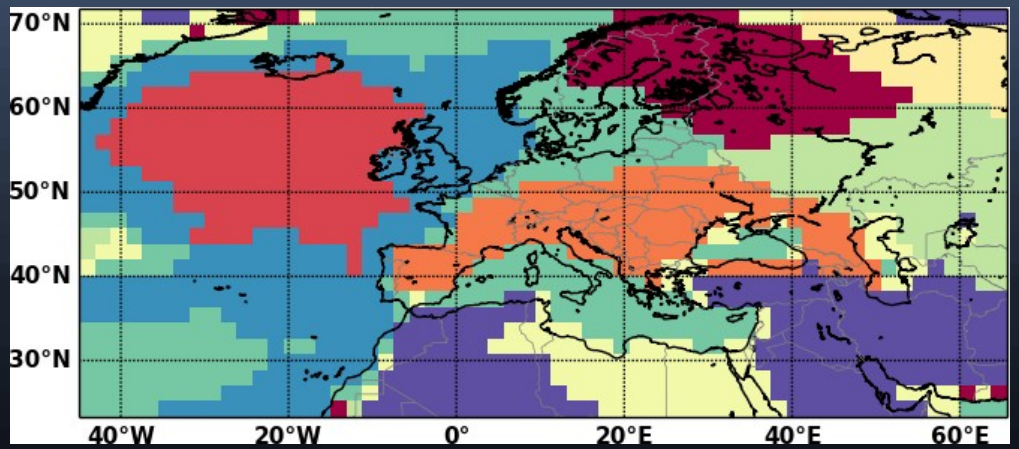
pr



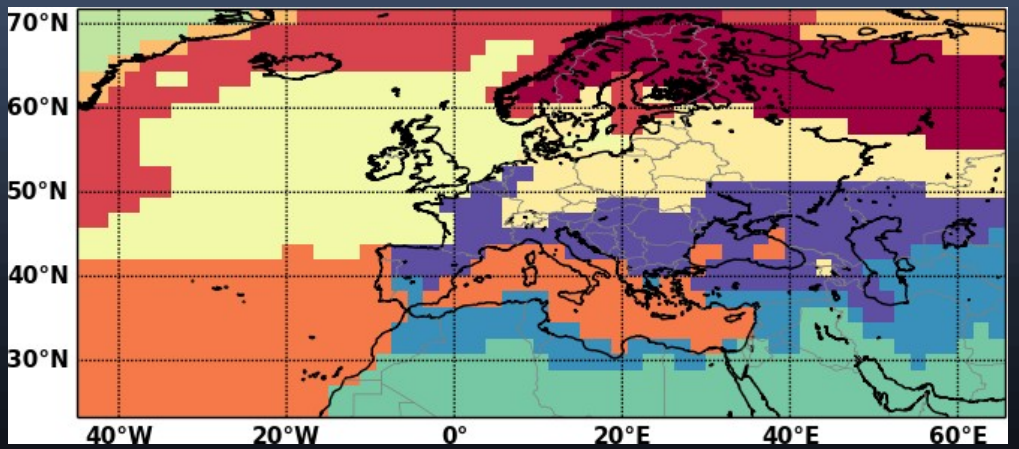
tasmin



tasmax



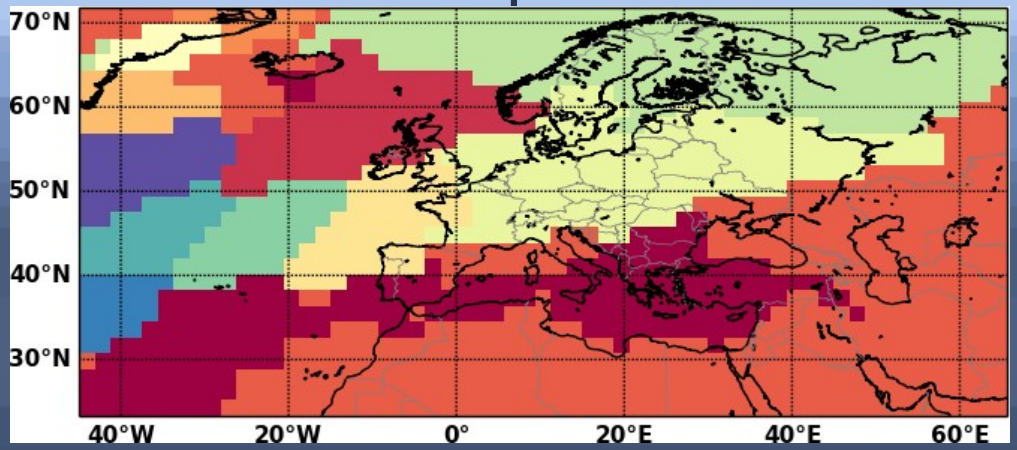
3-var



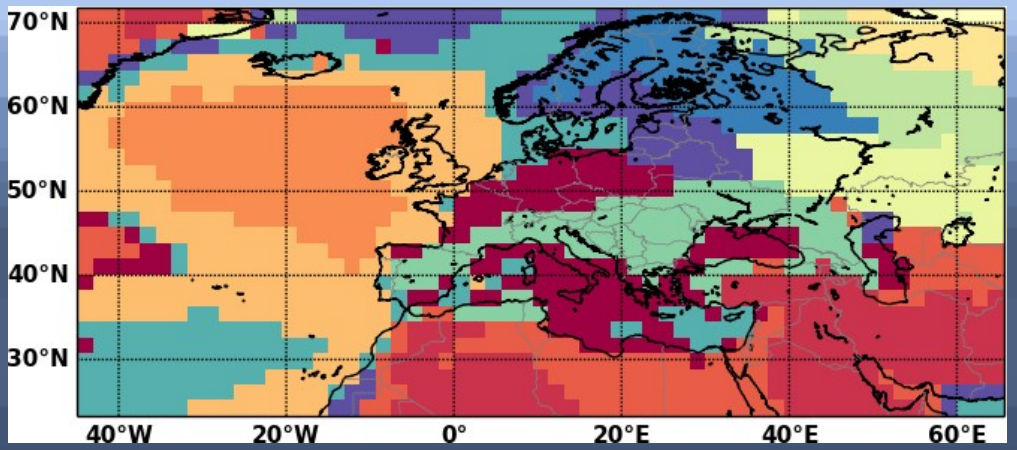
Results

K = 13

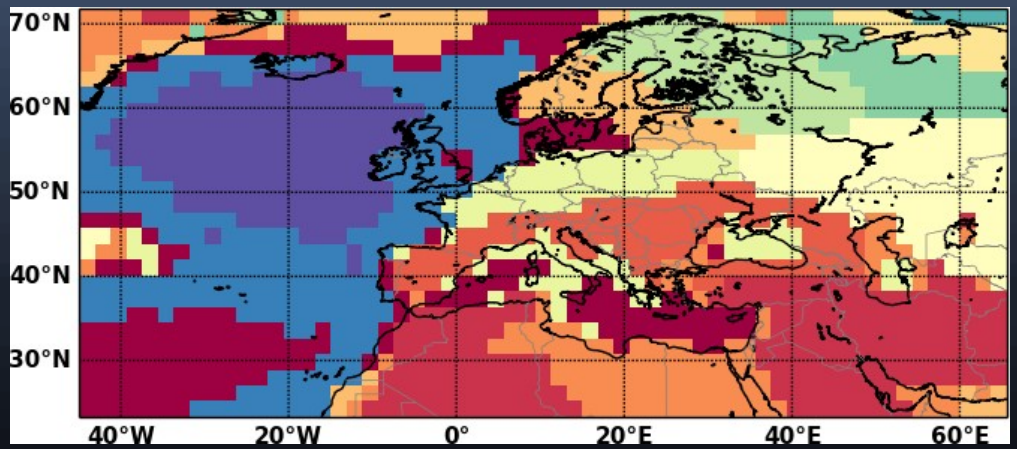
pr



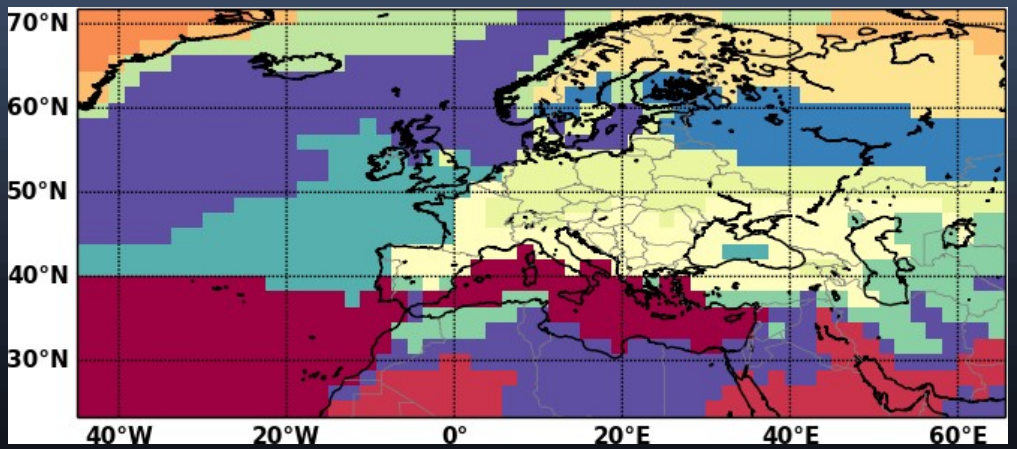
tasmin



tasmax

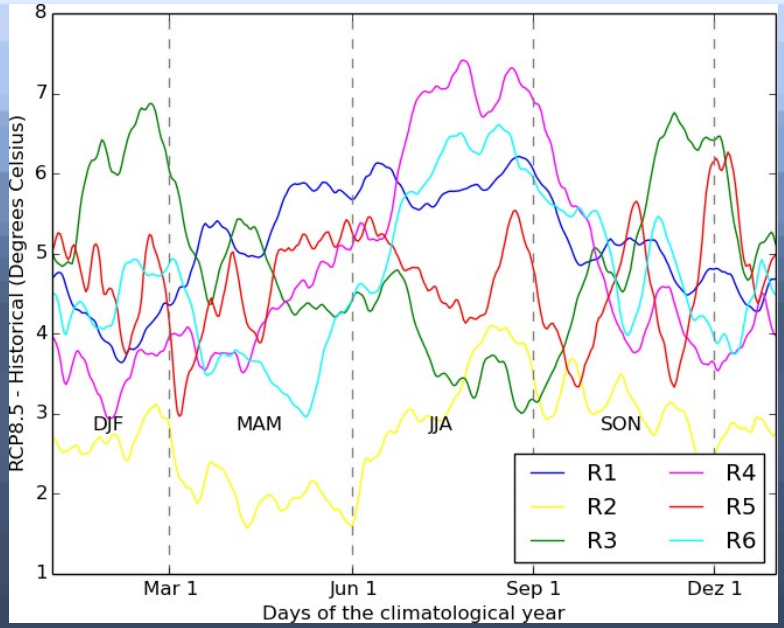


3-var

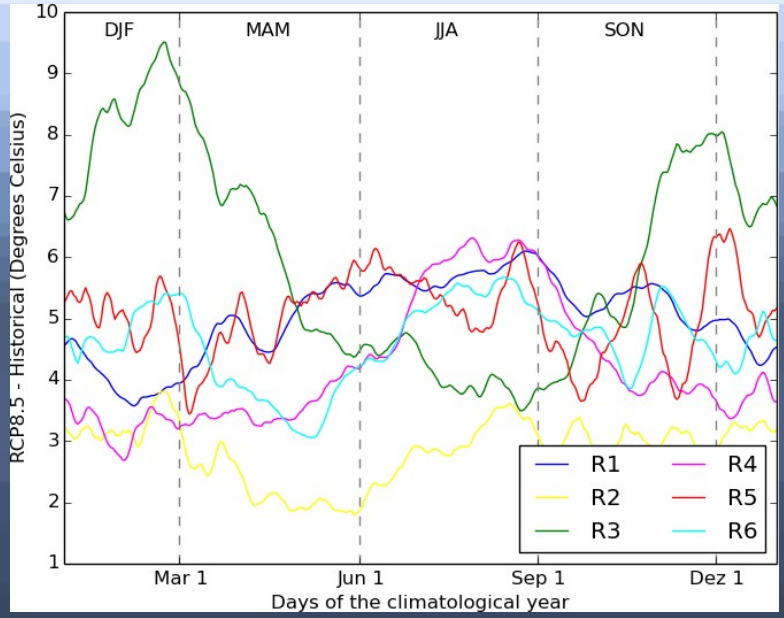


Validation of the *k* regions

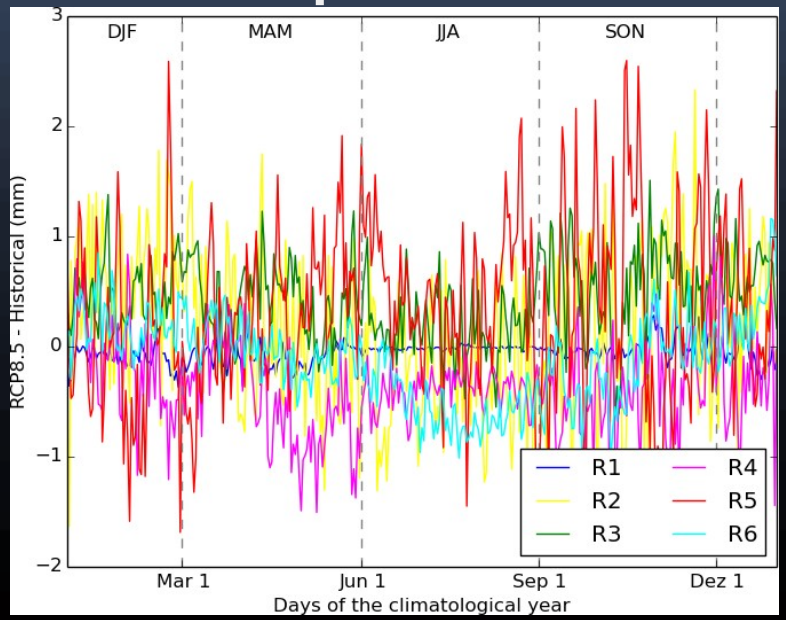
tasmax



tasmin



pr



Mean Climatology difference for all of each of the 6 region's grid points

Concluding Remarks

- Mathematical approach is, on a first look an effective way of determining *k*;
  - *k* = 3 → large variability within the clusters
  - *K* = 6 → optimal (for the used resolution)
  - *k* = 10/13 → new clusters are sometimes "*cell-thin*" and consequently not significant

(...)

Concluding Remarks

(...)

- Univariate K-means results vary for each variable, which was expectable specially for pr;
- Multivariate K-means analysis is consistent with the univariate versions;
- Daily climatology differences for each cluster are mostly outside the minimum-maximum range in-cluster differences.

Further work ...

- Sensitivity of *k* to horizontal resolution;
- Using defined regions for the climate change study in the regime of extreme events;
- Using other significant atmospheric variables such as wind intensity and direction as well as mean sea level pressure;
- Using an ensemble of the CMIP5 models instead of a single model approach.

This study was supported by FEDER funds through the Programa Operacional Factores de Competitividade – COMPETE and by Portuguese national funds through FCT – Fundaçao para a Ciência e a Tecnologia, within the framework of Projects:

RESORT Project Reference PTDC/CTE-ATM/111508/2009;

CLIPE Project Reference PTDC/ AAC-CLI/111733/2009;

Thank you for your time!