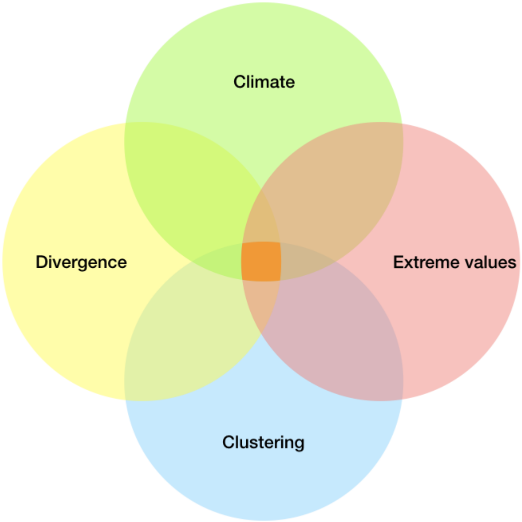
A new dissimilarity for extreme rainfall clustering, non-parametric and coupling bivariate extreme value theory and marginals

Margaux Zaffran, Philippe Naveau

Laboratoire des Sciences du Climat et de l'Environnement (ESTIMR)

JdS 2021 - 08/06/2021


Climate
Divergence
Extreme values
Clustering

Context

Dataset: daily France rainfall in mm ($1\text{mm} = 1\text{L}/\text{m}^2$), from 1976 to 2015, at 174 Meteo-France stations.

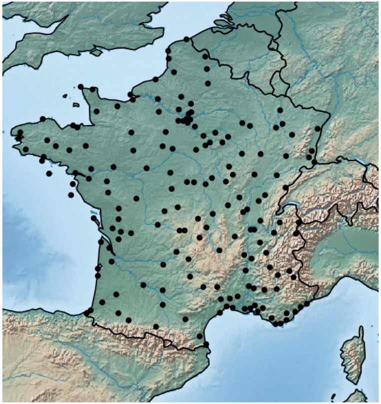


Figure: Weather stations

Objectives

Aim:

- Construct coherent groups of locations according to the nature of the extreme's rainfall/wind/temperature there (e.g. spatial group)

Objectives

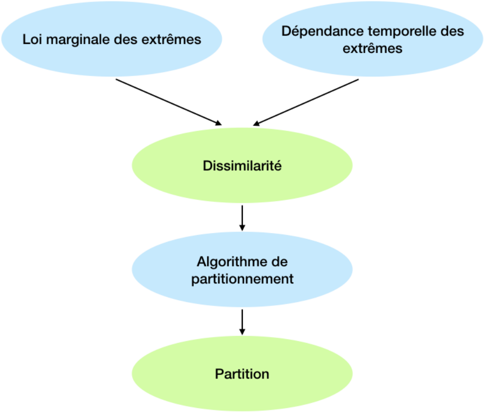
Aim:

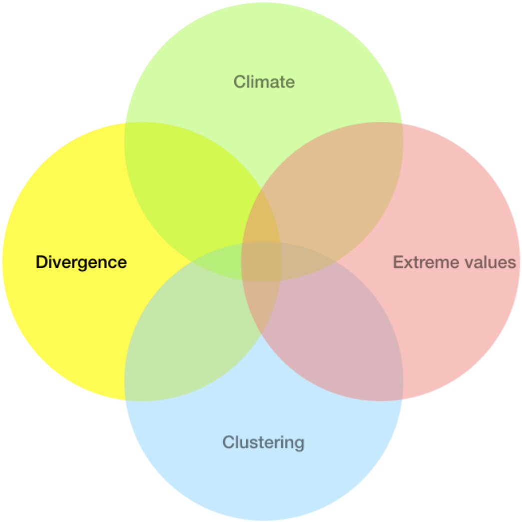
- Construct coherent groups of locations according to the nature of the extreme's rainfall/wind/temperature there (e.g. spatial group)

Constraints:

- Non-parametric approach (no fit, few assumptions)
- Good scaling to large datasets

Theoretical tools


Loi marginale des extrêmes
Dépendance temporelle des extrêmes
Dissimilarité
Algorithme de partitionnement
Partition


Climate
Divergence
Extreme values
Clustering

Divergence

Mathematical tool to quantify the similarity between two probability distributions.

Divergence

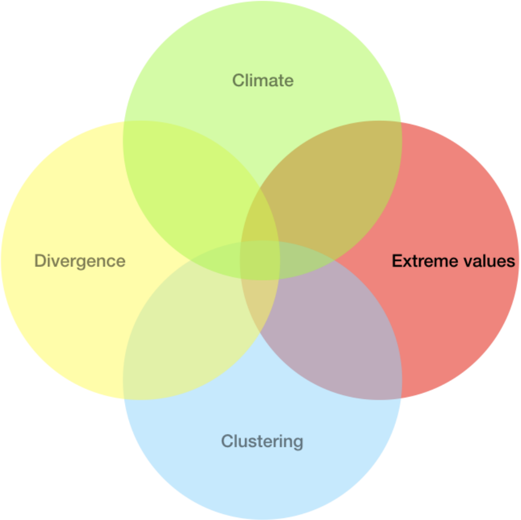
Mathematical tool to quantify the similarity between two probability distributions.

Kullback-Leibler directed divergence is asymmetric. Symmetrized quantity:

Definition (Kullback-Leibler divergence)

Let f and g be two probability density functions. Then:

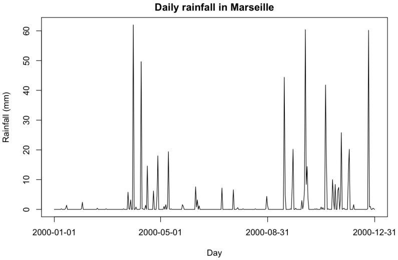
$$\begin{aligned} D(f,g) &= J(f;g) + J(g;f) \\ &= \mathbb{E}_f\left[\log\left(\frac{f(X)}{g(X)}\right)\right] + \mathbb{E}_g\left[\log\left(\frac{g(Y)}{f(Y)}\right)\right] \end{aligned}$$


Climate
Divergence
Extreme values
Clustering

Extreme

Excesses over a threshold.

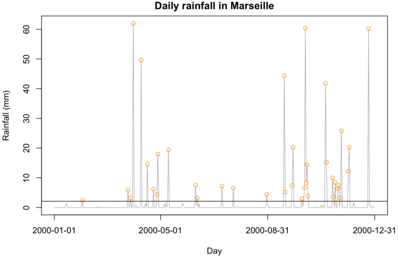
Variable of interest: $X_u = [X|X > u]$, for some well-chosen and high threshold u.



Extreme

Excesses over a threshold.

Variable of interest: $X_u = [X|X > u]$, for some well-chosen and high threshold u.



Extreme

Excesses over a threshold.

Variable of interest: $X_u = [X|X > u]$, for some well-chosen and high threshold u.

Let X r.v. with density f (and tail $\overline{F}$). Then, for any threshold u, has the following characteristics:

- density $f_u(x) = \frac{f(x)}{\overline{F}(u)} \mathbb{1}_{\{x>u\}}$
- tail $\overline{F}_u(x) = \frac{\overline{F}(x)}{\overline{F}(u)} \mathbb{1}_{\{x>u\}}$

Kullback-Leibler divergence tailored for large excesses

KL adapted for univariate extremes, in terms of excesses, in [Naveau et al., 2014]. We can consider the quantity:

$$D(f_u, g_u) = J(f_u; g_u) + J(g_u; f_u)$$

with $J(f_u; g_u) = \mathbb{E}_{f_u}\left[\log\left(\frac{f_u(X_u)}{g_u(X_u)}\right)\right]$.

Kullback-Leibler divergence tailored for large excesses

KL adapted for univariate extremes, in terms of excesses, in [Naveau et al., 2014]. We can consider the quantity:

$$D(f_u, g_u) = J(f_u; g_u) + J(g_u; f_u)$$

with $J(f_u; g_u) = \mathbb{E}_{f_u}\left[\log\left(\frac{f_u(X_u)}{g_u(X_u)}\right)\right]$.

Proposition (from [Naveau et al., 2014])

$D(f_u, g_u)$ is equivalent (under assumptions), as $u \to \tau$, to:

$$K(f_u, g_u) = -L(f_u; g_u) - L(g_u; f_u)$$

with: $L(f_u; g_u) = \mathbb{E}_f\left[\log\left(\frac{\overline{G}(X)}{\overline{G}(u)}\right)|X > u\right] - 1.$

Kullback-Leibler divergence tailored for large excesses

KL adapted for univariate extremes, in terms of excesses, in [Naveau et al., 2014]. We can consider the quantity:

$$D(f_u, g_u) = J(f_u; g_u) + J(g_u; f_u)$$

with $J(f_u; g_u) = \mathbb{E}_{f_u}\left[\log\left(\frac{f_u(X_u)}{g_u(X_u)}\right)\right]$.

Proposition (from [Naveau et al., 2014])

$D(f_u, g_u)$ is equivalent (under assumptions), as $u \to \tau$, to:

$$K(f_u, g_u) = -L(f_u; g_u) - L(g_u; f_u)$$

with: $L(f_u; g_u) = \mathbb{E}_f\left[\log\left(\frac{\overline{G}(X)}{\overline{G}(u)}\right) | X > u\right] - 1.$

→ easy construction of a plug-in estimator

Bivariate extreme dependence

Tail dependency

Take into account the tail dependence.

Definition (residual tail dependence coefficient)

$$\frac{\mathbb{P}\{A\}\mathbb{P}\{B\}}{\mathbb{P}\{A \cap B\}}$$

Tail dependency

Take into account the tail dependence.

Definition (residual tail dependence coefficient)

Let X, Y, two r.v. of c.d.f F and G respectively.

$$\frac{\mathbb{P}\{X > F^{-1}(q)\}\mathbb{P}\{Y > G^{-1}(q)\}}{\mathbb{P}\{X > F^{-1}(q), Y > G^{-1}(q)\}}$$

Tail dependency

Take into account the tail dependence.

Definition (residual tail dependence coefficient)

Let X, Y, two r.v. of c.d.f F and G respectively.

$$\bar{\chi} = \lim_{q \to 1} \frac{\log\left(\mathbb{P}\{X > F^{-1}(q)\}\mathbb{P}\{Y > G^{-1}(q)\}\right)}{\log\left(\mathbb{P}\{X > F^{-1}(q), Y > G^{-1}(q)\}\right)} - 1$$

Tail dependency

Take into account the tail dependence.

Definition (residual tail dependence coefficient)

Let X, Y, two r.v. of c.d.f F and G respectively.

$$\bar{\chi} = \lim_{q \to 1} \frac{\log\left(\mathbb{P}\{X > F^{-1}(q)\}\mathbb{P}\{Y > G^{-1}(q)\}\right)}{\log\left(\mathbb{P}\{X > F^{-1}(q), Y > G^{-1}(q)\}\right)} - 1$$

$\bar{\chi} \in [-1, 1]$

Dissimilarity

Dissimilarity

$$D(i,j) = \lambda \widehat{KL}_{i,j} + (1-\lambda)\left(1 - \widehat{\bar{\chi}}_{i,j}\right), \text{ with } \lambda \in [0,1].$$

Dissimilarity

$$D(i,j) = \lambda \widehat{KL}_{i,j} + (1-\lambda)\left(1 - \widehat{\bar{\chi}}_{i,j}\right), \text{ with } \lambda \in [0,1].$$

Some comments:

- we use $1 - \widehat{\bar{\chi}}$ and not $\widehat{\bar{\chi}}$
  $\Rightarrow$ 2 stations are close if they are dependent ;

Dissimilarity

$$D(i,j) = \lambda \widehat{KL}_{i,j} + (1-\lambda)\left(1 - \widehat{\bar{\chi}}_{i,j}\right), \text{ with } \lambda \in [0,1].$$

Some comments:

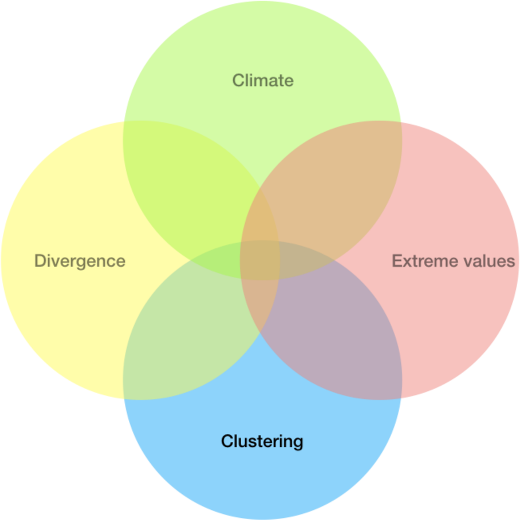
- we use $1 - \widehat{\bar{\chi}}$ and not $\widehat{\bar{\chi}}$
  $\Rightarrow$ 2 stations are close if they are dependent ;
- $\lambda = 1 \Leftrightarrow$ marginal law ;

Dissimilarity

$$D(i,j) = \lambda \widehat{KL}_{i,j} + (1-\lambda)\left(1 - \widehat{\bar{\chi}}_{i,j}\right), \text{ with } \lambda \in [0,1].$$

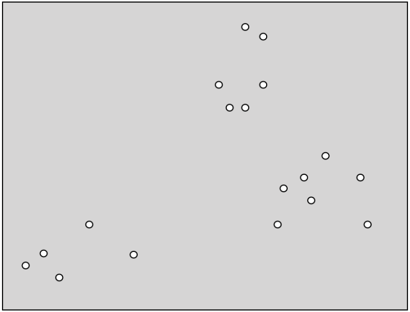
Some comments:

- we use $1 - \widehat{\bar{\chi}}$ and not $\widehat{\bar{\chi}}$
  $\Rightarrow$ 2 stations are close if they are dependent ;
- $\lambda = 1 \Leftrightarrow$ marginal law ;
- $\lambda = 0 \Leftrightarrow$ dependence structure.


Climate
Divergence
Extreme values
Clustering

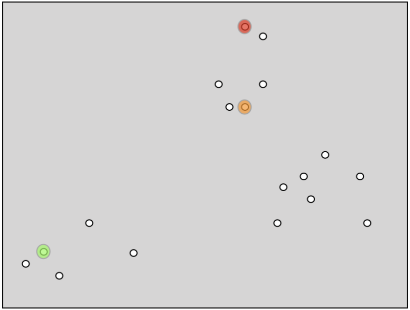
Clustering

We consider a simple and classic algorithm: `Partitioning Around Medoids (PAM)`, also called `k-medoids`, proposed by [Kaufman and Rousseeuw, 1987].



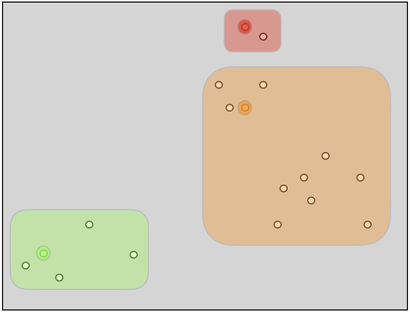
Clustering

We consider a simple and classic algorithm: `Partitioning Around Medoids (PAM)`, also called `k-medoids`, proposed by [Kaufman and Rousseeuw, 1987].



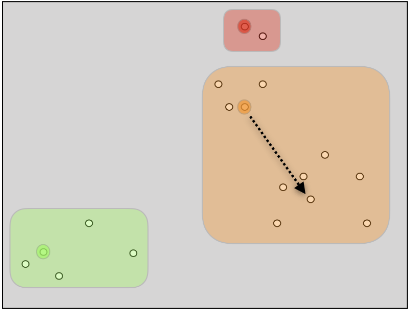
Clustering

We consider a simple and classic algorithm: `Partitioning Around Medoids (PAM)`, also called `k-medoids`, proposed by [Kaufman and Rousseeuw, 1987].



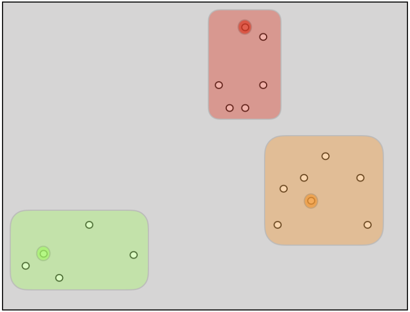
Clustering

We consider a simple and classic algorithm: `Partitioning Around Medoids (PAM)`, also called `k-medoids`, proposed by [Kaufman and Rousseeuw, 1987].



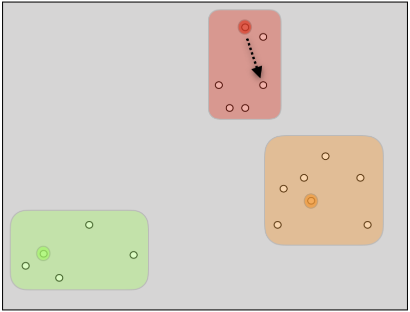
Clustering

We consider a simple and classic algorithm: `Partitioning Around Medoids (PAM)`, also called `k-medoids`, proposed by [Kaufman and Rousseeuw, 1987].



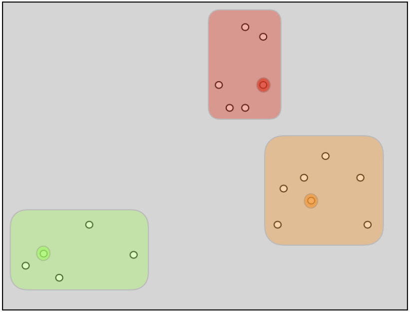
Clustering

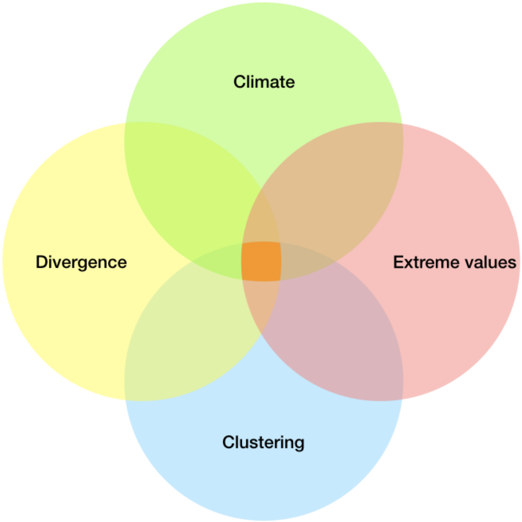
We consider a simple and classic algorithm: `Partitioning Around Medoids (PAM)`, also called `k-medoids`, proposed by [Kaufman and Rousseeuw, 1987].



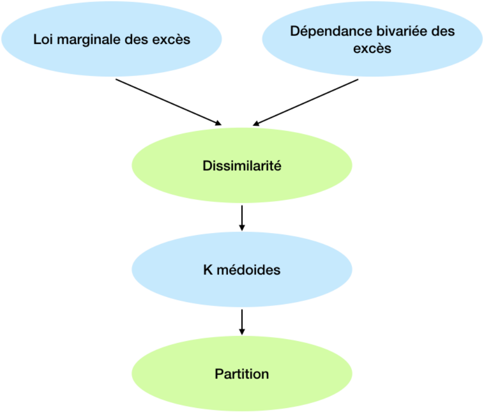
Clustering

We consider a simple and classic algorithm: `Partitioning Around Medoids (PAM)`, also called `k-medoids`, proposed by [Kaufman and Rousseeuw, 1987].




Climate
Divergence
Extreme values
Clustering

Algorithm


Loi marginale des excès
Dépendance bivariée des excès
Dissimilarité
K médoides
Partition

Back to rainfall

$q = 0.9, \lambda = 0.5$

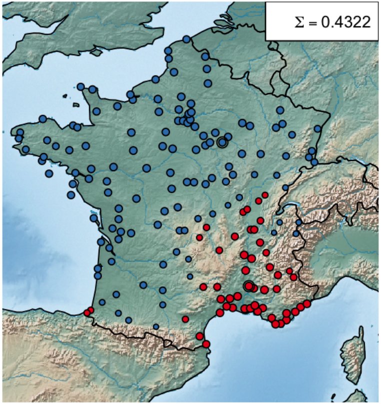


Figure: 2 clusters

$q = 0.9, \lambda = 0.5$

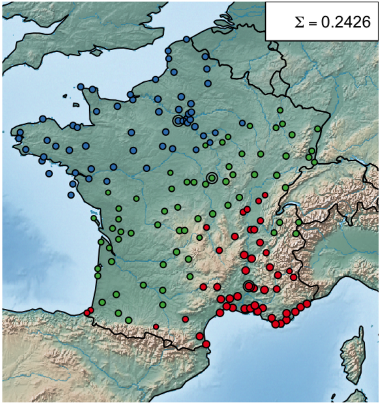


Figure: 3 clusters

$q = 0.9, \lambda = 0.5$

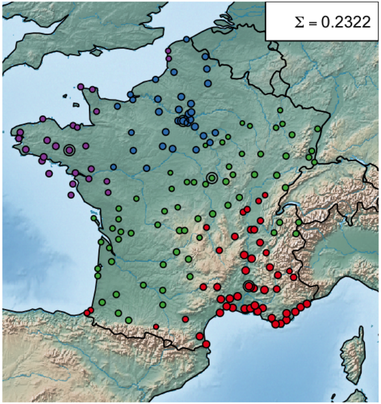


Figure: 4 clusters

Changing points: 3 clusters, $q = 0.9$

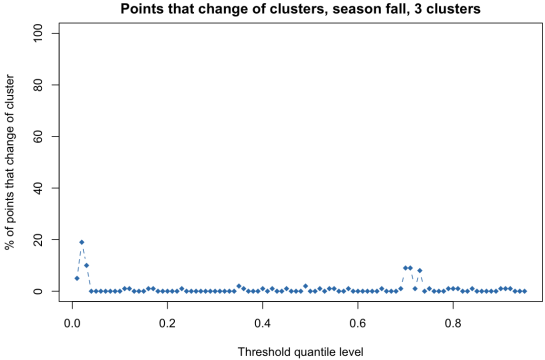


Figure: $\lambda = 0.5$

Constant points: $\lambda = 0.5$, $q_{\min} = 0.74$

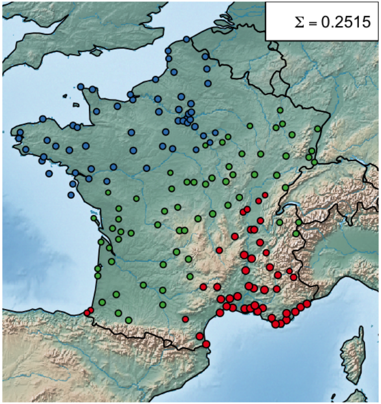


Figure: 3 clusters

Conclusion

Conclusion

- Climatologically coherent clusters
- Tool to help choosing the threshold level
- Implementation in a R package

Conclusion

- Climatologically coherent clusters
- Tool to help choosing the threshold level
- Implementation in a R package

Perspectives:

- Analyze whether the clusters are going to evolve in the climate change framework, using results of simulations
- Choice and cost of the clustering algorithm
- Statistical estimator study

Conclusion

Thanks for your attention!

Bernard, E., Naveau, P., Vrac, M., and Mestre, O. (2013).
Clustering of Maxima: Spatial Dependencies among Heavy Rainfall in France.
Journal of Climate, 26(20):7929–7937.

Kaufman, L. and Rousseeuw, P. J. (1987).
Clustering by means of medoids.

Naveau, P., Guillou, A., and Rietsch, T. (2014).
A non-parametric entropy-based approach to detect changes in climate extremes.
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 76(5):861–884.

Rousseeuw, P. J. (1987).
Silhouettes: A graphical aid to the interpretation and validation of cluster analysis.
Journal of Computational and Applied Mathematics, 20:53 – 65.

Tail dependency, gaussian case

Example (Bivariate Gaussian distribution)

Let $X = (X_1, X_2)$ bivariate gaussian distributed random variable, of correlation parameter ρ.
Then $\bar{\chi}(X_1, X_2) = \rho$

Tail dependency, gaussian case

Example (Bivariate Gaussian distribution)

Let $X = (X_1, X_2)$ bivariate gaussian distributed random variable, of correlation parameter ρ.
Then $\bar{\chi}\,(X_1, X_2) = \rho$

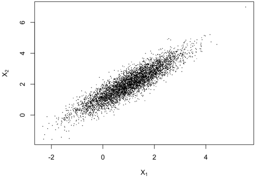


Figure: $\rho = 0.9$

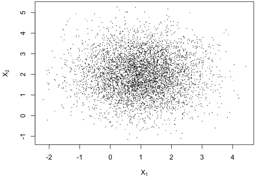


Figure: $\rho = 0$

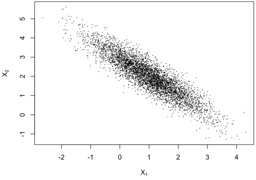
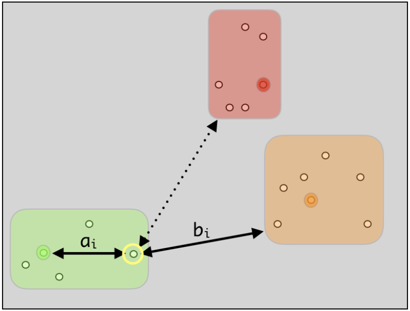


Figure: $\rho = -0.9$

Clustering: evaluation



Clustering: evaluation

Definition (Silhouette coefficient, [Rousseeuw, 1987])

Let's consider a point i.

- a_i the average dissimilarity of point i with the other points of its cluster;
- b_i the minimal average dissimilarity of point i to any other cluster.

We can now define the silhouette coefficient of i by:

$$s_i = \begin{cases} 1 - \frac{a_i}{b_i} & \text{if } a_i < b_i \text{ and } |C_i| > 1 \\ 0 & \text{if } a_i = b_i \text{ or } |C_i| = 1 \\ \frac{b_i}{a_i} - 1 & \text{if } a_i > b_i \text{ and } |C_i| > 1 \end{cases}$$

Changing points: method

Algorithm 1 Detect changing points between C_1 and C_2

Require: 2 clusterings of n points, C_1 and C_2

Initialize empty list: `moving_points`
Build A_1 and A_2, adjacency matrices of C_1 and C_2
$D = A_1 - A_2$
$S_i = \sum_{j=1}^{n} \mathbb{1}_{D_{i,j} \in \{-1,1\}}$
while $S \neq 0 \in \mathbb{R}^n$ **do**
    $u = \underset{i \in [\![1,n]\!]}{\operatorname{argmax}} \{S_i\}$
    Append u to `moving_points`
    $D_{u,\cdot} = 0$ and $D_{\cdot,u} = 0$
    Recompute S
end while
return `moving_points`

$q = 0.9, \lambda = 1$

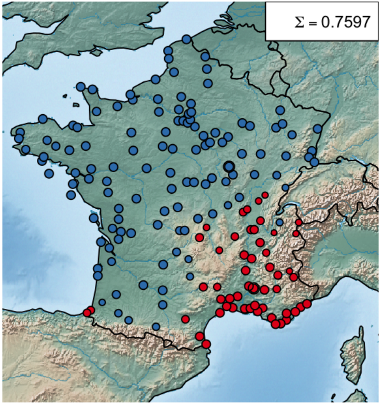


Figure: 2 clusters

$q = 0.9, \lambda = 1$

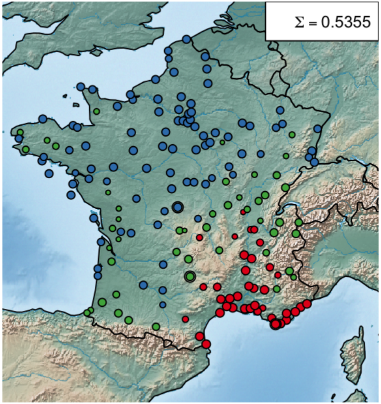


Figure: 3 clusters

$q = 0.9, \lambda = 1$

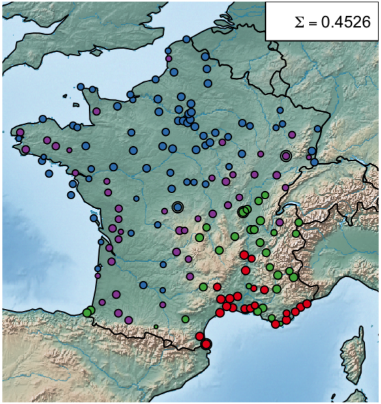


Figure: 4 clusters

$q = 0.9, \lambda = 1$

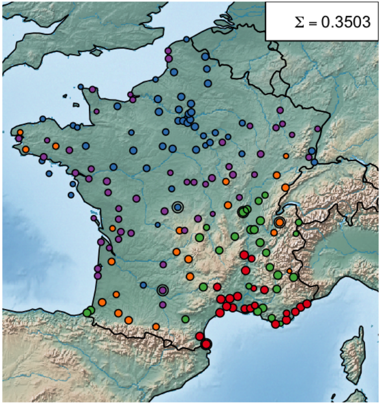


Figure: 5 clusters

q = 0.9, $\lambda = 1$

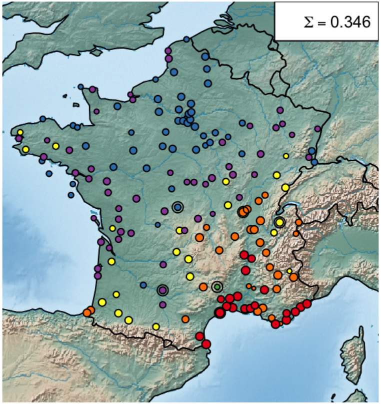


Figure: 6 clusters

$q = 0.9, \lambda = 1$

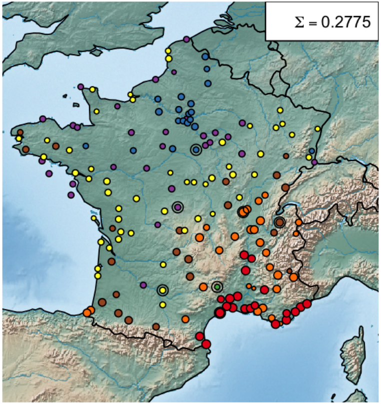


Figure: 7 clusters

q = 0.9, $\lambda = 1$

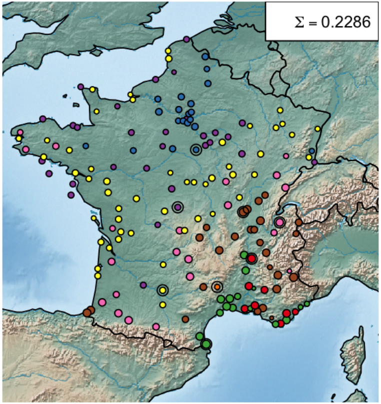


Figure: 8 clusters

$q = 0.9$, $\lambda = 1$

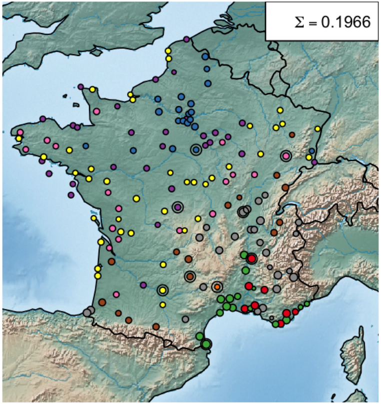


Figure: 9 clusters

$q = 0.9, \lambda = 0.75$

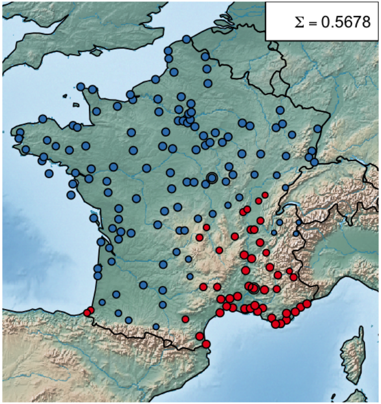


Figure: 2 clusters

$q = 0.9, \lambda = 0.75$

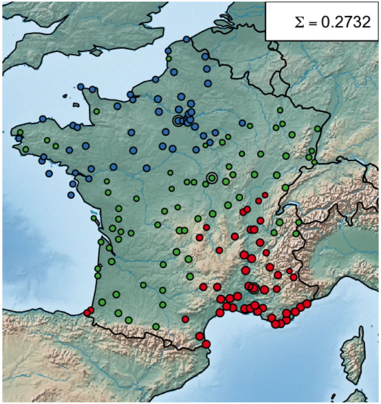


Figure: 3 clusters

$q = 0.9, \lambda = 0.75$

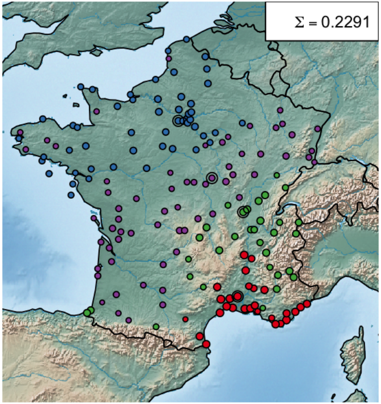


Figure: 4 clusters

q = 0.9, $\lambda = 0.75$

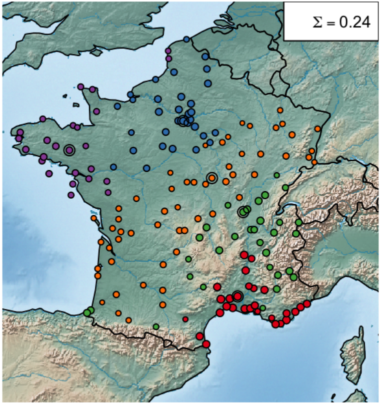


Figure: 5 clusters

$q = 0.9,\ \lambda = 0.75$

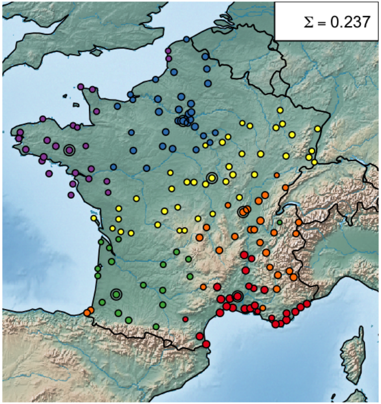


Figure: 6 clusters

$q = 0.9, \lambda = 0.75$

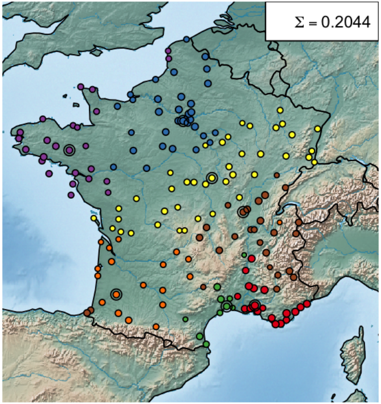


Figure: 7 clusters

$q = 0.9, \lambda = 0.75$

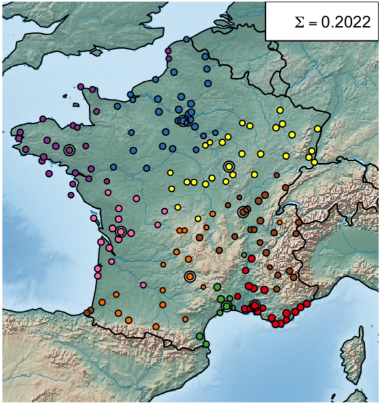


Figure: 8 clusters

$q = 0.9, \lambda = 0.75$

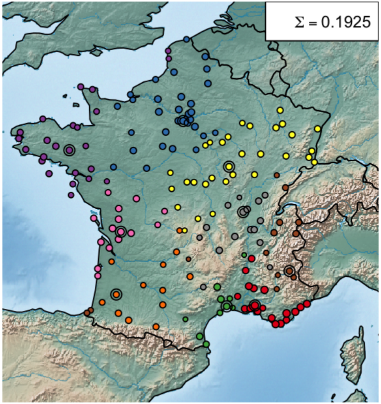


Figure: 9 clusters

q = 0.9, λ = 0.5

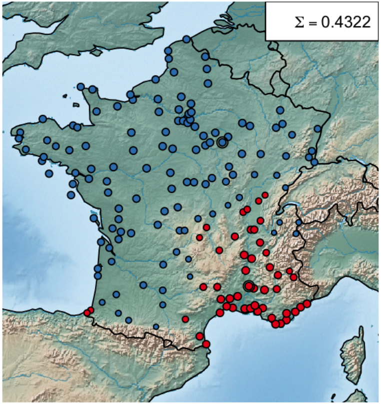


Figure: 2 clusters

q = 0.9, $\lambda = 0.5$

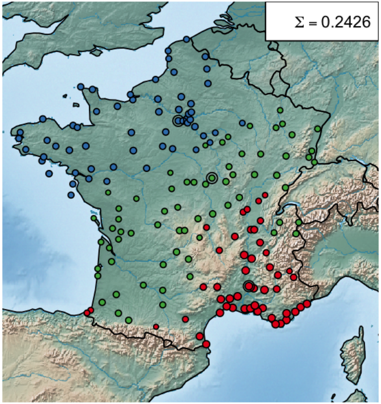


Figure: 3 clusters

$q = 0.9, \lambda = 0.5$

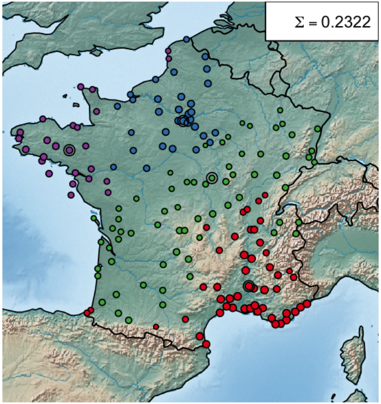


Figure: 4 clusters

$q = 0.9, \lambda = 0.5$

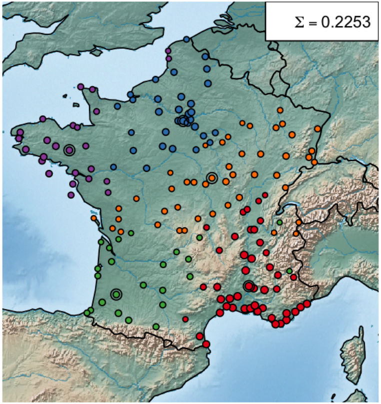


Figure: 5 clusters

$q = 0.9, \lambda = 0.5$

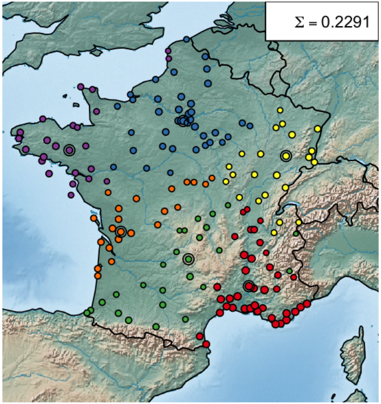


Figure: 6 clusters

$q = 0.9, \lambda = 0.5$

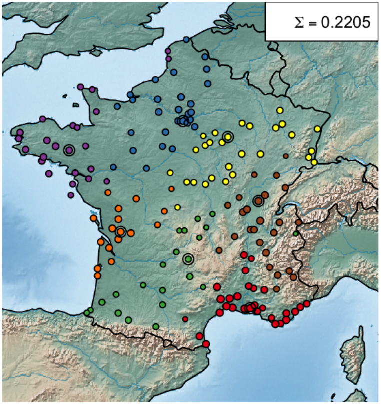


Figure: 7 clusters

$q = 0.9, \lambda = 0.5$

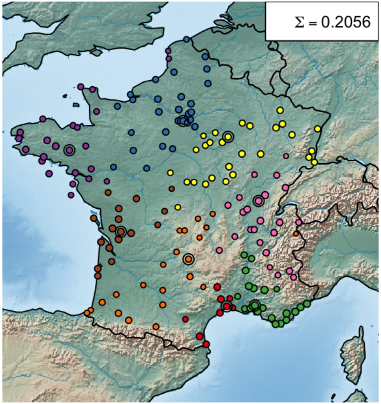


Figure: 8 clusters

$q = 0.9$, $\lambda = 0.5$

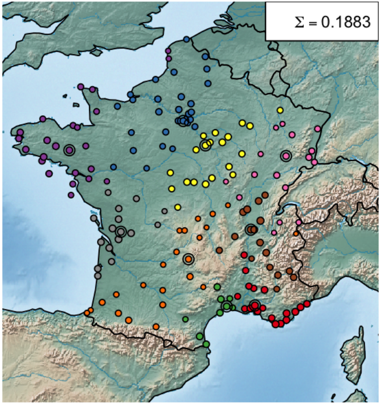


Figure: 9 clusters

q = 0.9, $\lambda = 0.25$

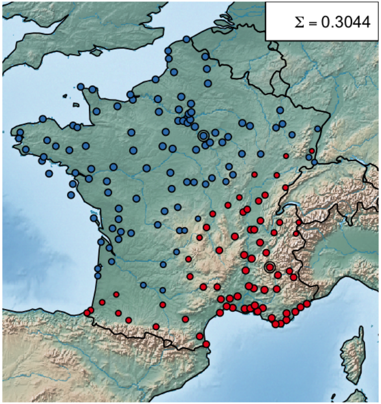


Figure: 2 clusters

$q = 0.9, \lambda = 0.25$

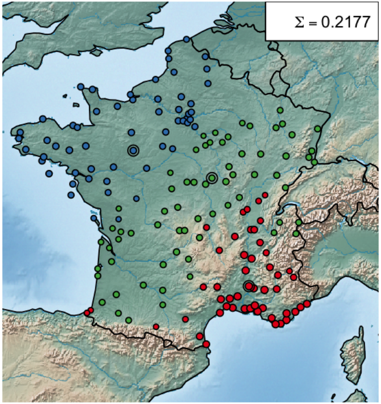


Figure: 3 clusters

$q = 0.9, \lambda = 0.25$

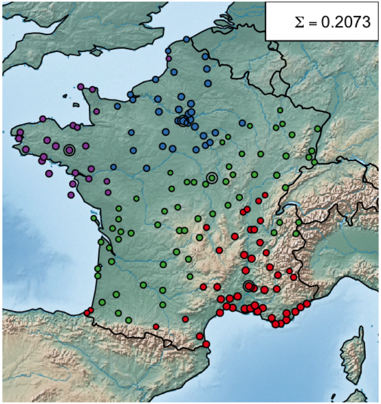


Figure: 4 clusters

$q = 0.9, \lambda = 0.25$

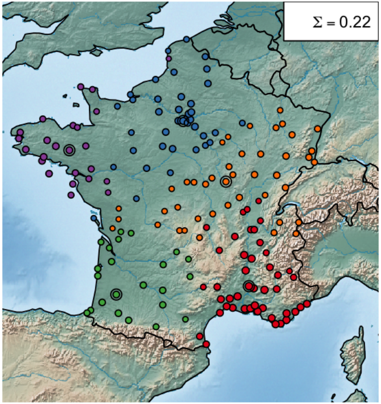


Figure: 5 clusters

$q = 0.9, \lambda = 0.25$

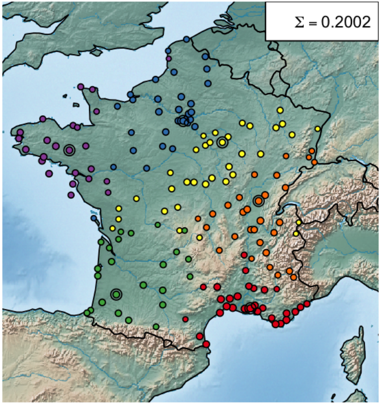


Figure: 6 clusters

$q = 0.9, \lambda = 0.25$

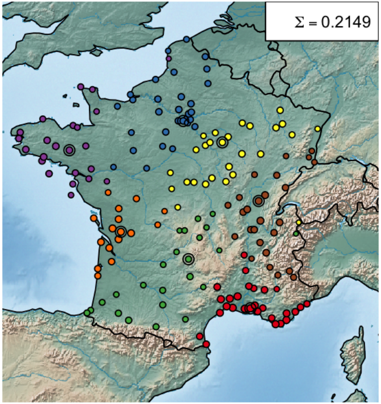


Figure: 7 clusters

$q = 0.9, \lambda = 0.25$

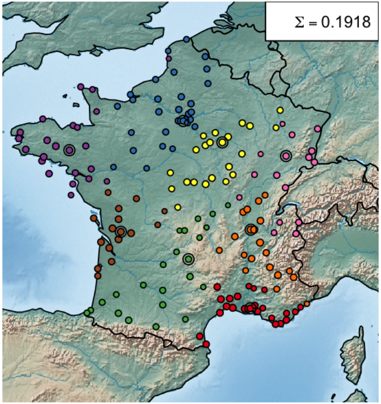


Figure: 8 clusters

$q = 0.9, \lambda = 0.25$

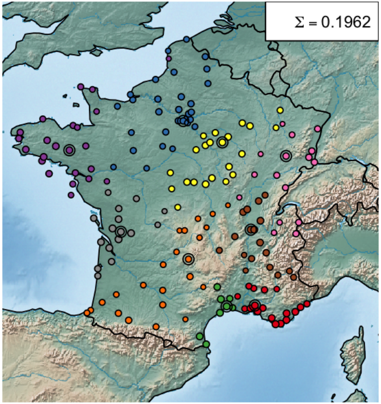


Figure: 9 clusters

$q = 0.9, \lambda = 0$

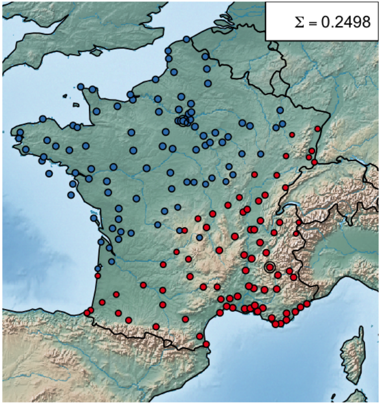


Figure: 2 clusters

$q = 0.9, \lambda = 0$

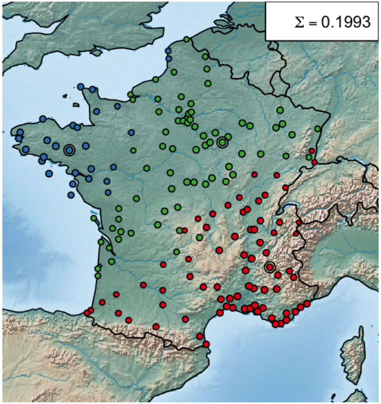


Figure: 3 clusters

$q = 0.9, \lambda = 0$

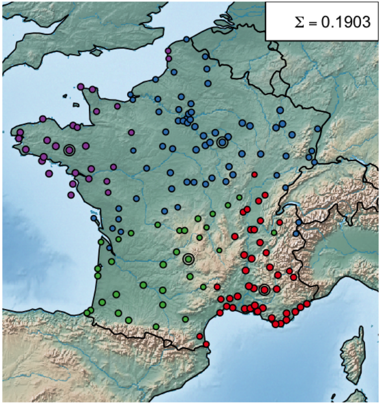


Figure: 4 clusters

$q = 0.9, \lambda = 0$

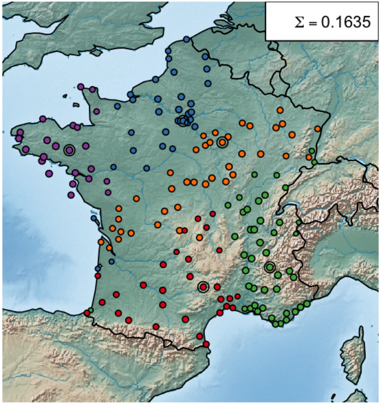


Figure: 5 clusters

$q = 0.9, \lambda = 0$

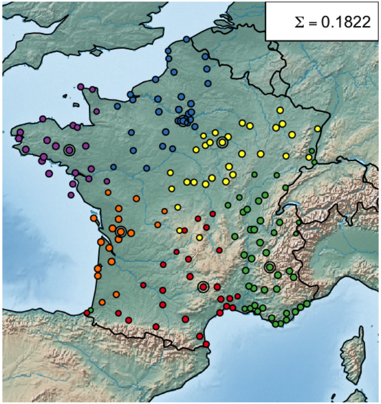


Figure: 6 clusters

$q = 0.9, \lambda = 0$

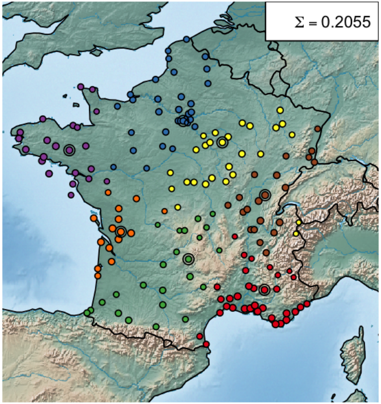


Figure: 7 clusters

$q = 0.9, \lambda = 0$

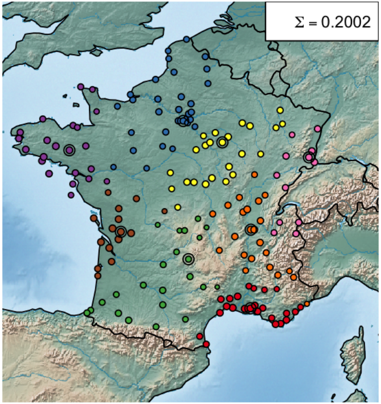


Figure: 8 clusters

$q = 0.9$, $\lambda = 0$

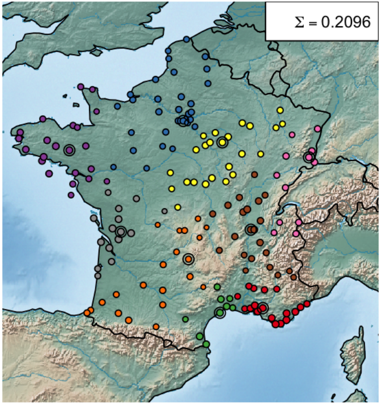


Figure: 9 clusters

Silhouette with number of clusters

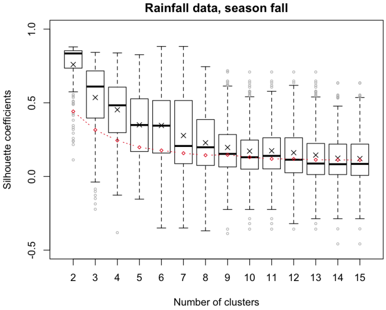


Figure: $\lambda = 1$

Silhouette with number of clusters

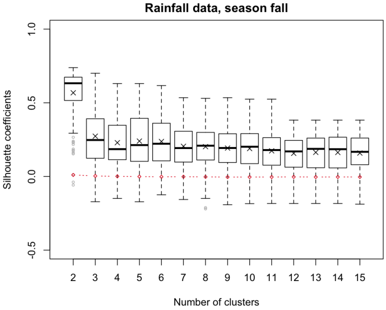


Figure: $\lambda = 0.75$

Silhouette with number of clusters

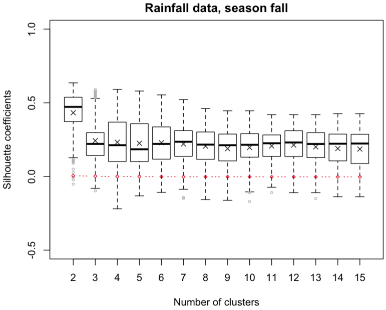


Figure: $\lambda = 0.5$

Silhouette with number of clusters

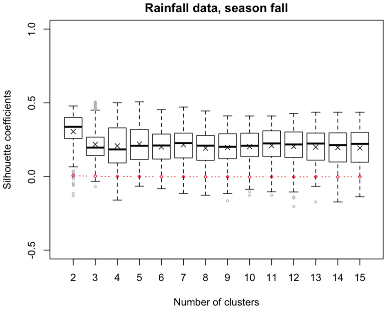


Figure: $\lambda = 0.25$

Silhouette with number of clusters

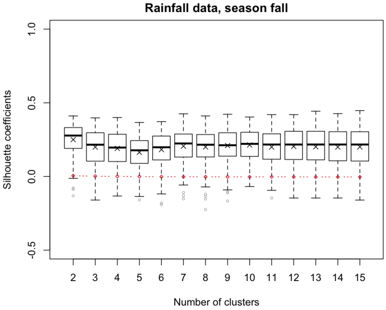
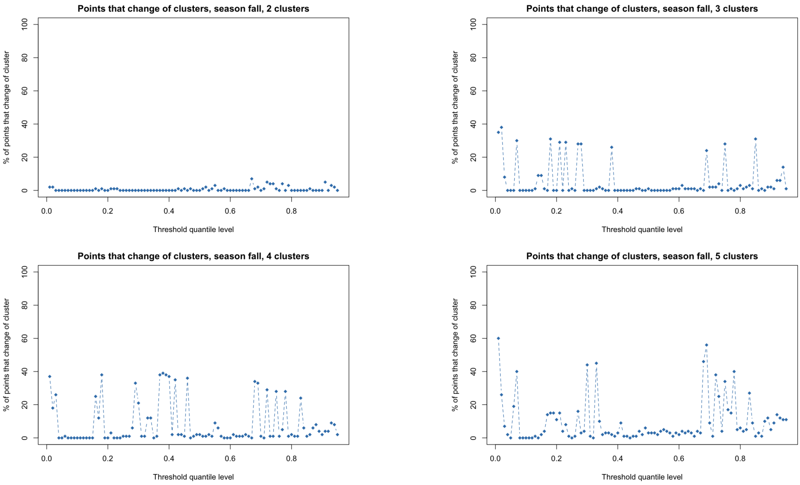
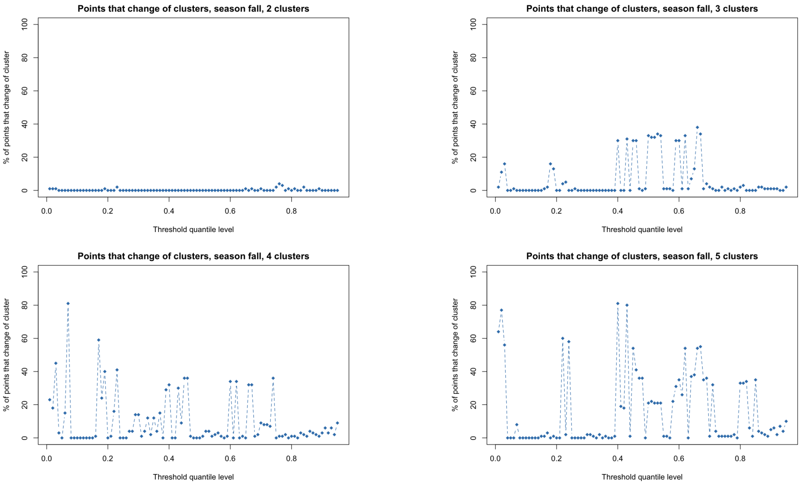


Figure: $\lambda = 0$

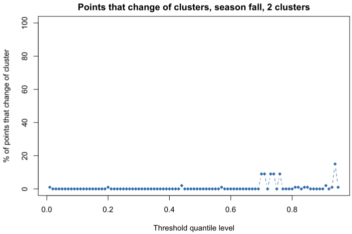
Changing points, $\lambda = 1$

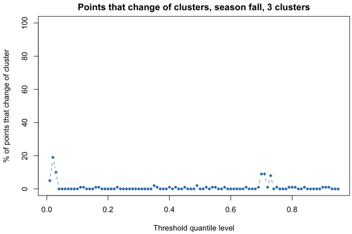


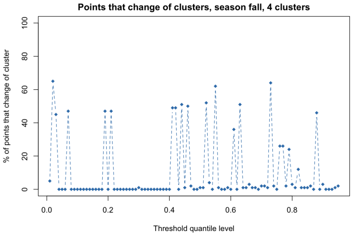
Changing points, $\lambda = 0.75$

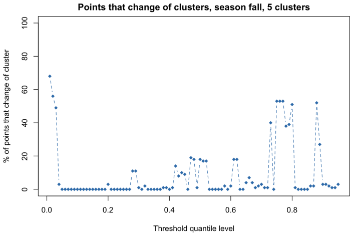


Changing points, $\lambda = 0.5$

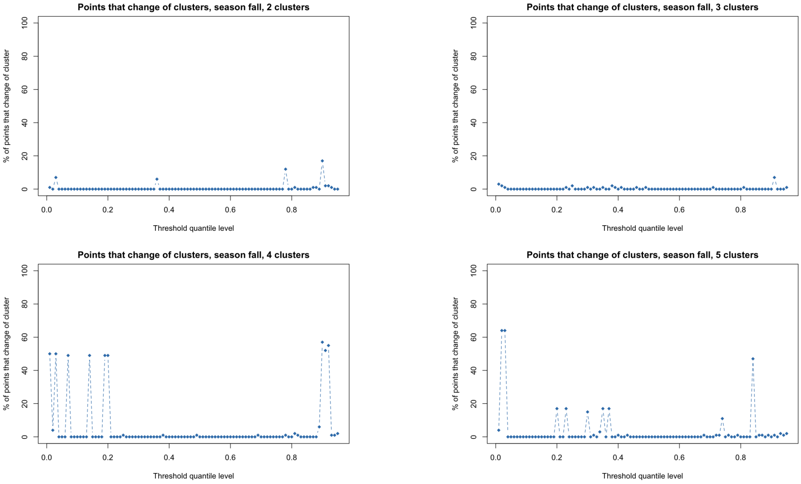








Changing points, $\lambda = 0.25$



No changing, 2 clusters

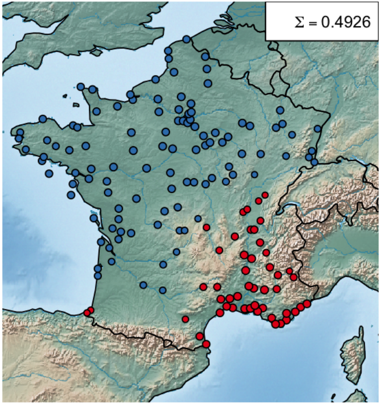


Figure: $q_{\min} = 0.77$

No changing, 3 clusters

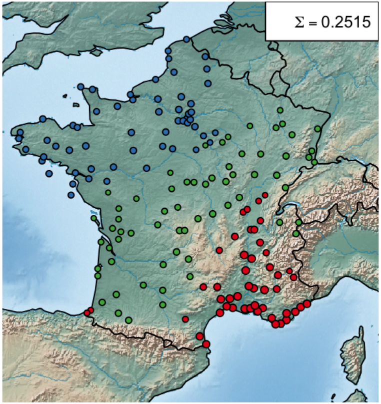


Figure: $q_{\text{min}} = 0.74$

No changing, 4 clusters

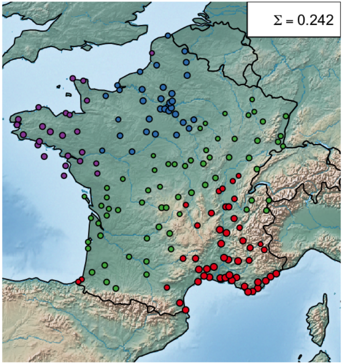


Figure: $q_{\text{min}} = 0.89$

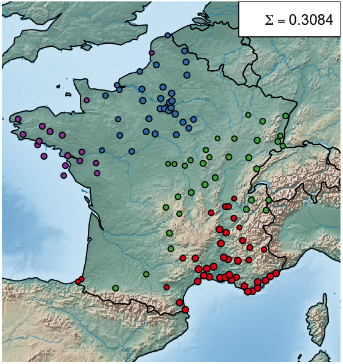


Figure: $q_{\text{min}} = 0.83$

No changing, 5 clusters

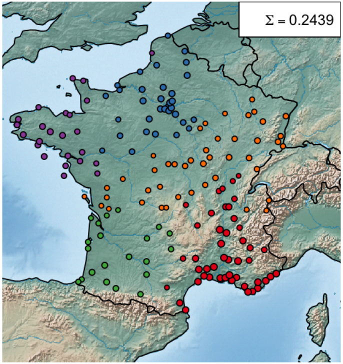


Figure: $q_{\text{min}} = 0.9$

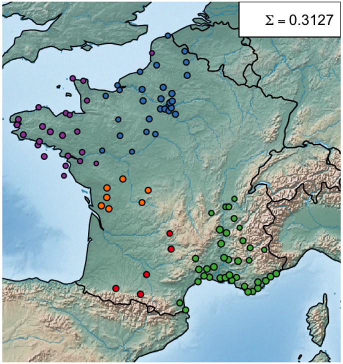


Figure: $q_{\text{min}} = 0.81$

[Bernard et al., 2013]

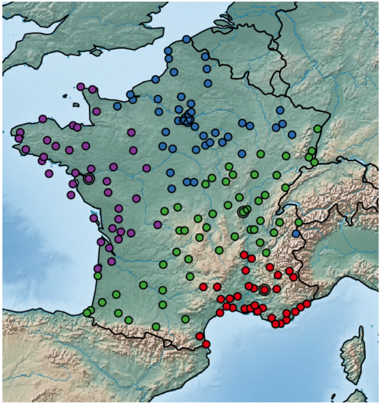


Figure: 4 clusters