Multi-level spectral clustering

RainsMore – 24/10/2022

Kelly Grassi

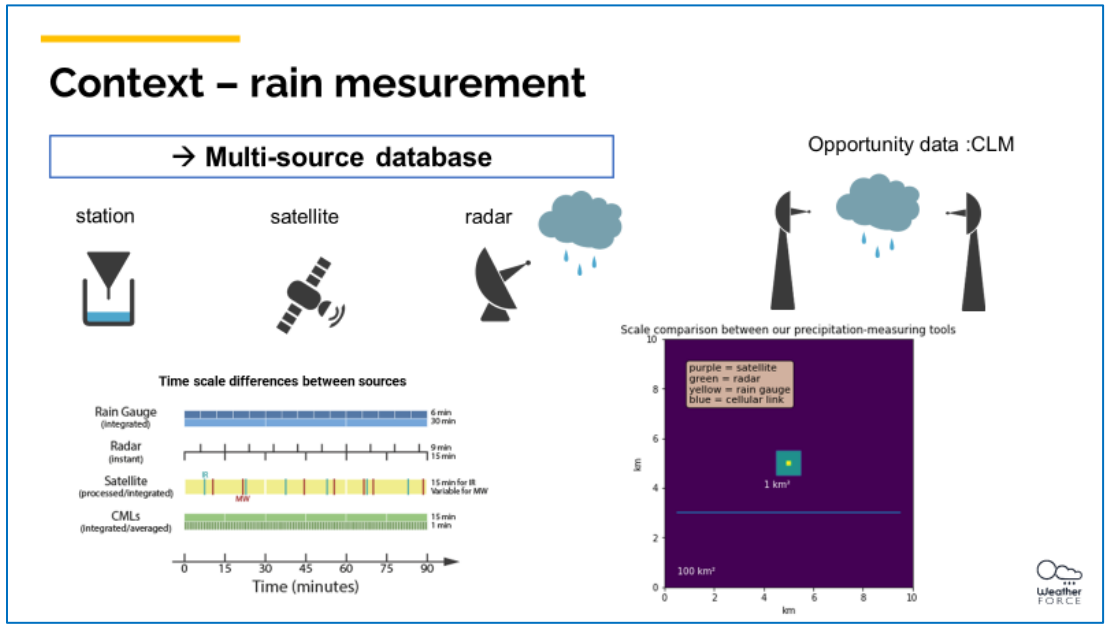
Weather FORCE

Introduction

1. concept and creation process of the M-SC method
2. Exemple of application 1 : dynamique of phytoplancton
3. Exemple of application 2 : water bodies detection

Weather FORCE

Need approaches - Integrated and multi-varied



Development of new digital tools to define environmental states

- **Detect environmental states** in complex **multi-variate and multi-source series**
- **Characterize multi-scale events** ranging from long-term to extreme events and their dynamics in these series
- **Predict** these events for other data sets

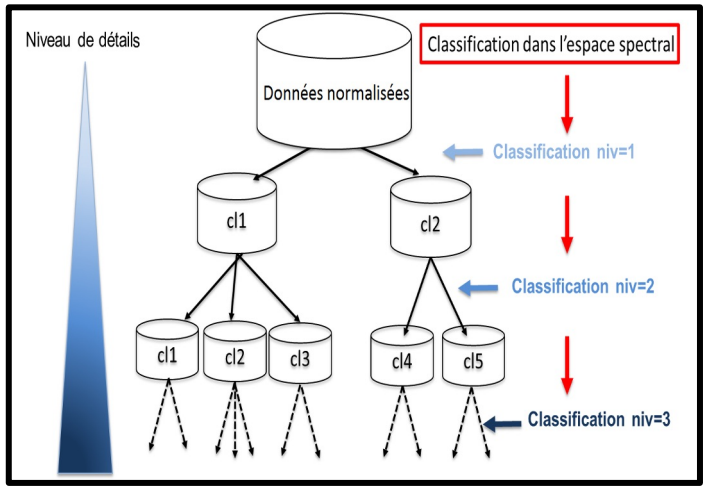
Weather FORCE

Multi-level spectral clustering (M-SC)

Creation of a new method : M-SC

Combination of 3 approaches :

1. Spectral / K'means
2. Hierarchical
3. Density

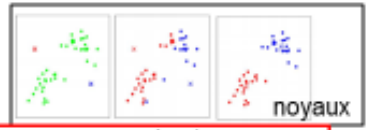


Machine Learning Base

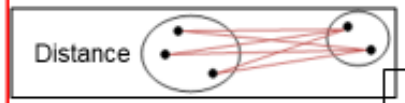
Unsupervised

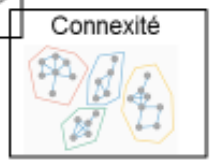
Identification of natural structures

→ Core/Convexity :
K-Means - EM

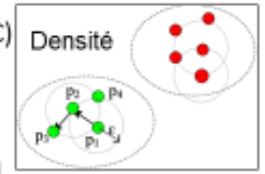


→ Range : *Hierarchical Clustering (HC)*





→ Connexity/Graph:
Spectral Clustering (SC)

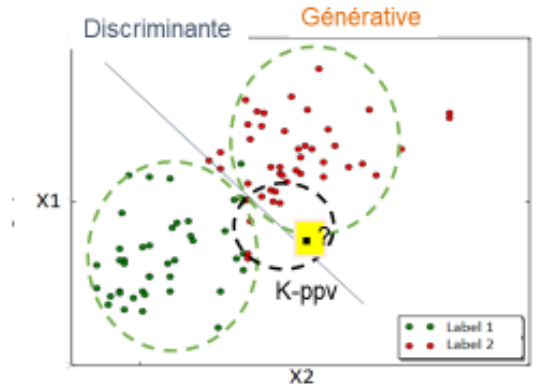


→ **Density : DBSCAN**

Supervised

Identification of boundaries or patterns

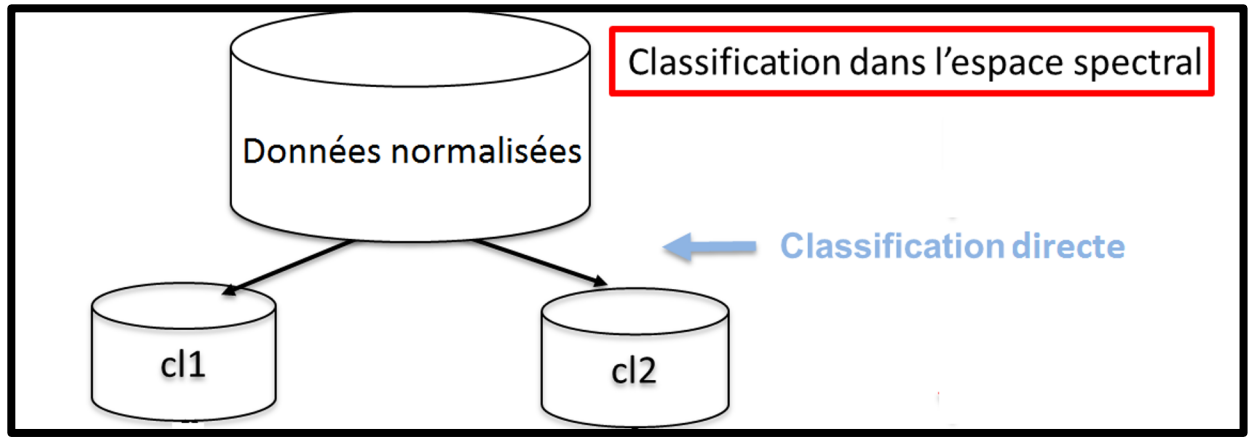
→Discriminant : RF, SVM, MLP, ...
→ Generative : HMM, ...
→ **Other(s) : K-ppv**



Spectral clustering

Large database with complex distribution
stochastic processes
nested, non-linear and non-stationary processes

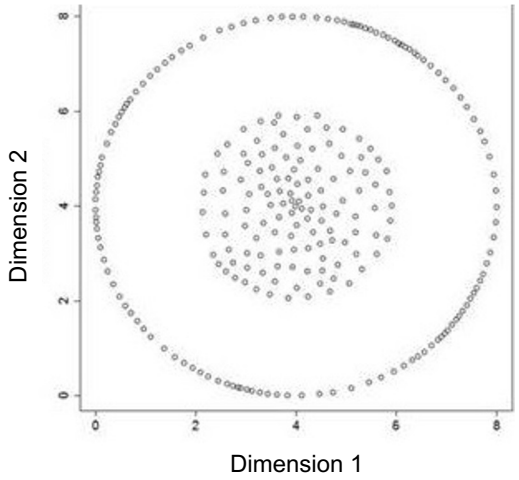
M-SC → Spectral clustering



→ independent of the shape of the data
→ Independent of spatial, temporal and threshold scales

Weather FORCE

Spectral clustering - theoretical case - concept



1

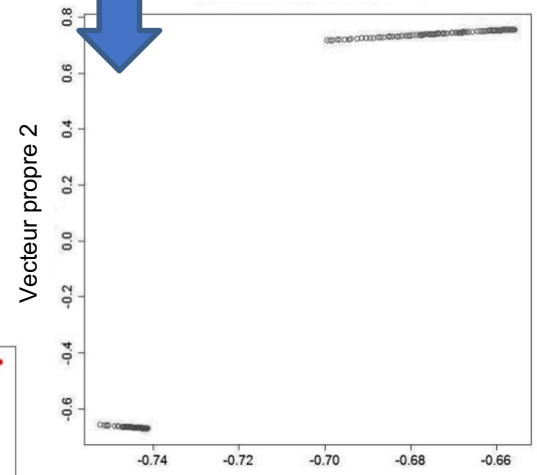
similarity matrix

a_j / a_i	a_{11}	a_{12}
a_{11}	$sim_{a11,a11}$	$sim_{a11,a12}$
a_{12}	$sim_{a12,a11}$	$sim_{a12,a12}$

Calculation of the Similarity matrix

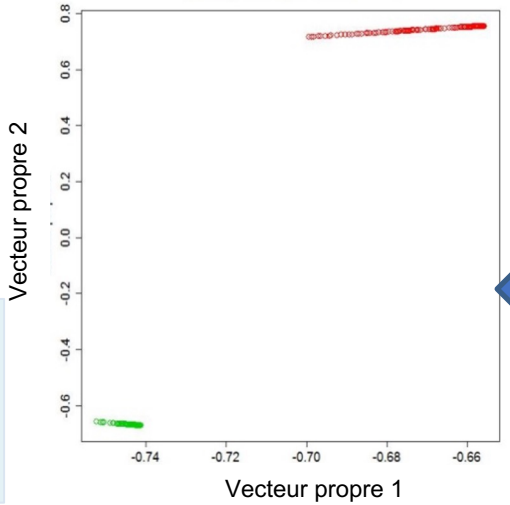
2

Projection of the data in the spectral space (space of K eigenvectors)



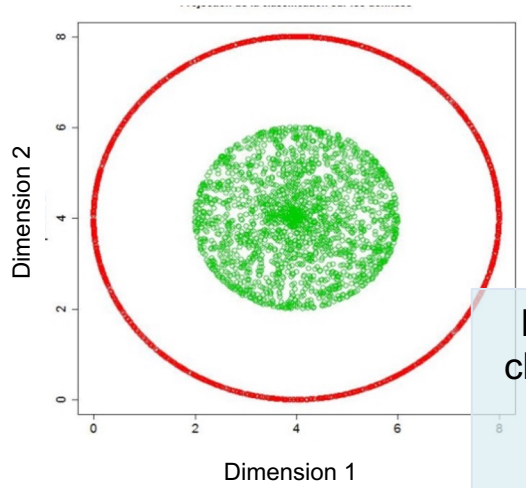
3

Classification K-means



4

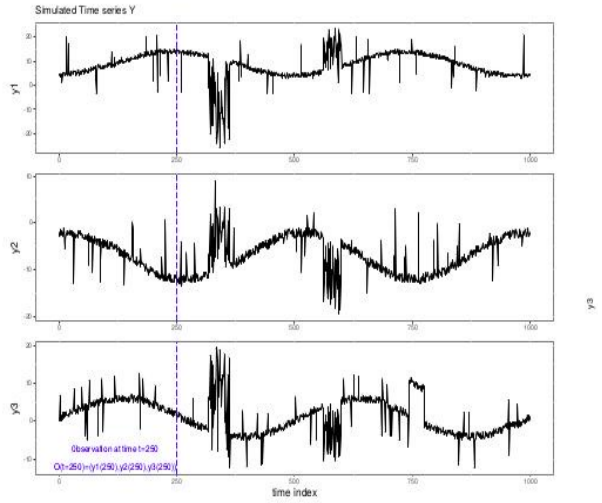
Projection of the classification in the initial (sampled) space



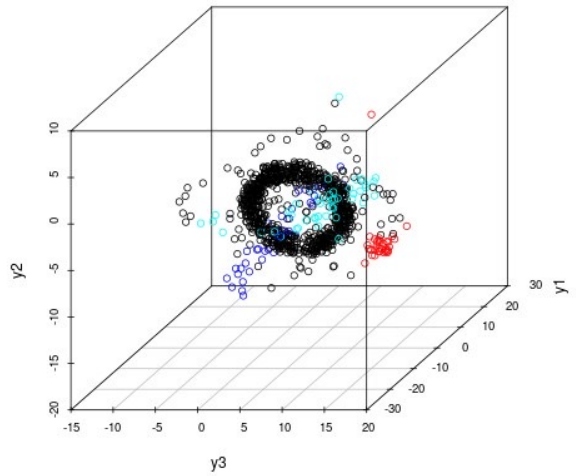
Classification spectrale NJW

(Ng et al., 2001)

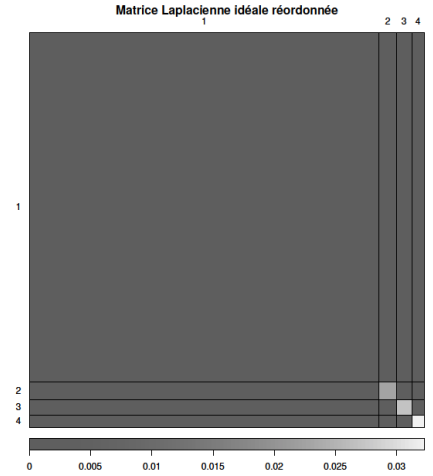
1- Initial signal Y(t)



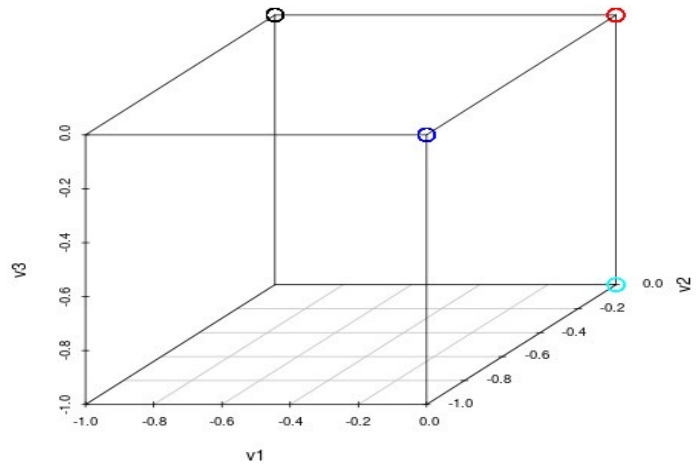
2- Observations Space Y_i



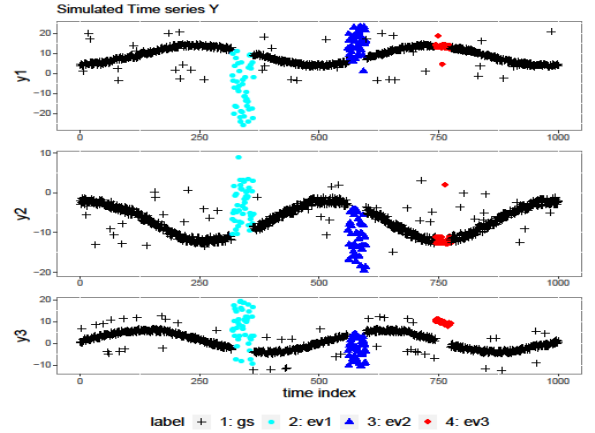
3- Laplacian Matrix



4- Clustering in spectral space



5-Labeling on initial signal

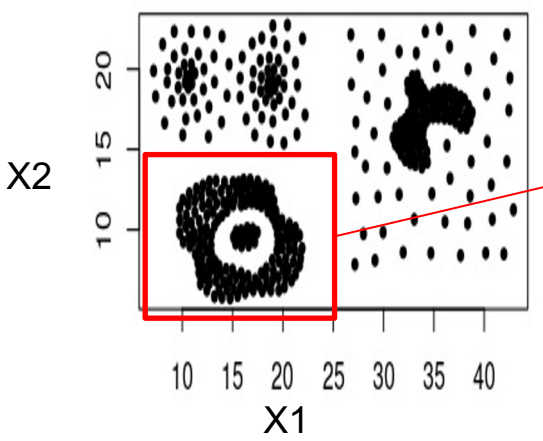


Weather FORCE

Spectral clustering

Exemple : Benchmark

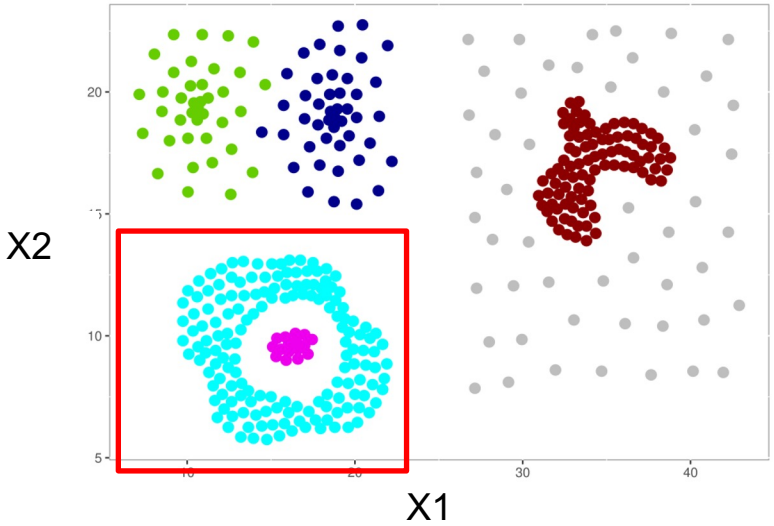
Compound" dataset
UCI data
N=399, k=6, D=2
(source C.T. Zahn, Compound, 1971)



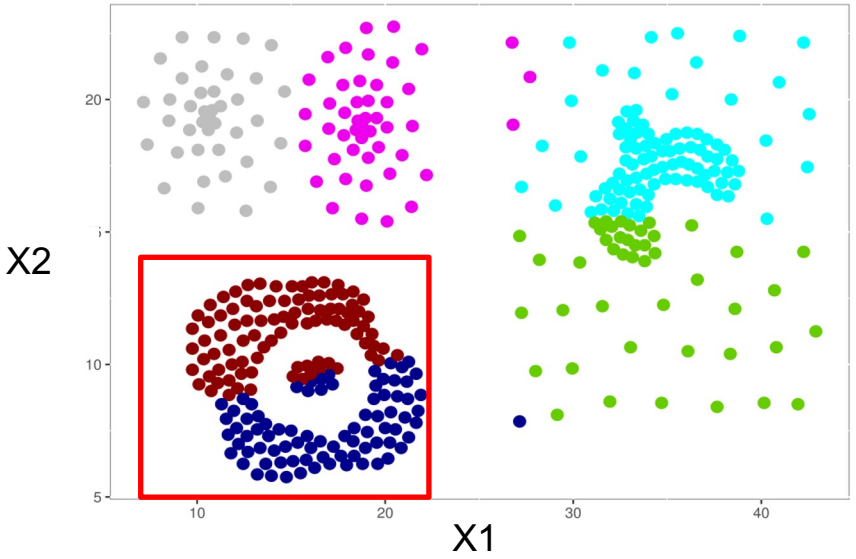
M-SC : Multi-level Spectral Clustering
KM: *K-means* (Hartigan et Wong, 1979)

Nested classes

M-SC Niveau 3



KM K=6



→ M-SC: isolates nested classes

→ KM : tendency to over-segment

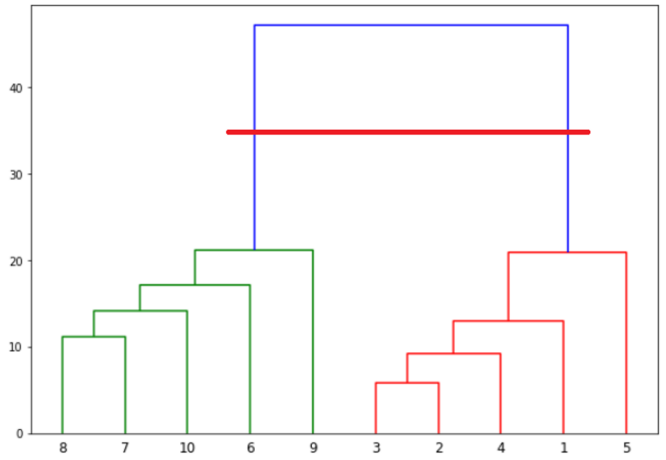
Weather FORCE

Hierarchique approch

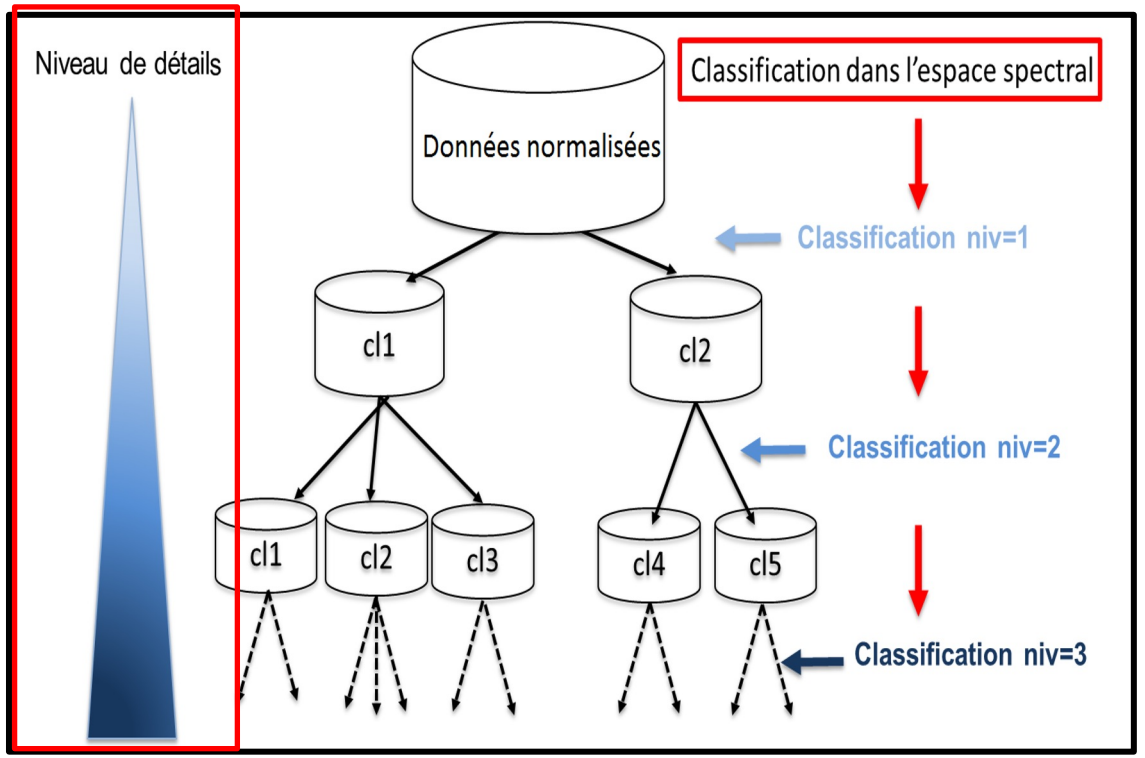
Moving towards the identification of extreme events

M-SC → Spectrale + Hiérarchique clustering

Exemple tree
Hierarchique Clustering



→ Increased level of detail

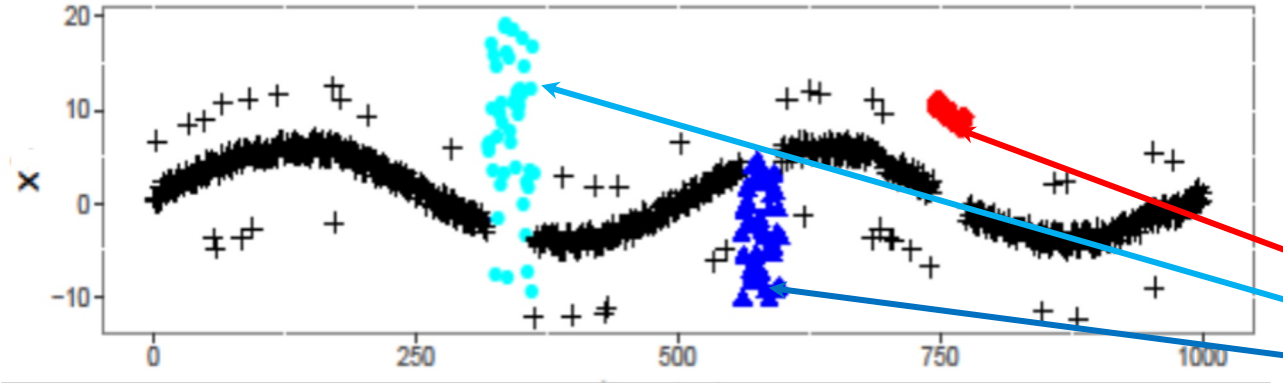


Weather FORCE

Hierarchique approch

M-SC : Multi-level Spectral Clustering
PAM-SC : *Partition Around Medoids (K-medoid)-spectral clustering*
Bi-SC : Bi-parted Spectral Clustering (Garcia et al, 2014)

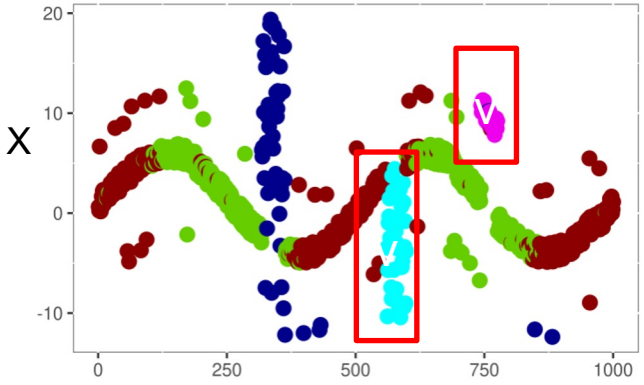
Exemple :
Simulated data



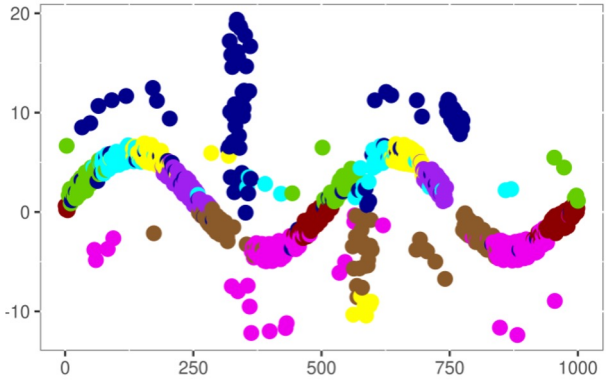
Simulated dataset:
4 components:

Global Signal
Shift
Variation 1
Variation 2

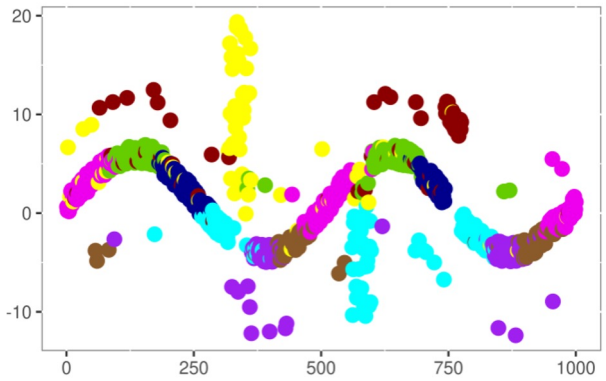
M-SC



PAM-SC



Bi-SC



→ M-SC : definition of extreme states
→ SC-PAM, Bi-SC : Confusion with the global signal

Weather FORCE

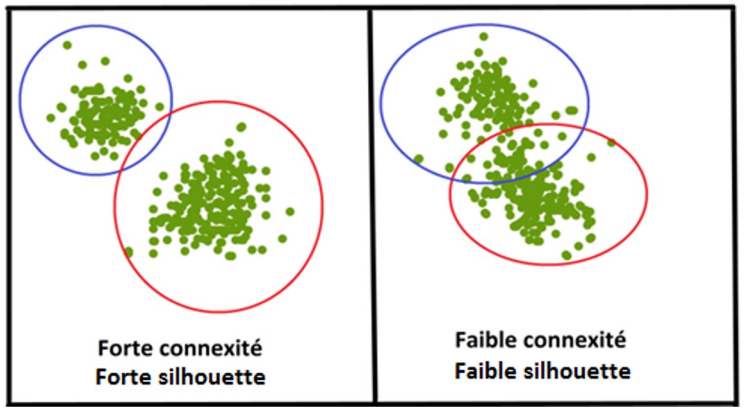
Density

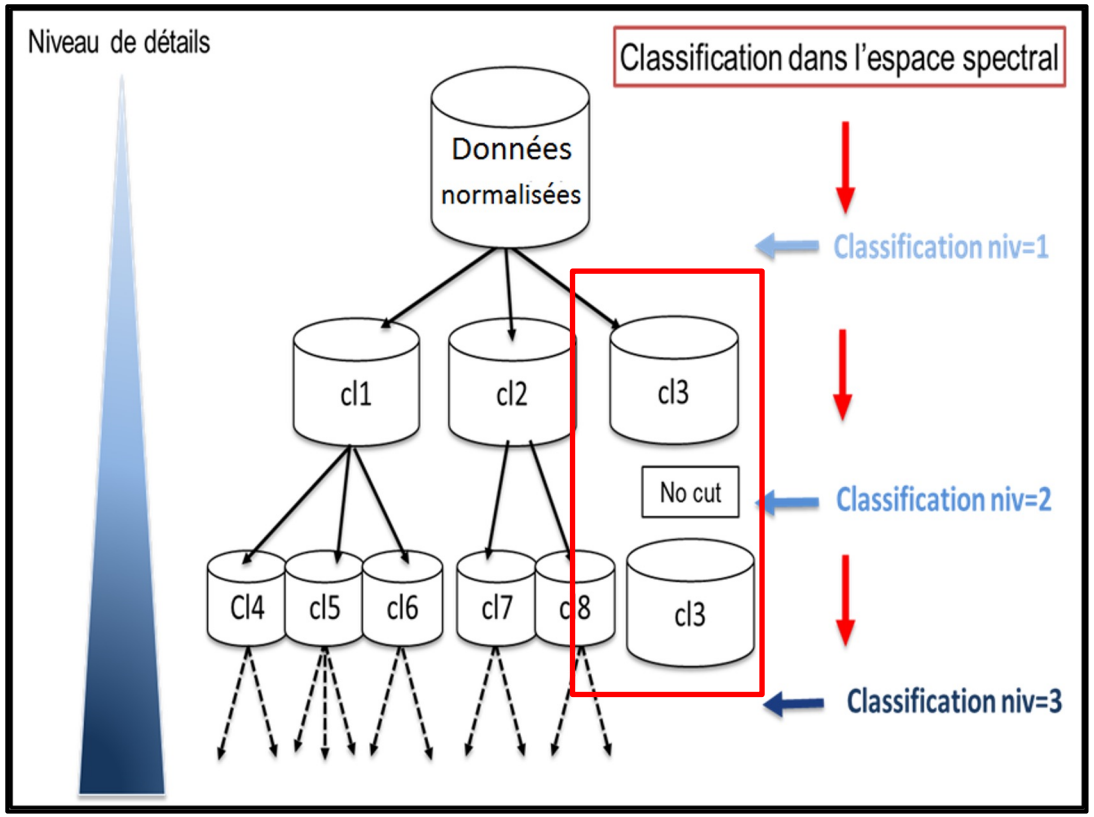
→ Limitation of over-segmentation in deeper levels

→ Automatic calculation of K-classes

M-SC → Spectral + Hierarchique clustering + density

Stopping point : Silhouette





Application example - phytoplankton

Weather FORCE

Dataset

Data Base
MAREL Carnot

High frequency (HF)
automated measurement system

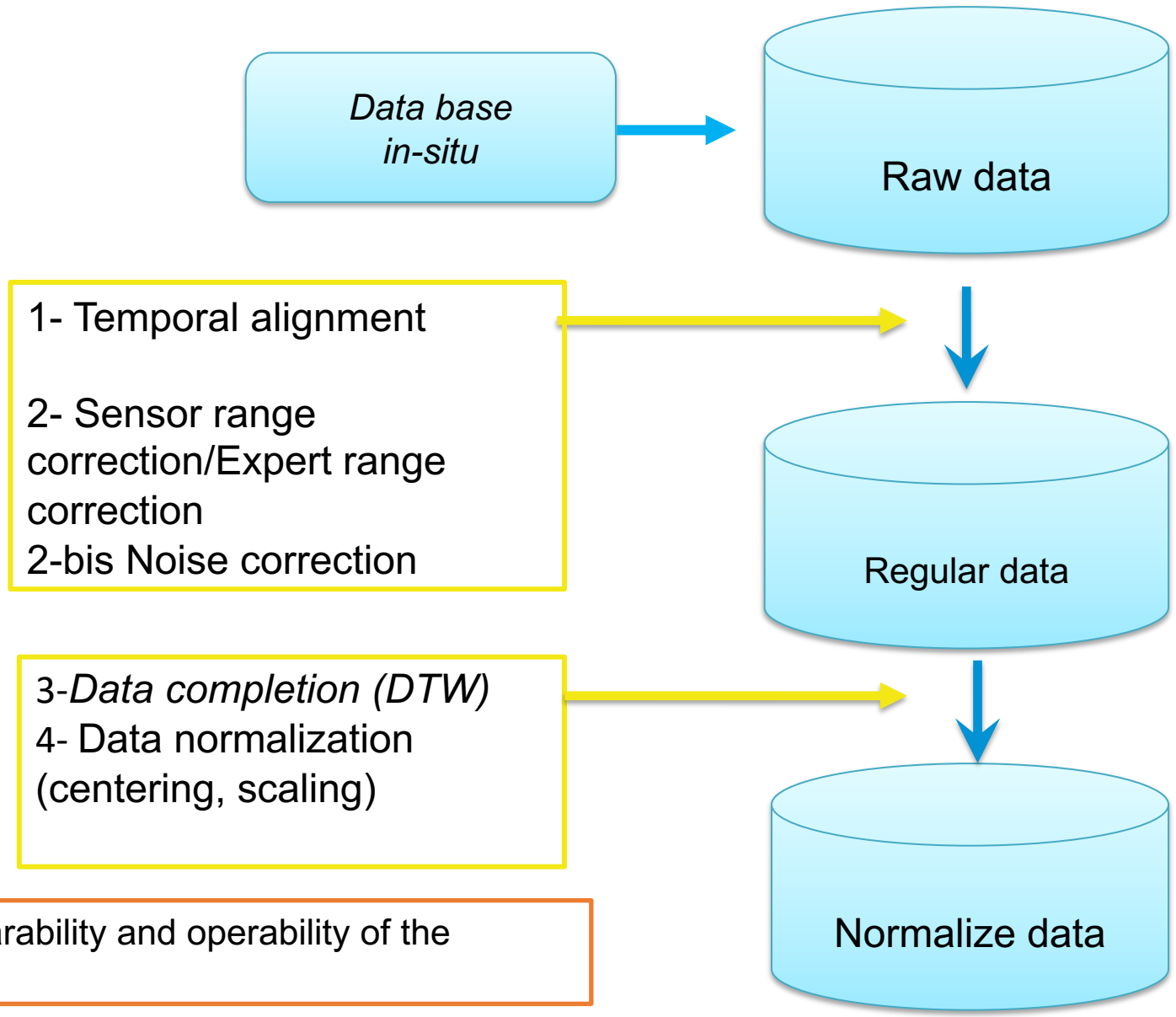
Measurements every 20 minutes
4 years of measurements (2005-2009)

Raw data
*size : 92 968 * 9*
#NA: 320 401 : 38 %

9 explanatory variables - EOVs :

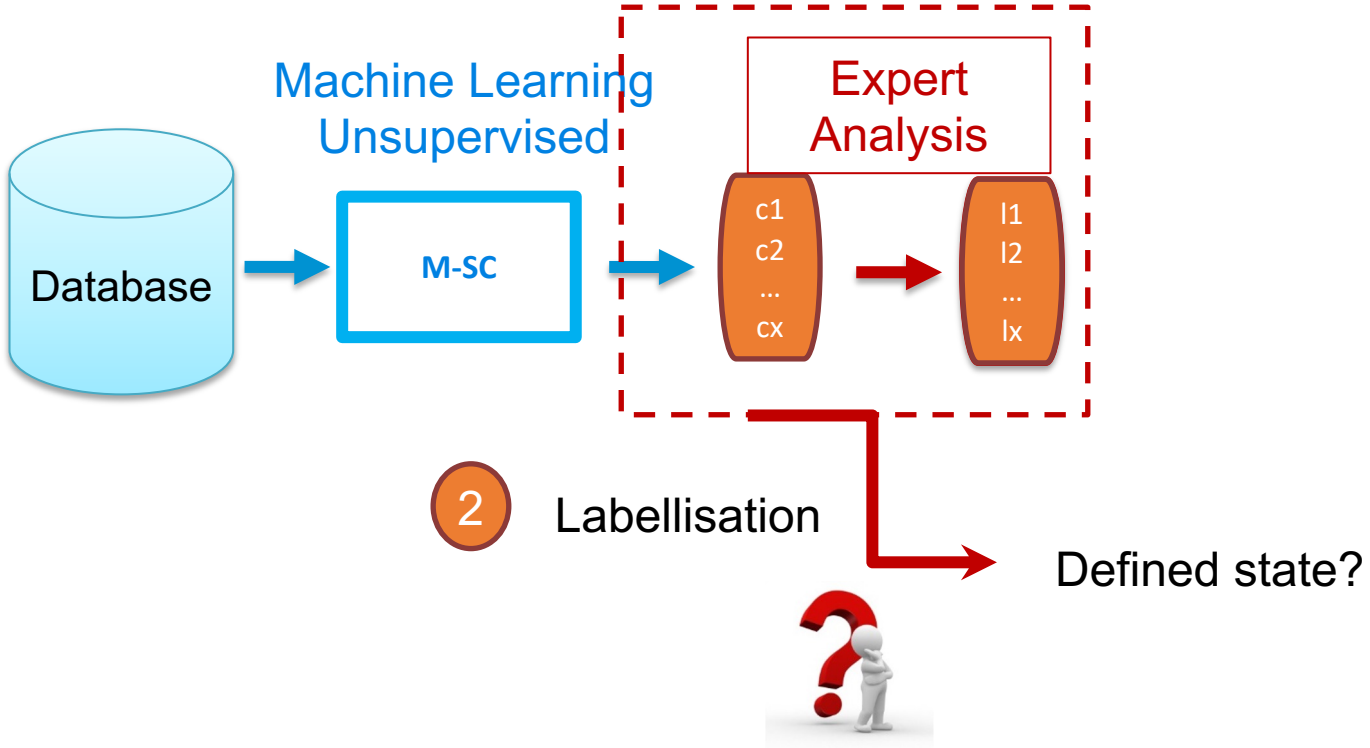
- **Temperature**
- **Salinity**
- **Dissolved Oxygen**
- **Nitrate**
- **Phosphate**
- **Silicate**
- **Turbidity**
- **PAR (Photosynthetically Active Radiation)**
- **Fluorescence**

http://data.coriolis-cotier.org/

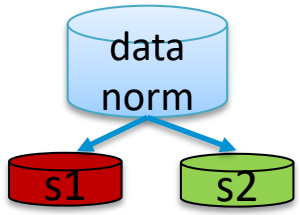


→ Facilitate inter-comparability and operability of the different variables

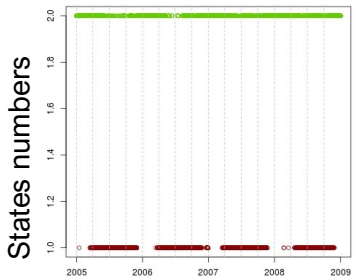
Definition of environmental states of phytoplankton blooms



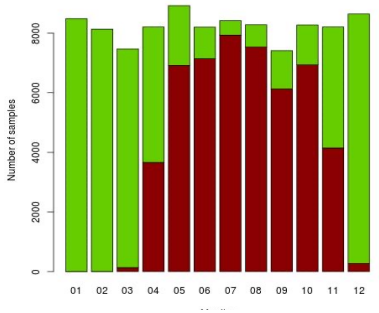
Level 1



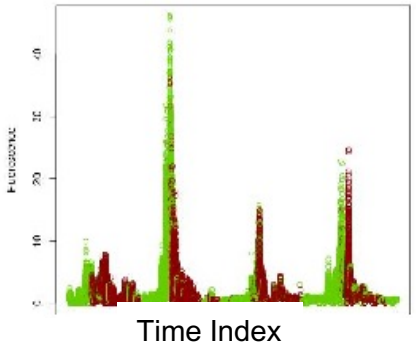
Dynamic



Frequency

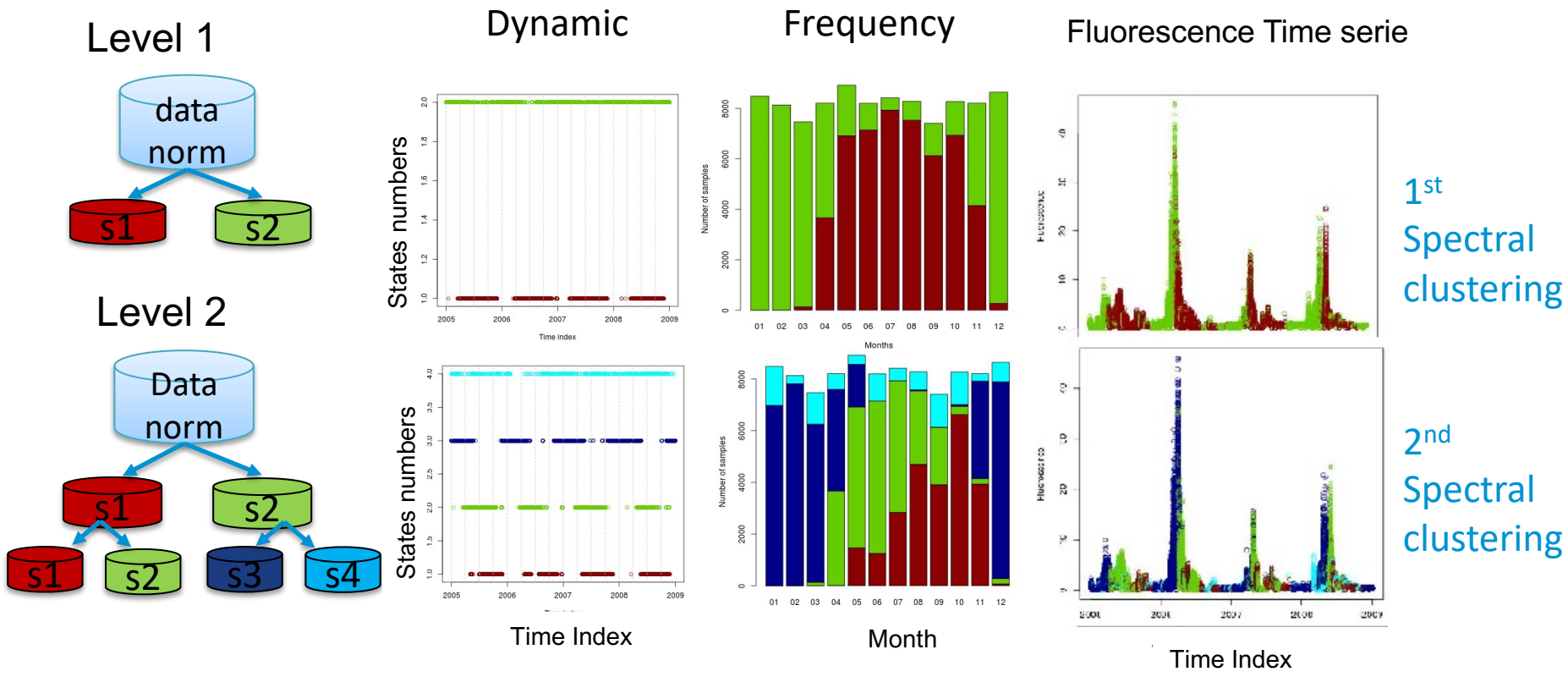


Fluorescence Time serie



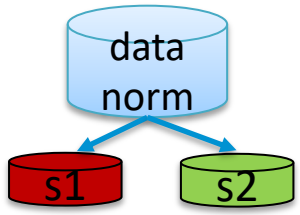
1st Spectral clustering

→ contrasted ecological states are identified: a productive and non-productive period.

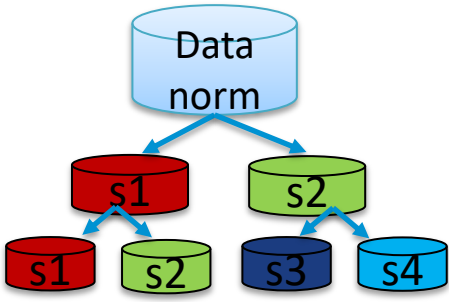


→ two main periods are cut into sub-period, pre-bloom, bloom and post bloom.

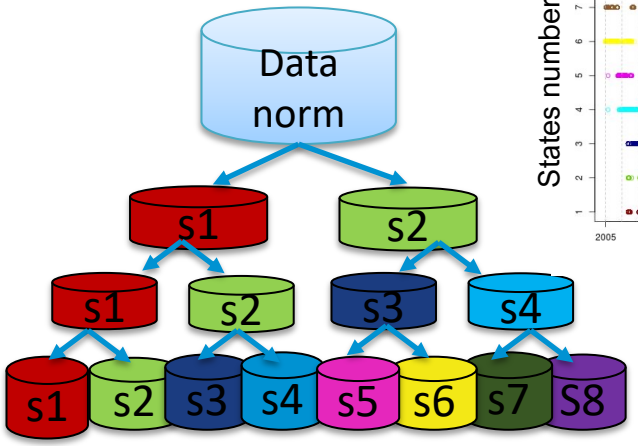
Level 1



Level 2



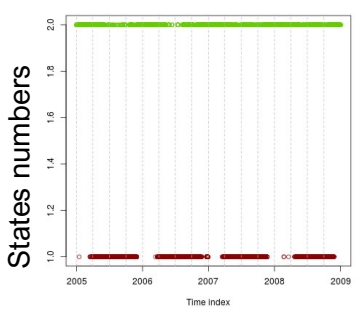
Level 3

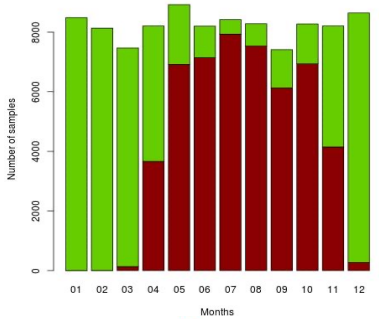


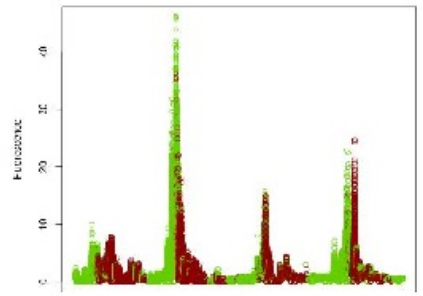
Dynamic

Frequency

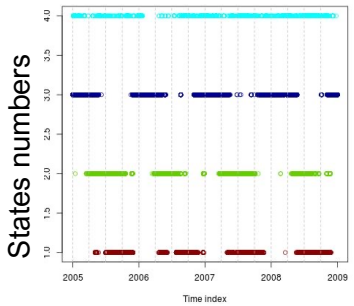
Fluorescence Time serie

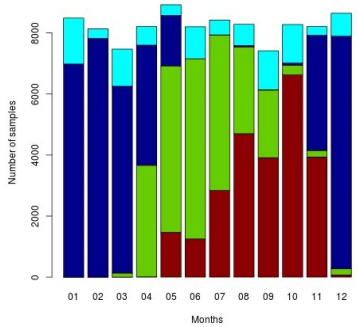


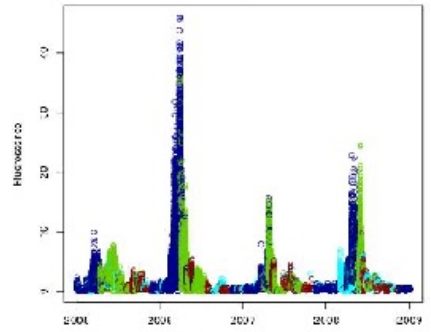




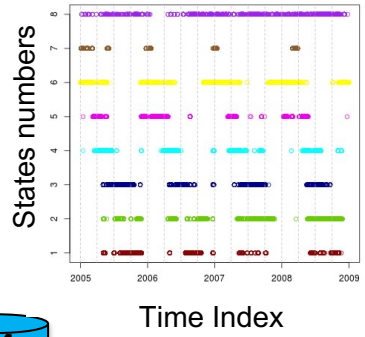
1st Spectral clustering

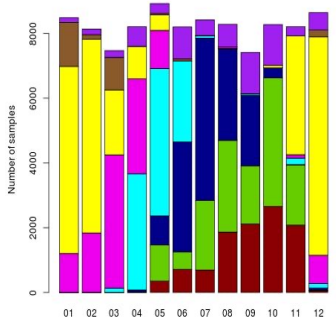


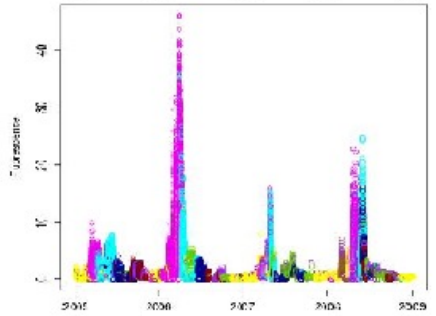




2nd Spectral clustering







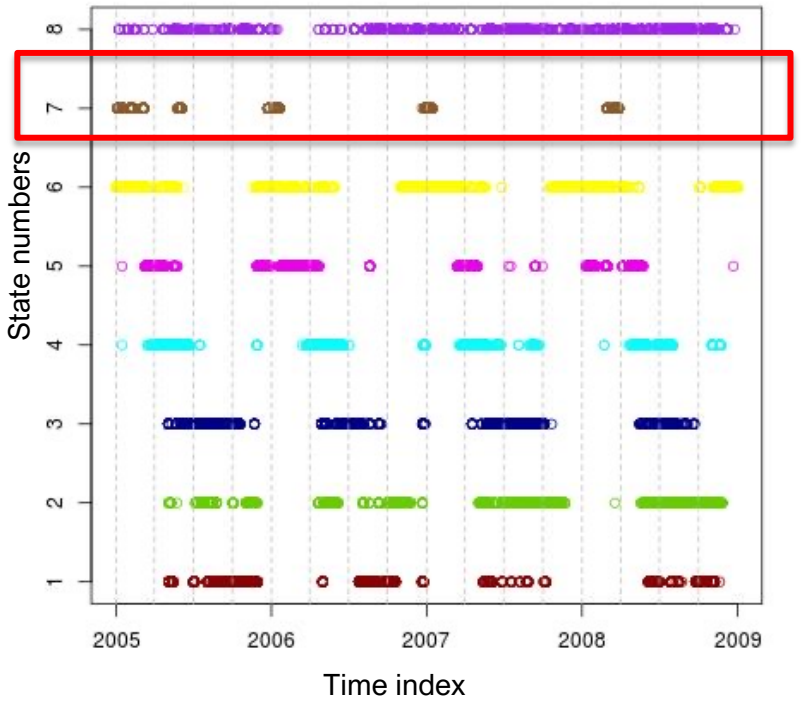
3rd Spectral clustering

→ intra-weekly or hourly events could be identified.

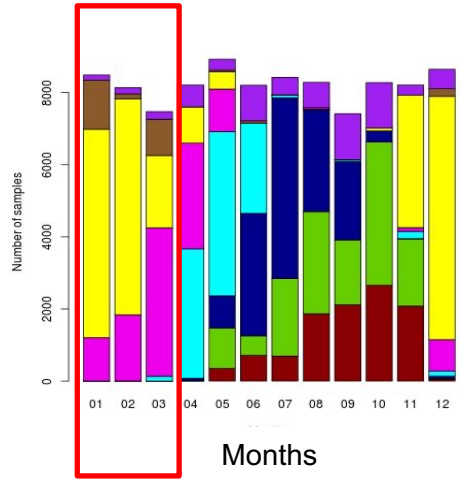
Weather FORCE

Extreme event

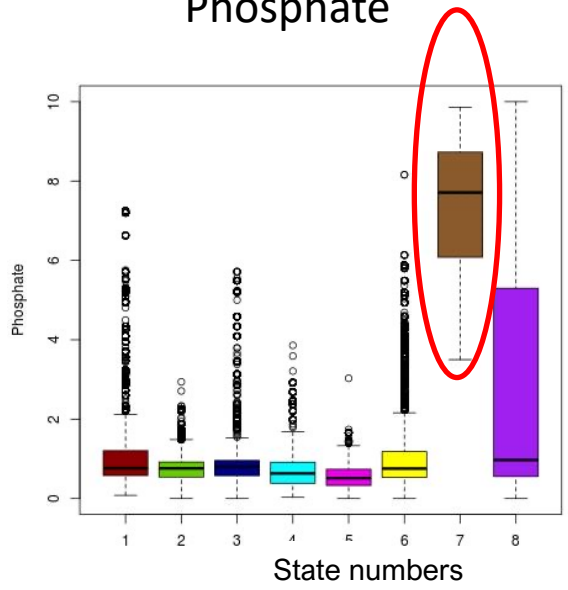
3rd Spectral clustering



Rare/Extreme events



Phosphate



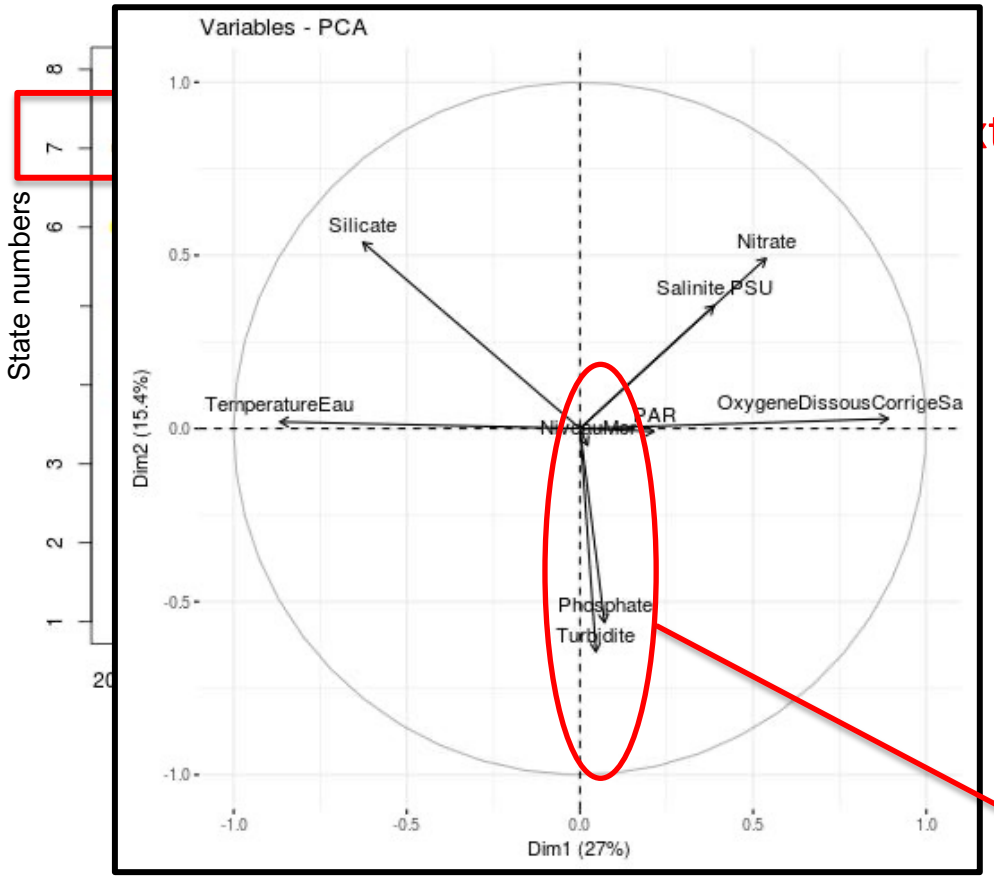
these state have a punctual dynamic more structuring by high phosphate values

Phosphate Correlation State 7 = 0.62

Weather FORCE

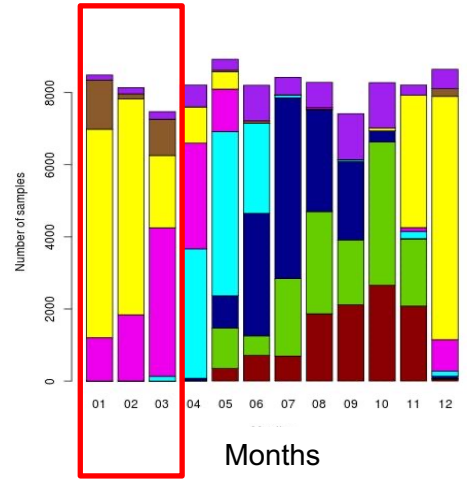
Extreme event

3rd Spectral clustering

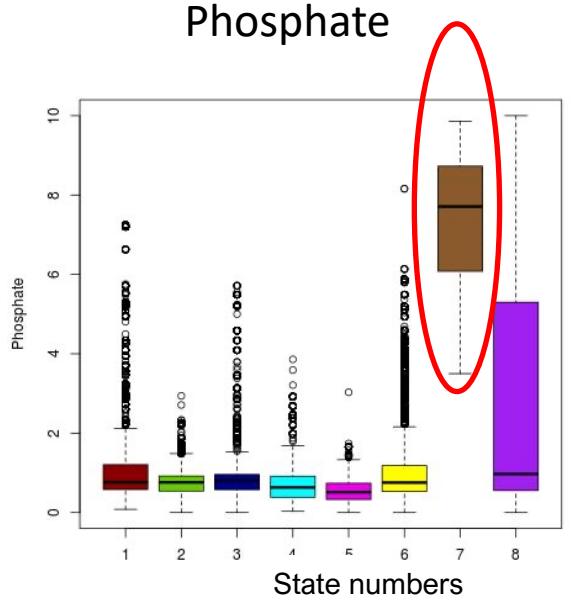


PCA of s7 variables

xtreme events



Phosphate

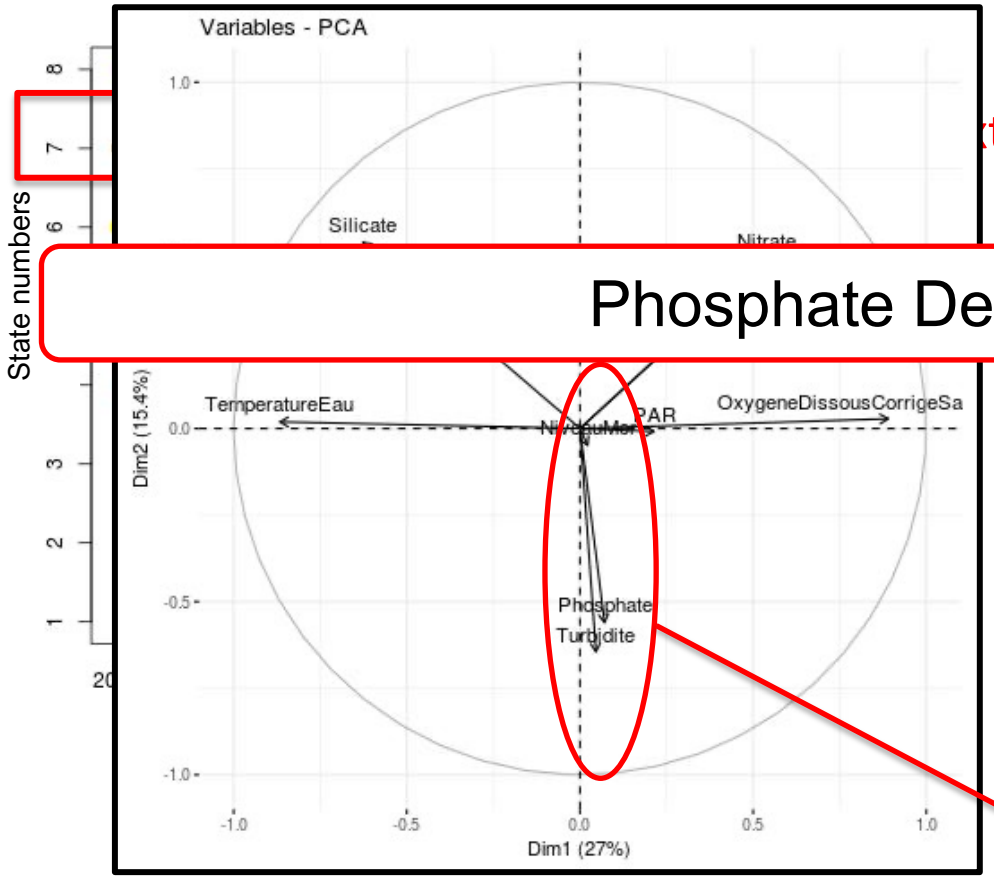


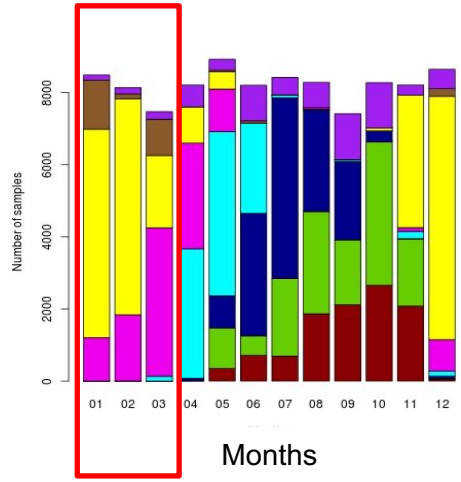
phosphate and turbidity correlation

Weather FORCE

Extreme event

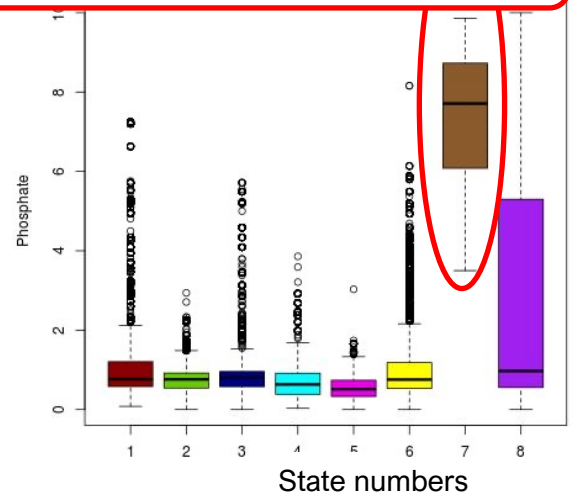
3rd Spectral clustering





xtreme events

Phosphate Desorption

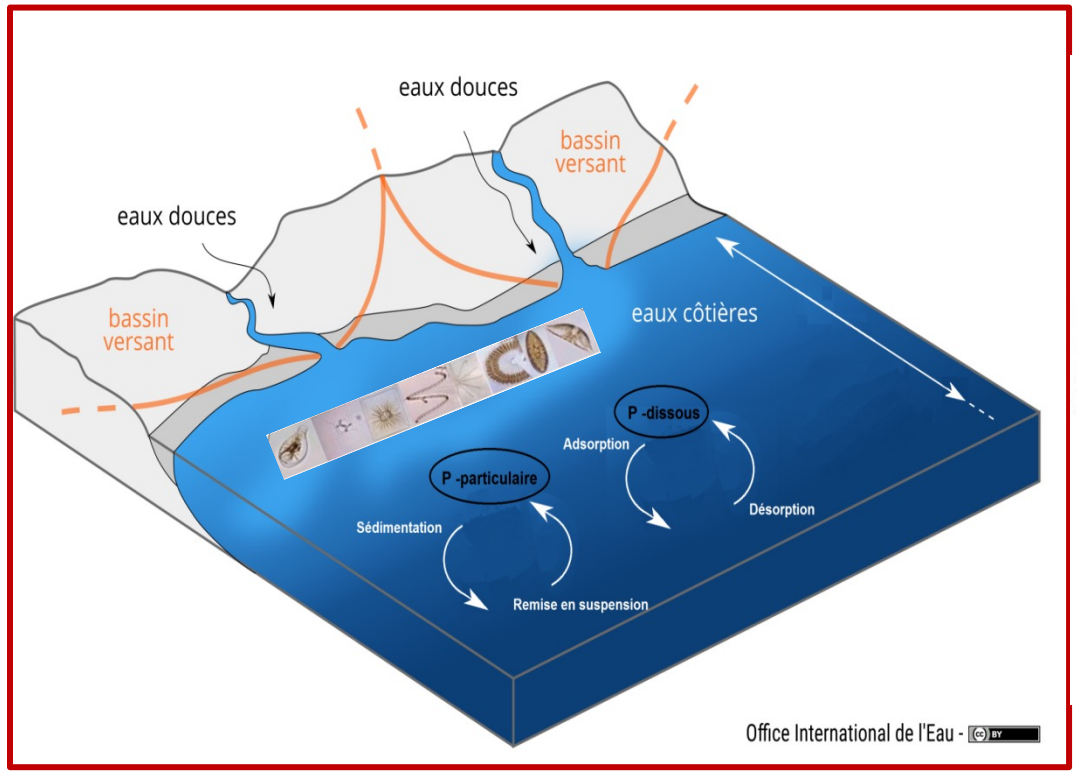


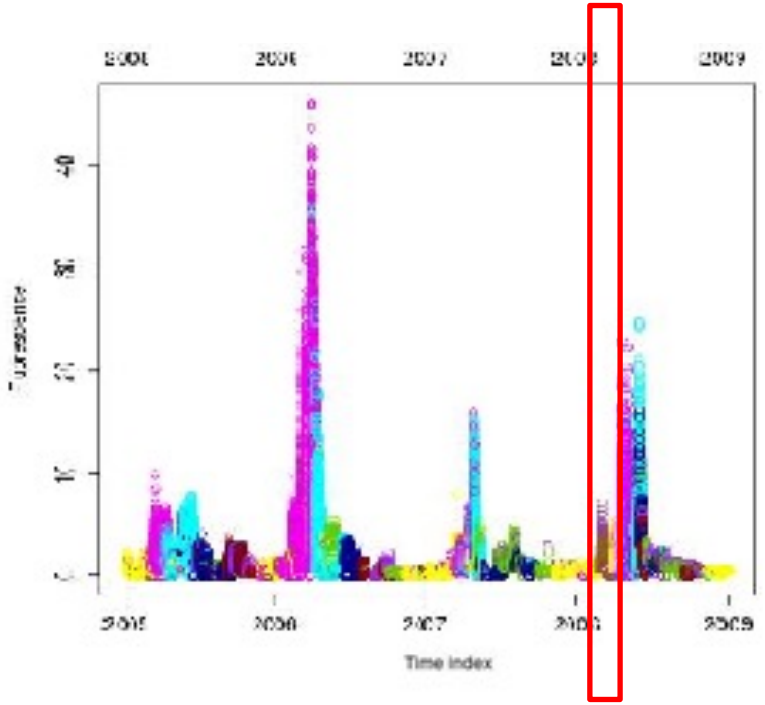
phosphate and turbidity correlation

Extreme event

Phosphate Desorption

New stocks available





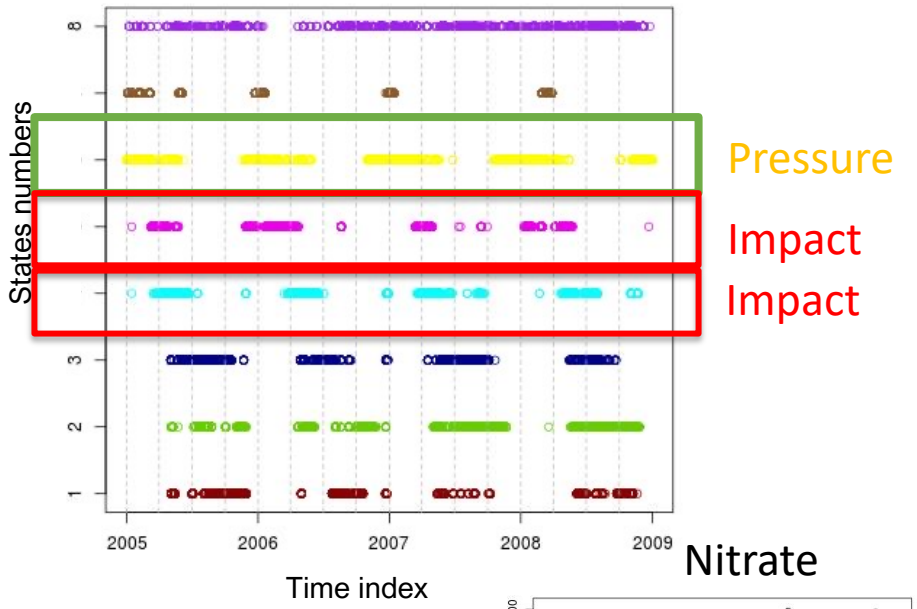
secondary bloom

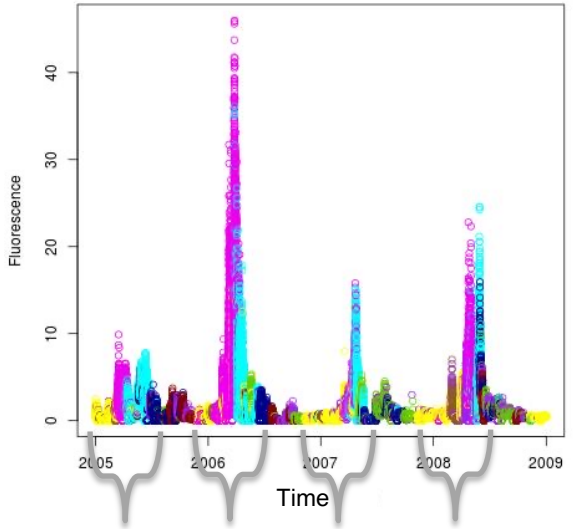
Normaly, it’s difficult to isolate this secondary bloom from the main dynamics.

Weather FORCE

Events : Pressure and Impacts

3rd Spectral clustering

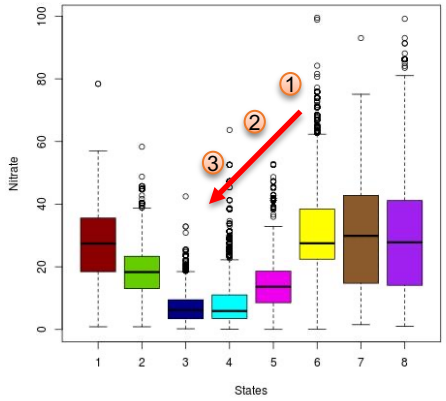




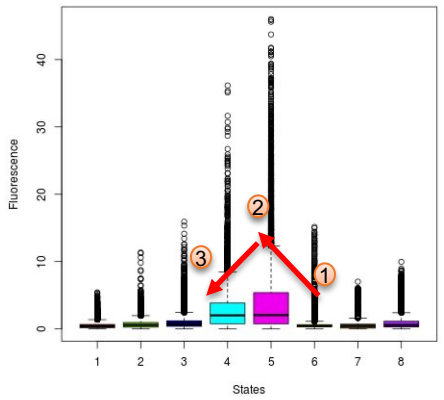
detect pressures and impacts states

1 - decrease of nutrients
2 - Increase of phytoplankton
3 - Re-balancing

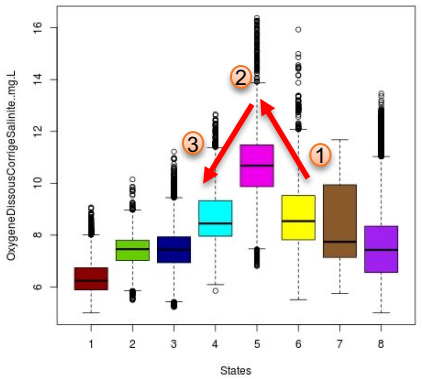
Nitrate



Fluorescence

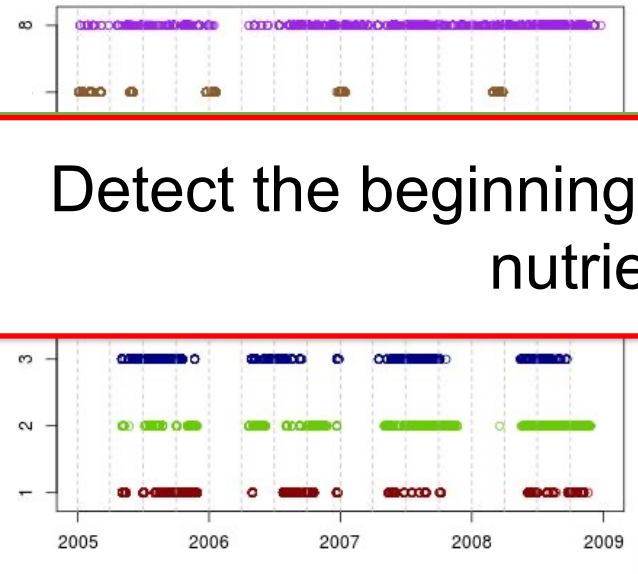


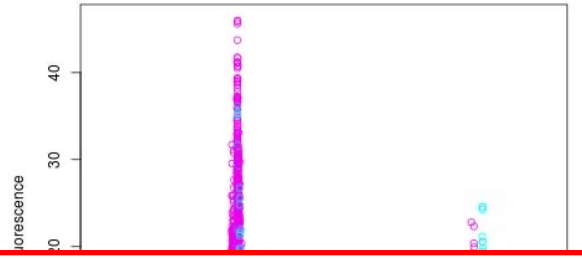
Dissolved Oxygen



Events : Pressure and Impacts

3rd Spectral clustering

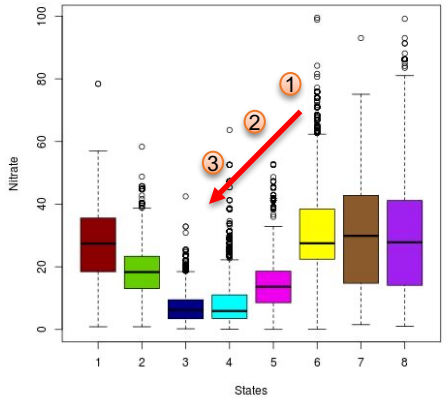




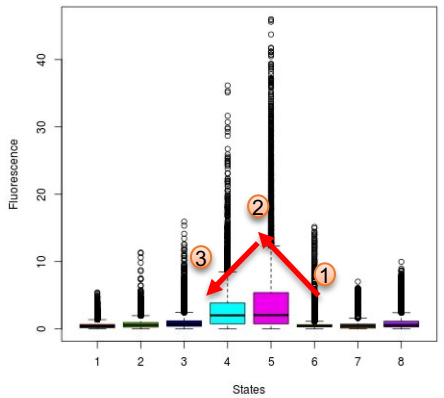
Detect the beginning of a phytoplankton bloom when nutrients are supplied

Succession events

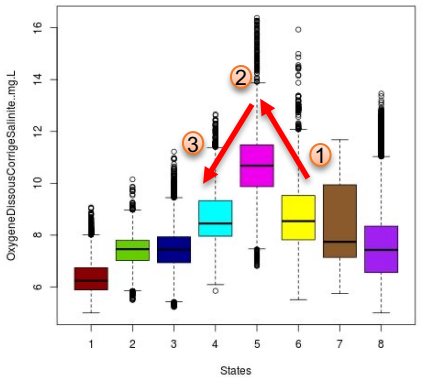
Nitrate



Fluorescence



Dissolved Oxygen



Weather FORCE

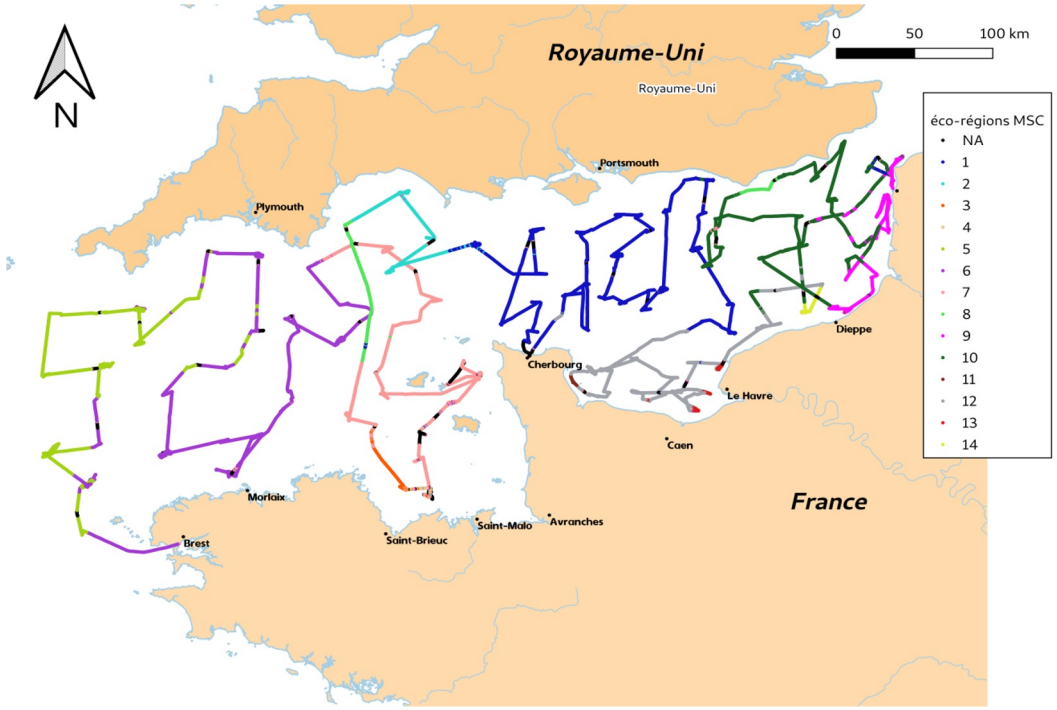
Application examples - detection of water bodies

Weather FORCE

- Contribution to new definitions and interpretations of processes

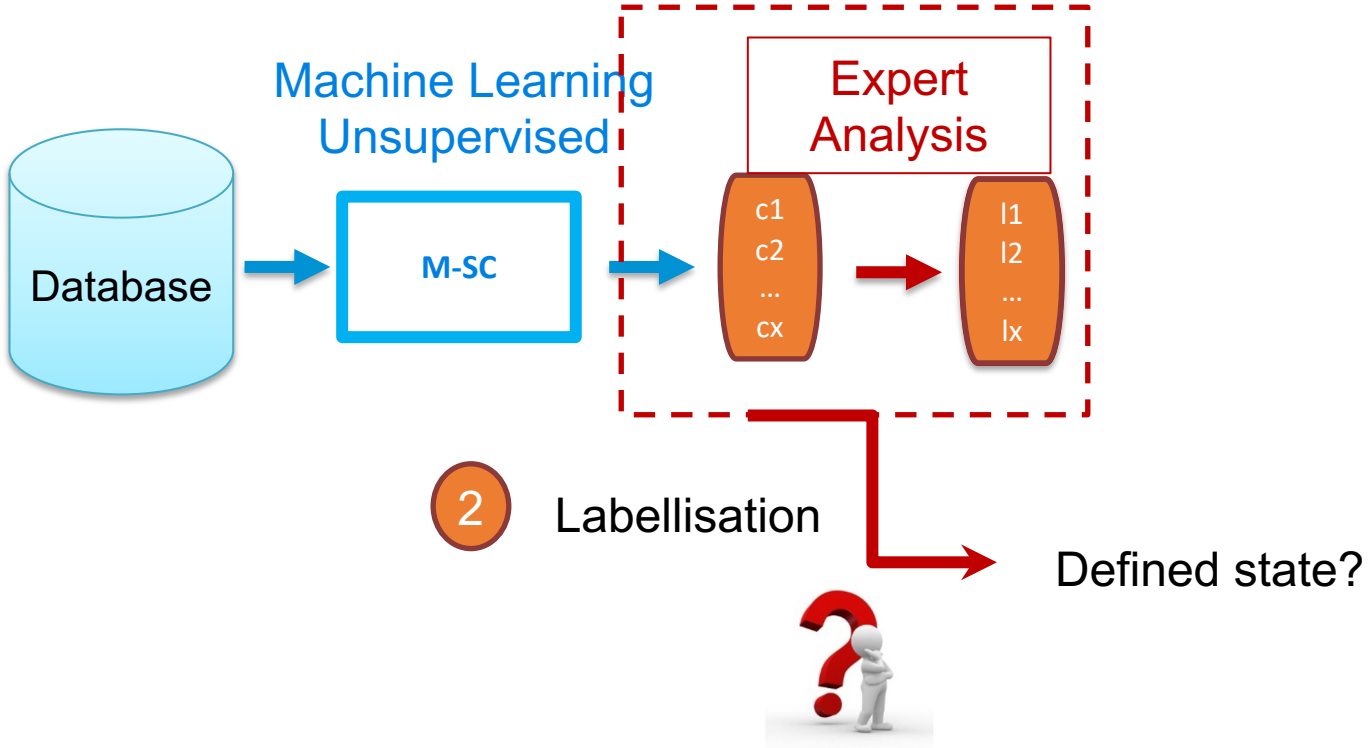
- Contribution to the definition of adaptive sampling strategies during sea campaigns

→ Identification of water bodies (eco-hydrodynamic landscapes)

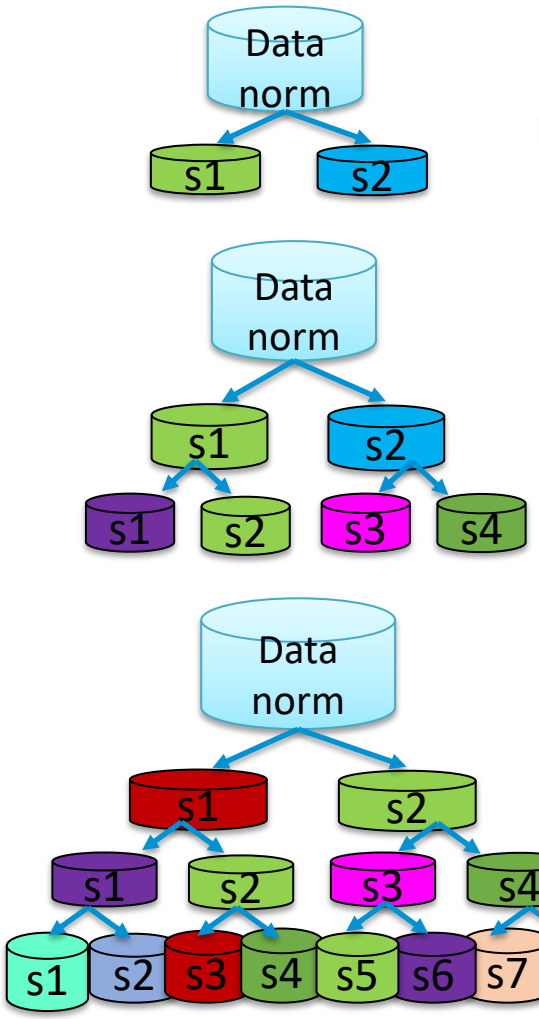


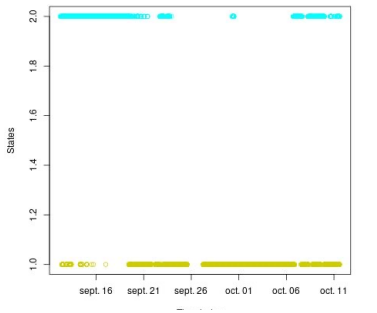
FerryBox Data
Campagne CGFS

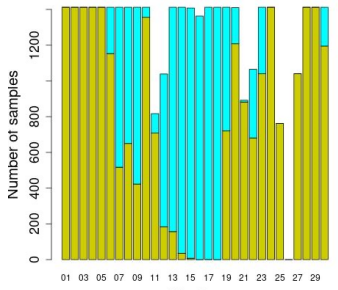
Definition of environmental states of phytoplankton blooms

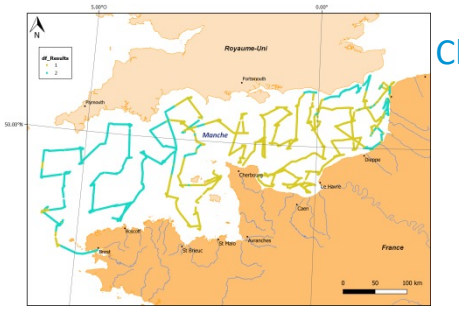


Application: Multi-level spectral clustering

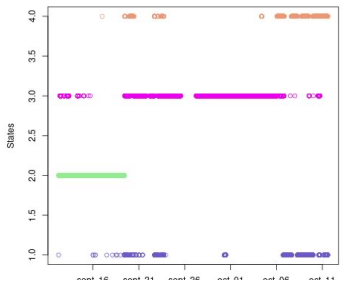


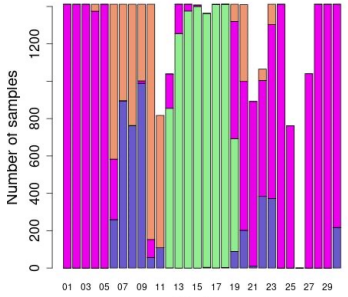


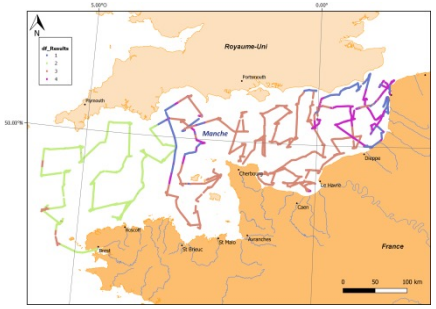




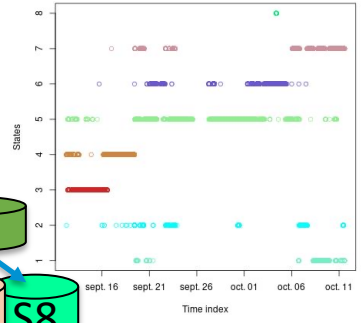
1ere Classification Spectrale

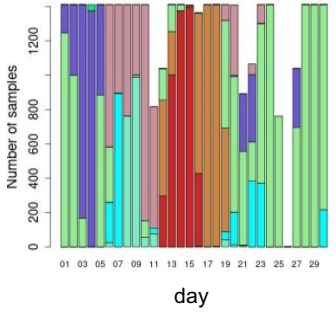


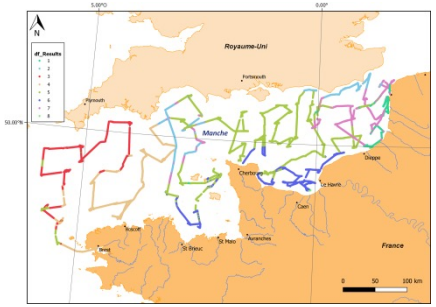




2nd Classification Spectrale







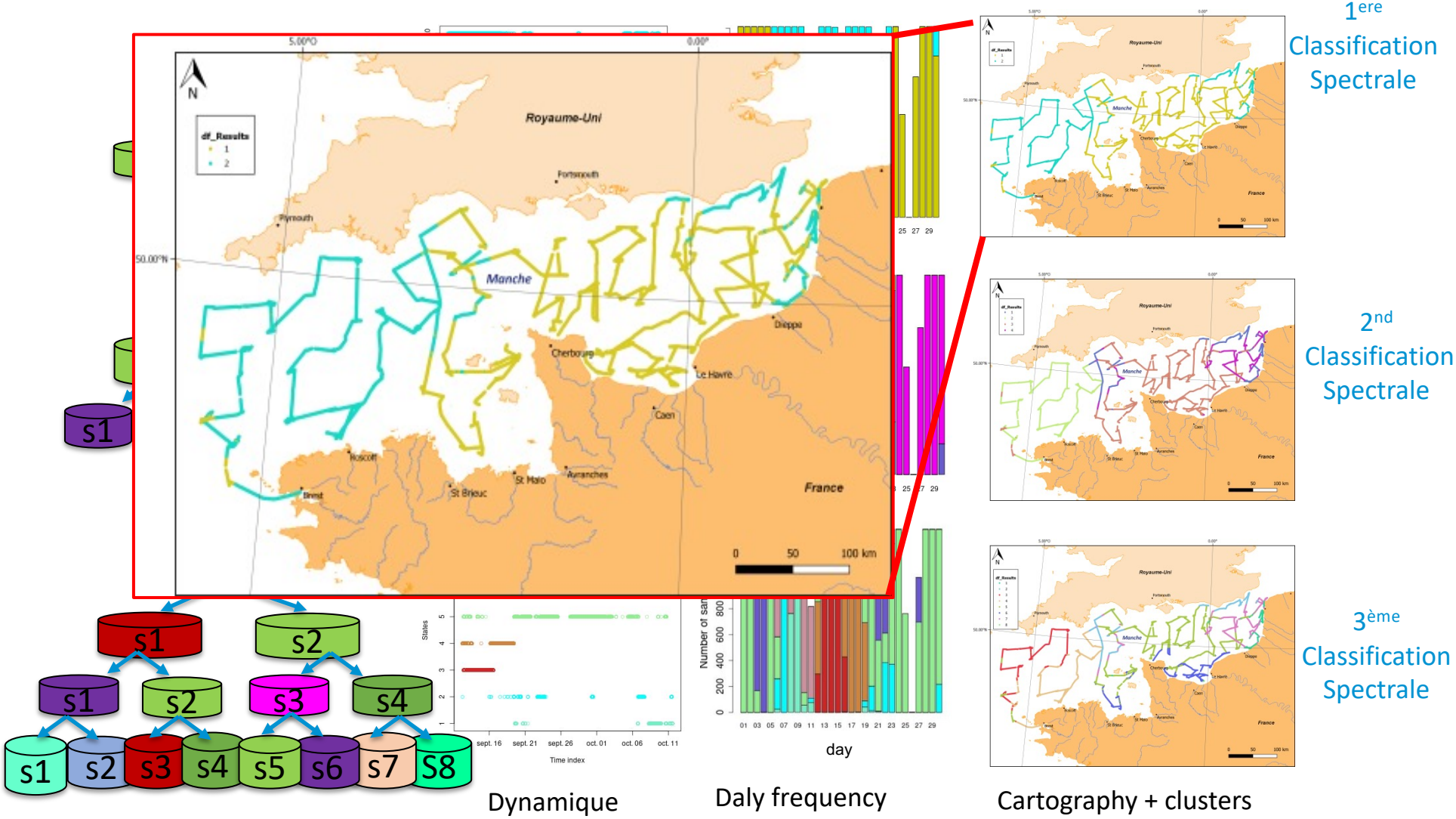
3ème Classification Spectrale

Dynamique

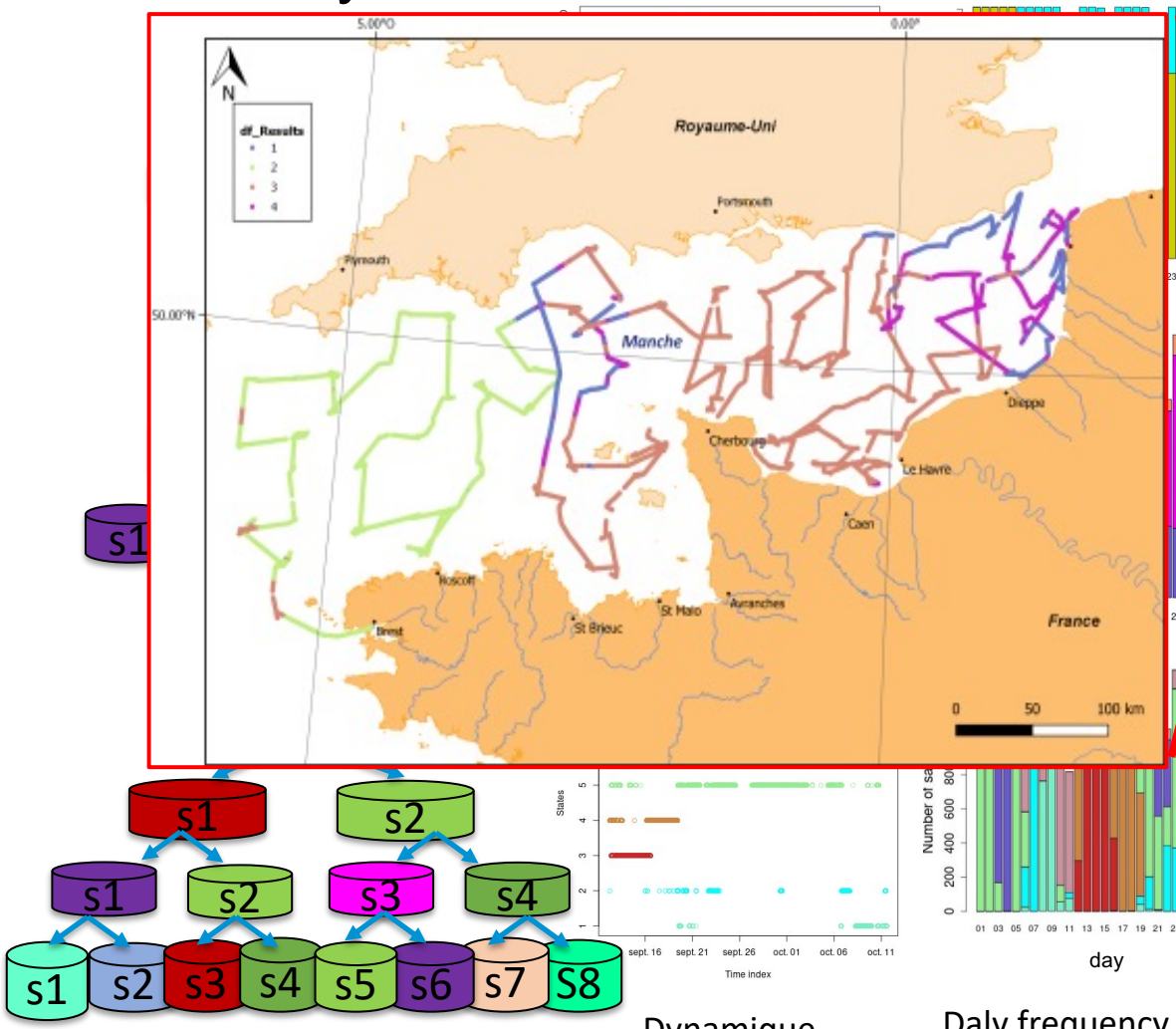
Daly frequency

Cartography + clusters

→ a distinction between the two Channel basins

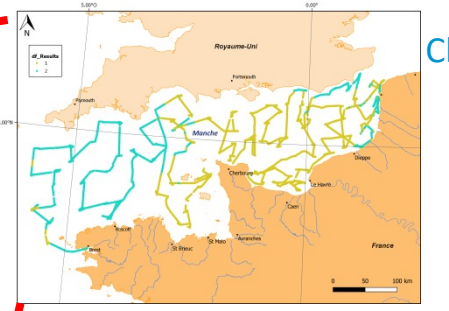


→ clearly identifies 3 water bodies: offshore water body, water body in the central, a water body further west

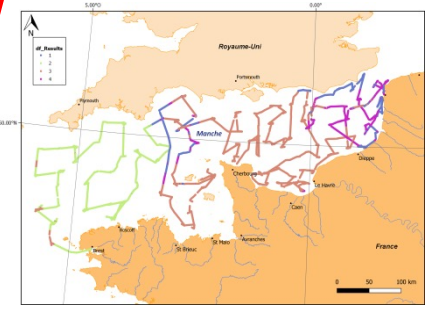


Dynamique

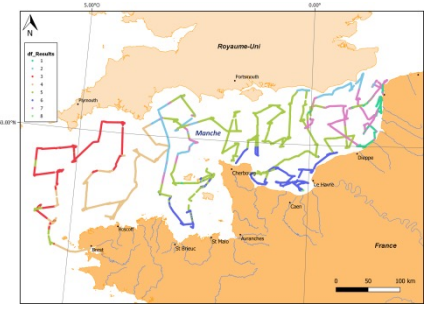
Daly frequency



1ere Classification Spectrale



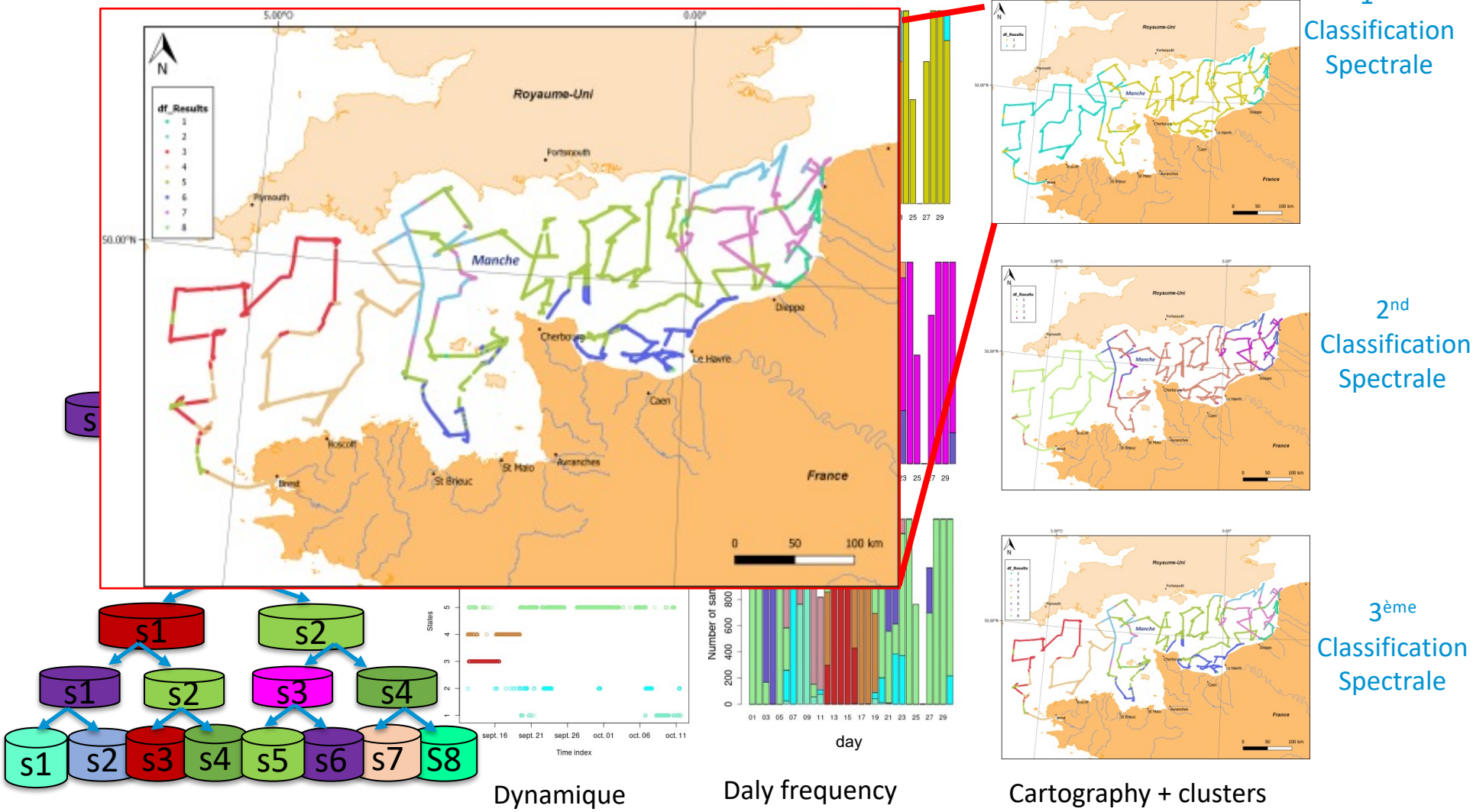
2nd Classification Spectrale



3ème Classification Spectrale

Cartography + clusters

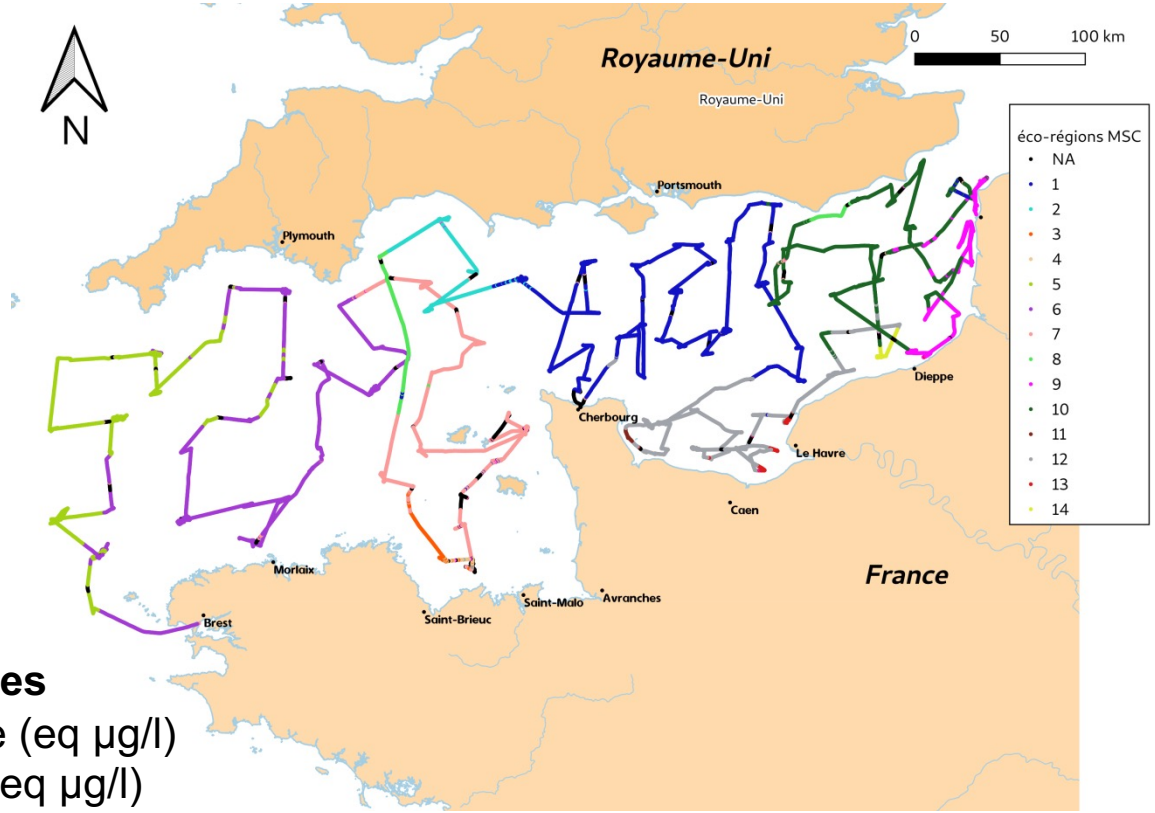
→ highlights 8 areas with an environmental difference



Analysis: Eco-hydro-regions vs Seascapes

Environmental variables :

- Temperature (°C)
- Salinity
- Turbidity
- Oxygen (µmol/l)

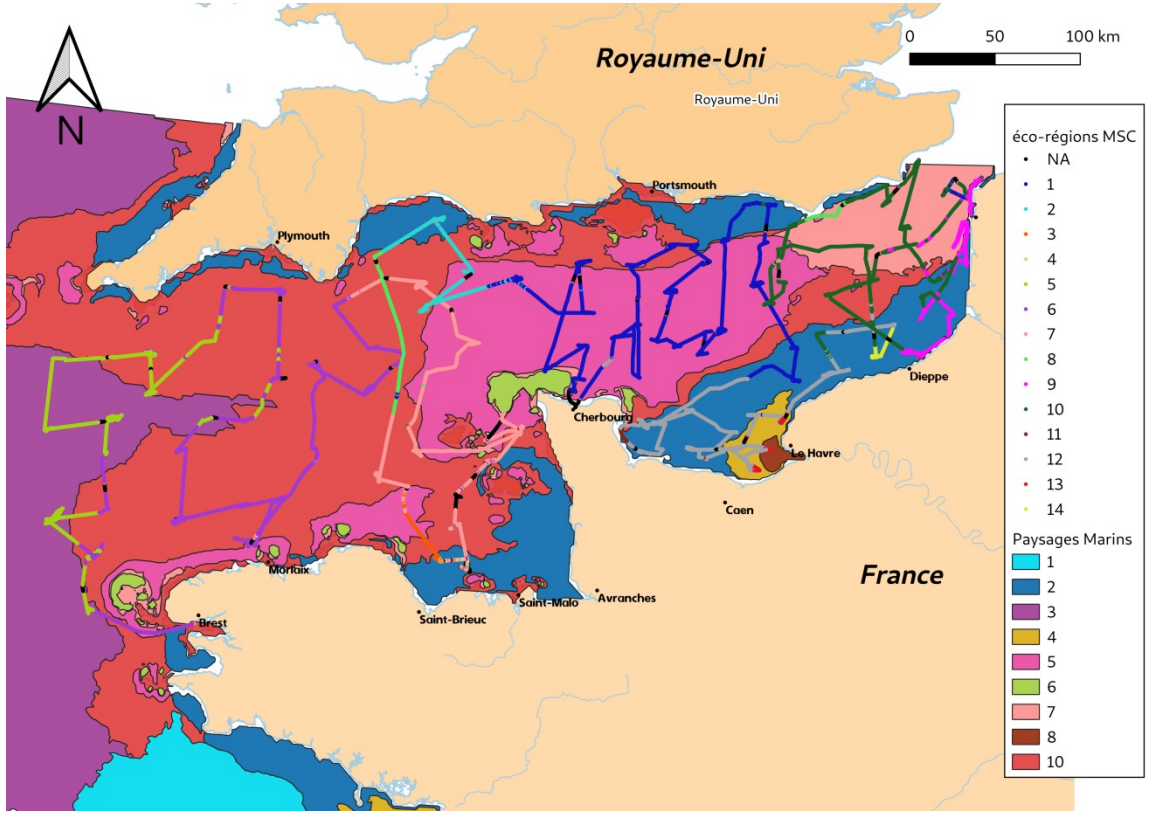


Biological variables

- AOA Green Algae (eq µg/l)
- AOA Blue Algae (eq µg/l)
- AOA Brown Algae (eq µg/l)
- AOA Cryptophyte (eq µg/l)

Weather FORCE

Analysis: Eco-hydro-regions vs Seascapes



Sources: Seascapes: DCSMM D7 teams, SHOM

Weather FORCE

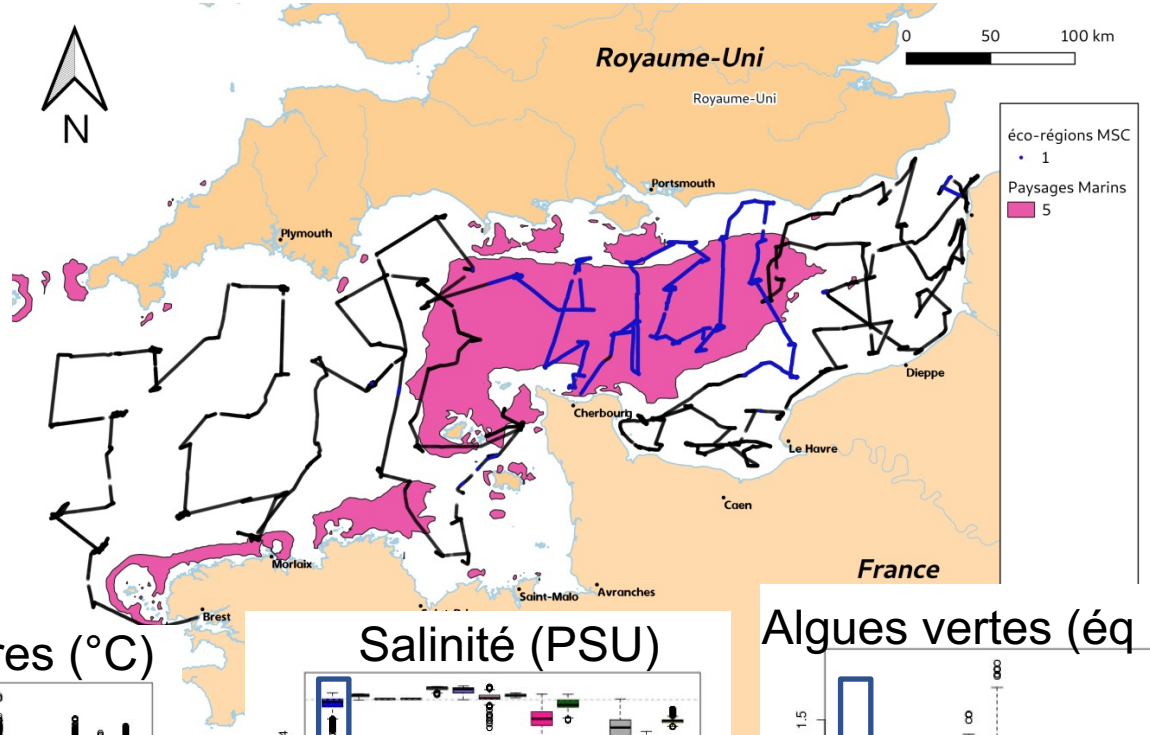
Analysis: Eco-hydro-regions vs Seascapes

→ Classe 1 – PM5 : mixed waters under tidal influence

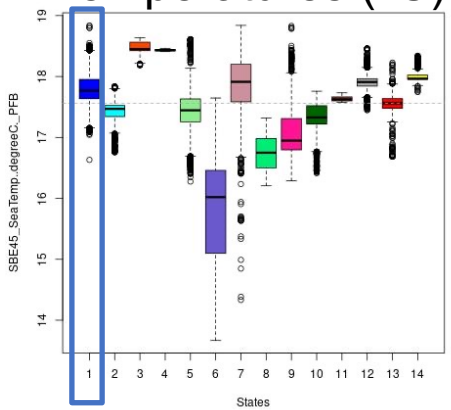
- above average temperatures and salinities

- high concentration of green algae

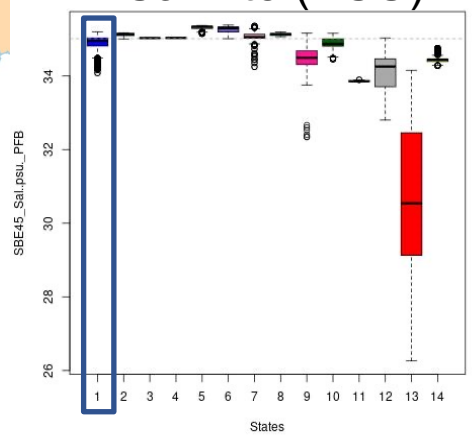
- consistent with the characteristics of the mixed waters of the center.



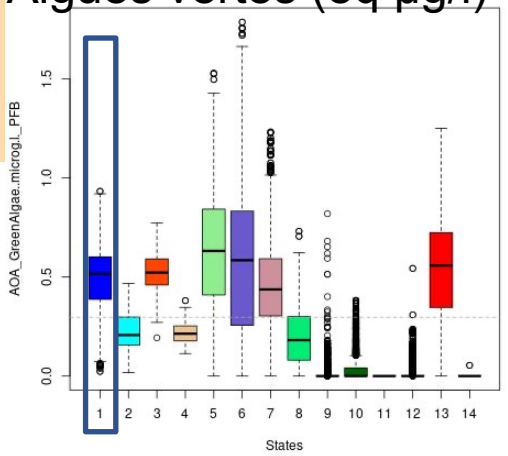
Températures (°C)



Salinité (PSU)



Algues vertes (éq µg/l)



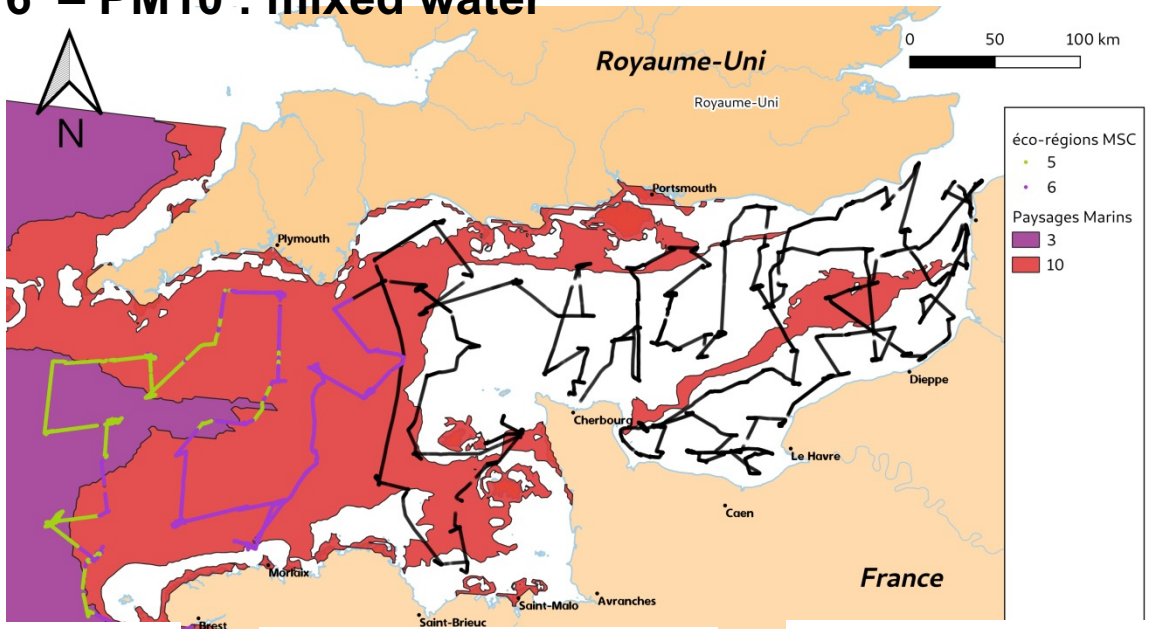
Analysis: Eco-hydro-regions vs Seascapes

→ Classe 5 – PM3 : offshore waters with seasonal stratification
→ Classe 6 – PM10 : mixed water

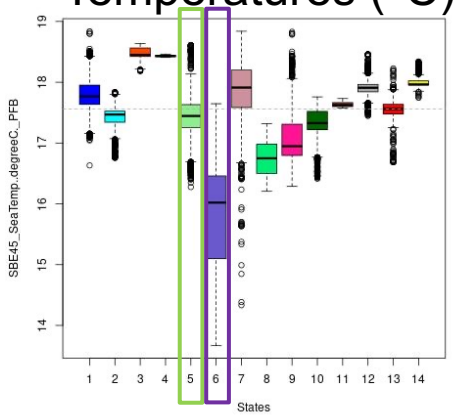
- a dominance of green and brown algae

- representative of the zones of the open sea

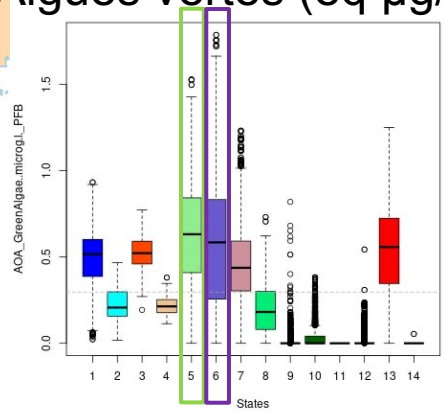
- difference in temperature



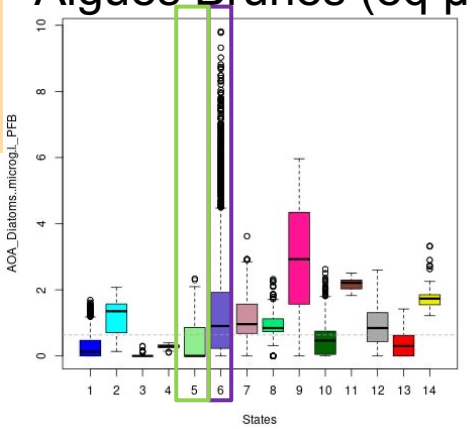
Températures (°C)



Algues vertes (éq µg/l)



Algues Brunes (éq µg/l)



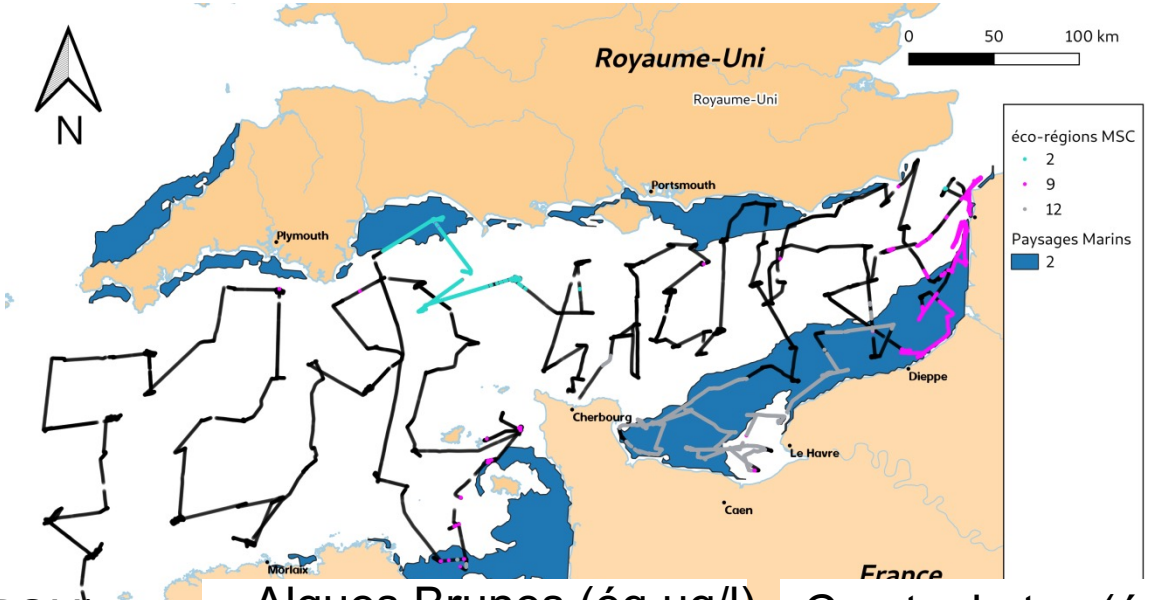
Weather FORCE

Analysis: Eco-hydro-regions vs Seascapes

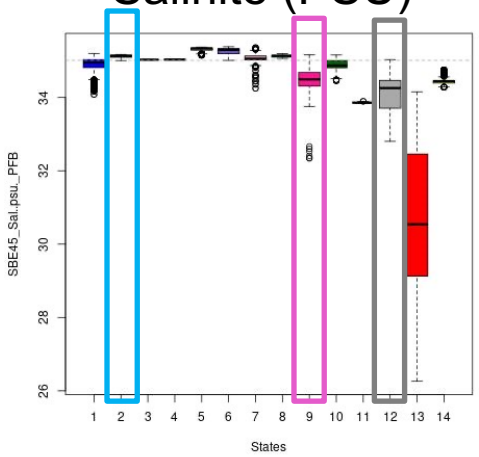
→ **Classe 2-9-12 – PM2 : coastal and shallow waters under the influence of freshwater**

3 coastal ecoregions:
- the English coasts (cl2),
- the Seine Bay (cl9)
- the Somme Bay (cl12)

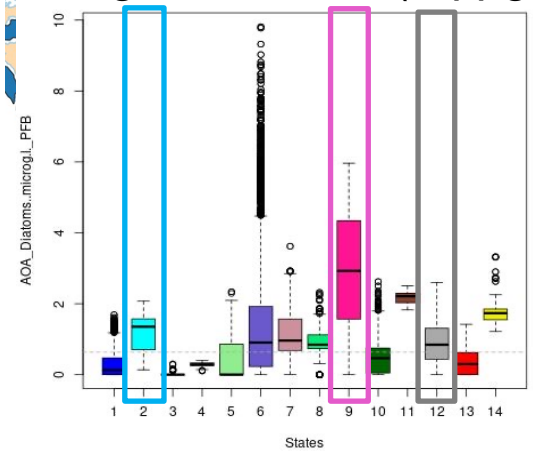
with a variation in salinity (symbol of freshwater input) and a difference in phytoplanktonic composition characteristic of the area.



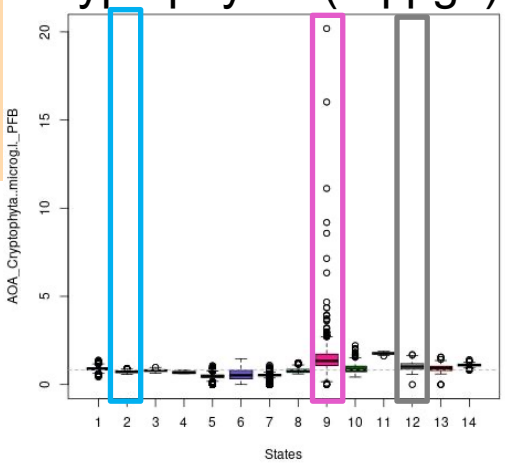
Salinité (PSU)



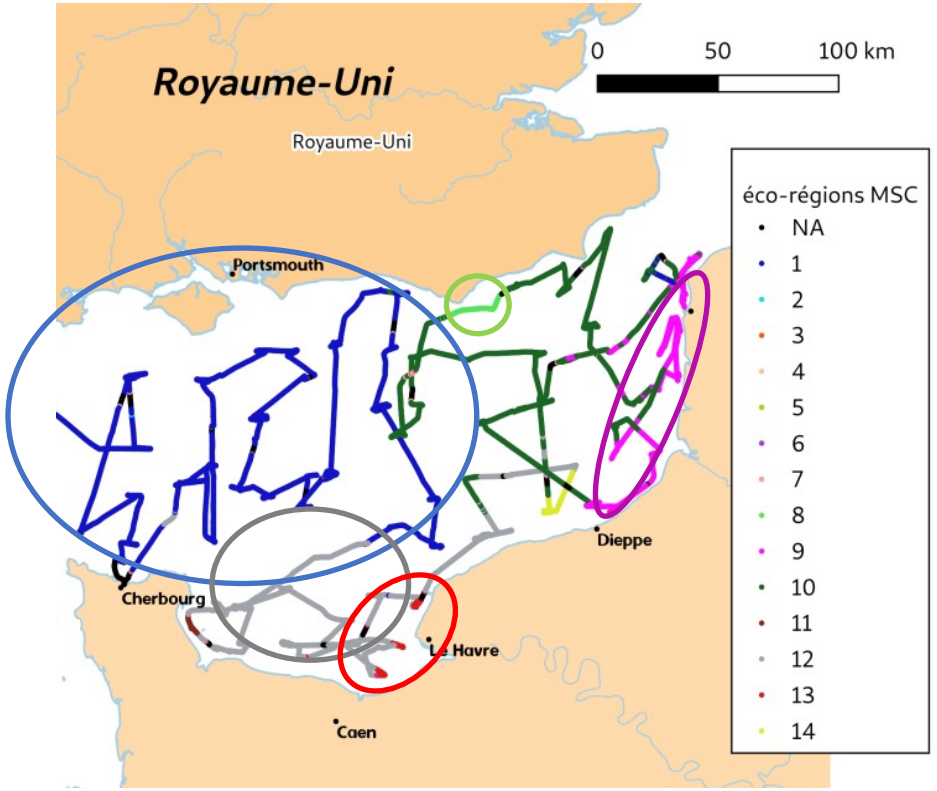
Algues Brunes (éq µg/l)

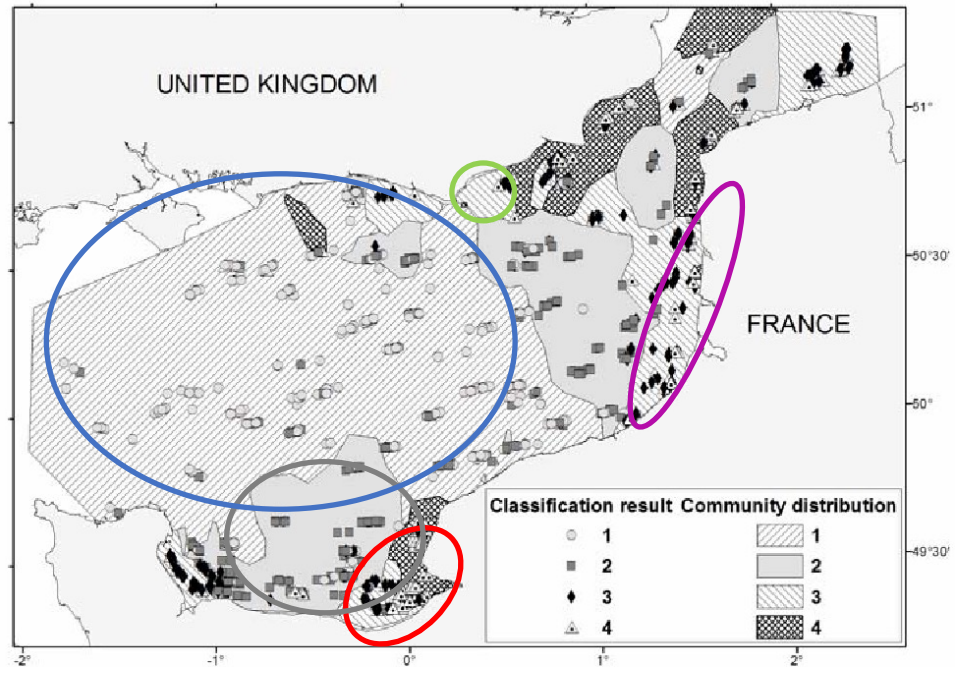


Cryptophytes (éq µg/l)



Analysis: eco-hydro-regions vs. fish community distribution

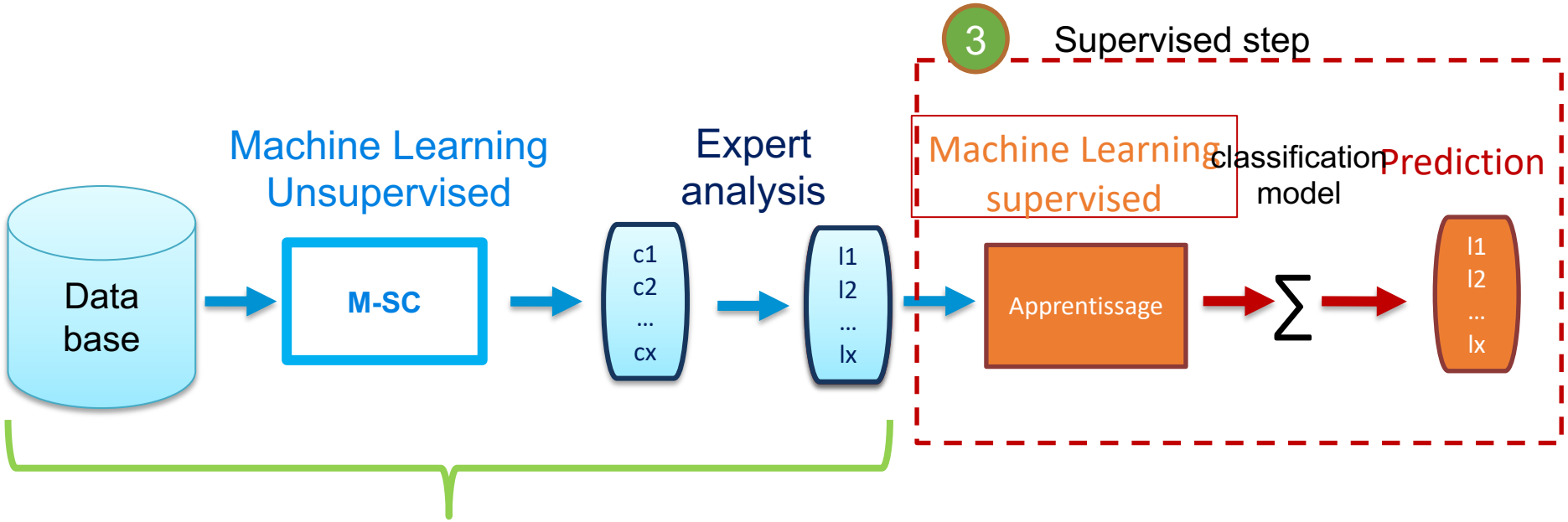




Source : Vaz et al, 2007. Eastern English Channel fish assemblages: measuring the structuring effect of habitats on distinct sub-communities

Weather FORCE

Towards a system to help recognize environmental conditions

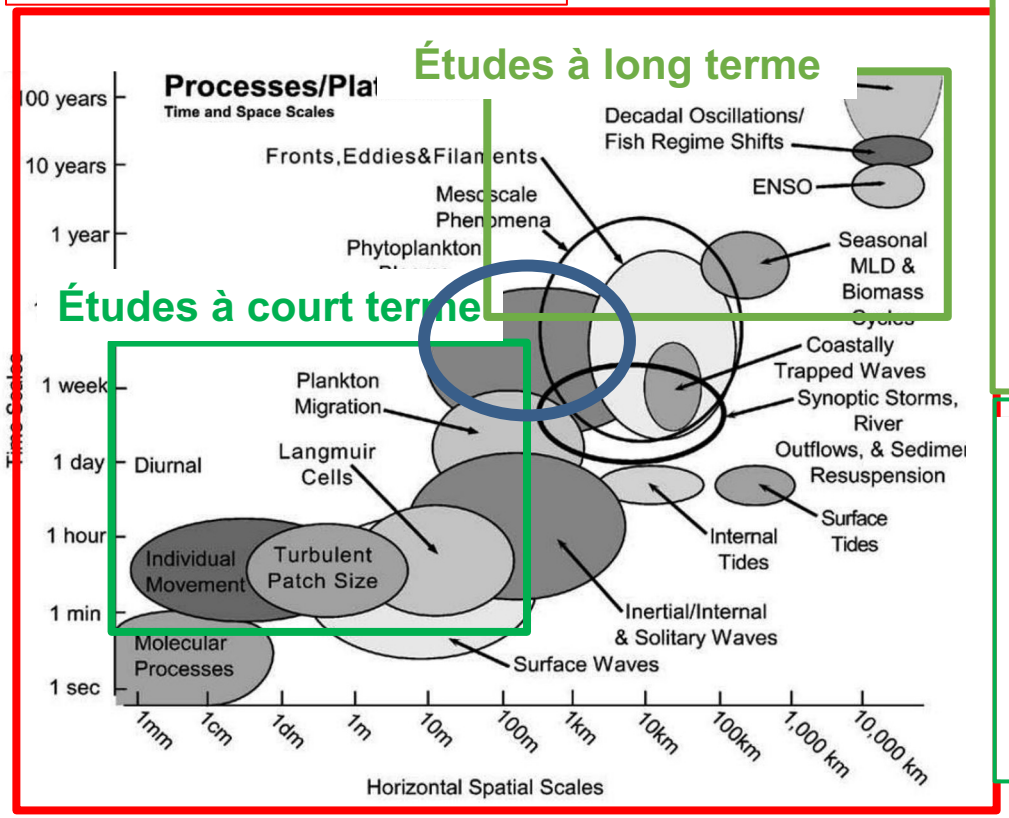


→ Building a learning model

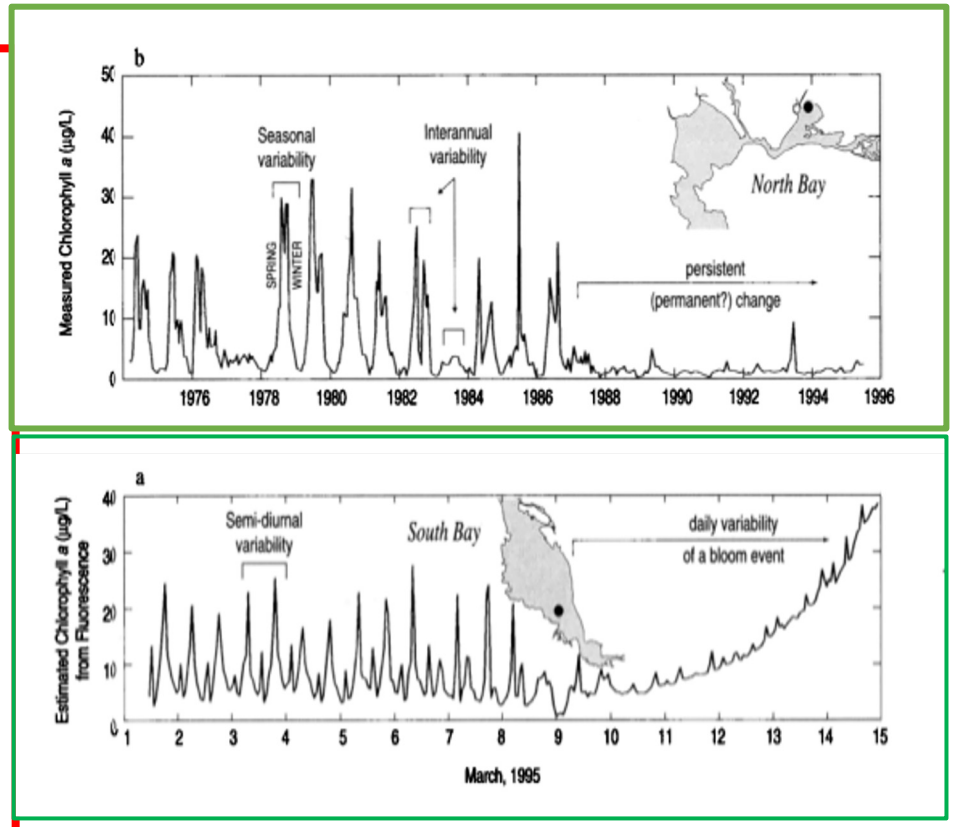
Thank for ours attention

→ Détection des variations multi-échelles

Approche intégrée



(Source : Dickey, 2003)

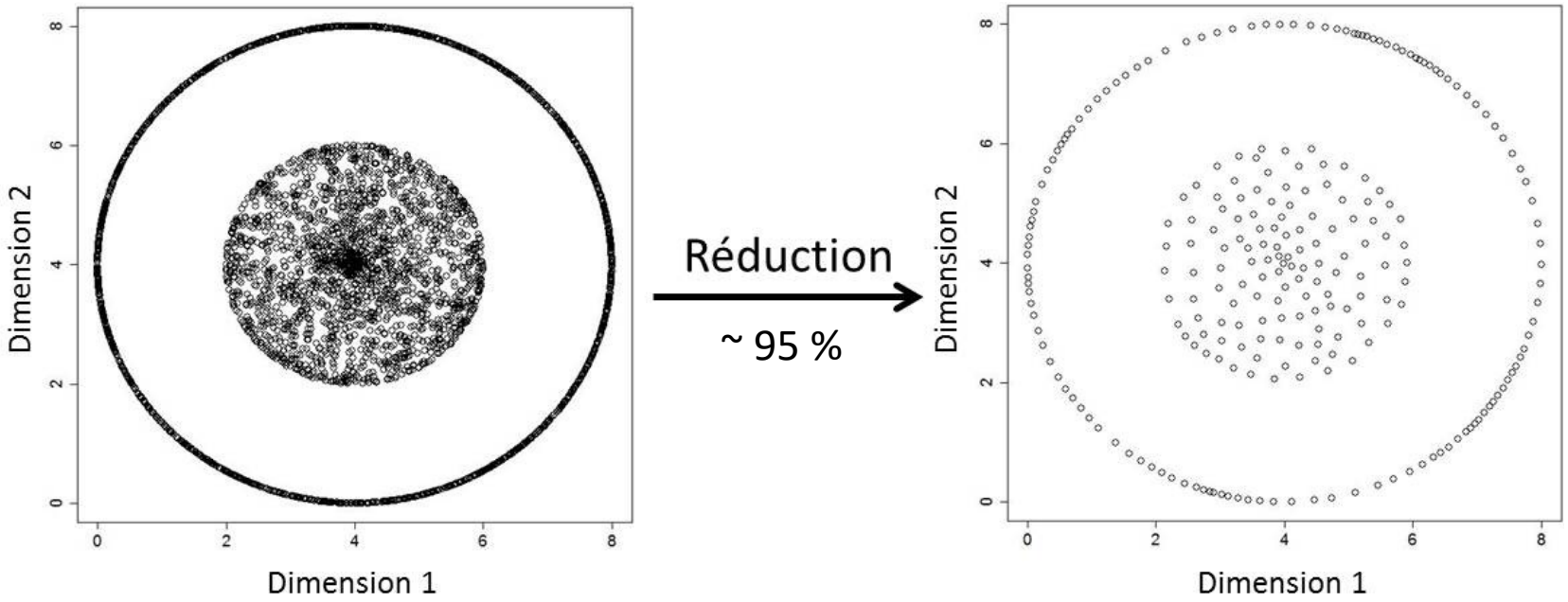


(Source : Cloern, 1996)

Weather FORCE

reduction step :
selected representative dataset by a vectors quantification of space : use K-means and Elbow criterion

- Reduction 4 000 to 260 points
  - Elbow criterion:
    - 95 % of variance explaned



K-means (KM)

Classification Hiérarchique (HC)

Classification Spectrale (SC)