Utilisation du Machine Learning pour la caractérisation et la prédiction des blooms phytoplanctoniques (y compris les HAB) et des états environnementaux associés.

Lefebvre A.[1], Wacquet G.[1], Halawi Ghosn R.[1], Poisson-Caillault E.[2]

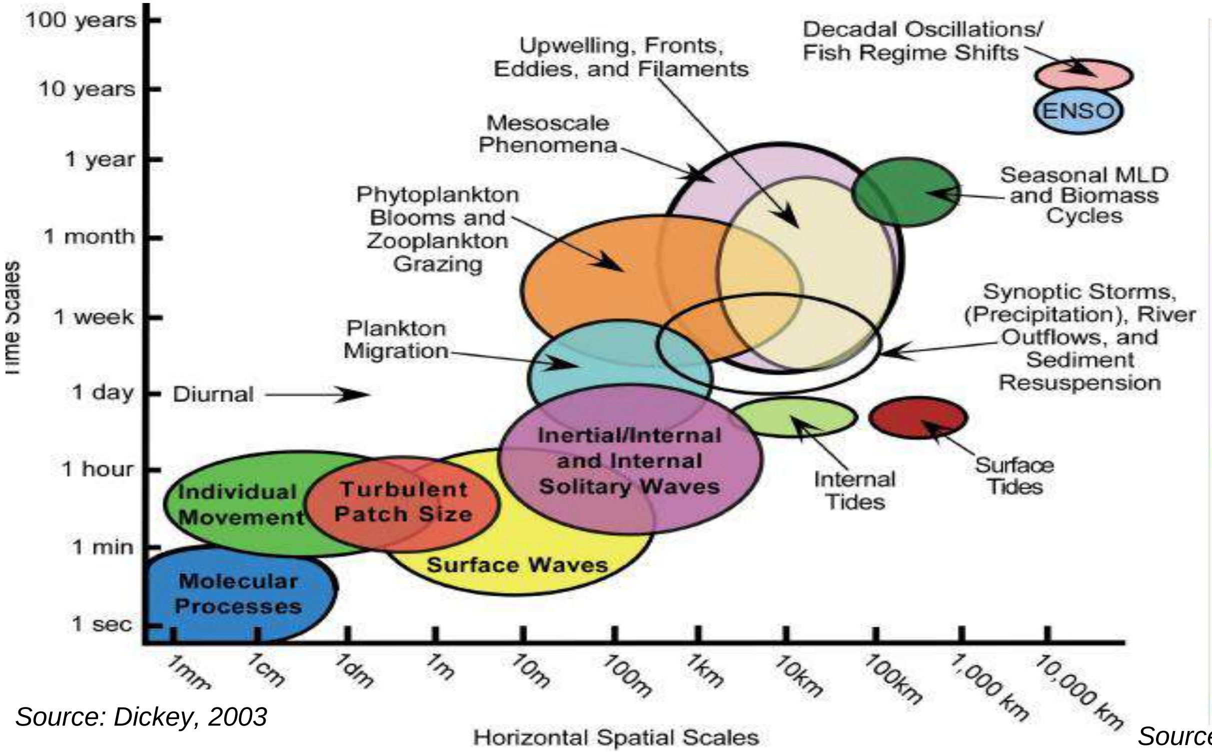
1- Ifremer, 2 - ULCO/LISIC

Atelier IA/ML, CS LEFE/CYBER – IR ILICO

4-5 juin 2024

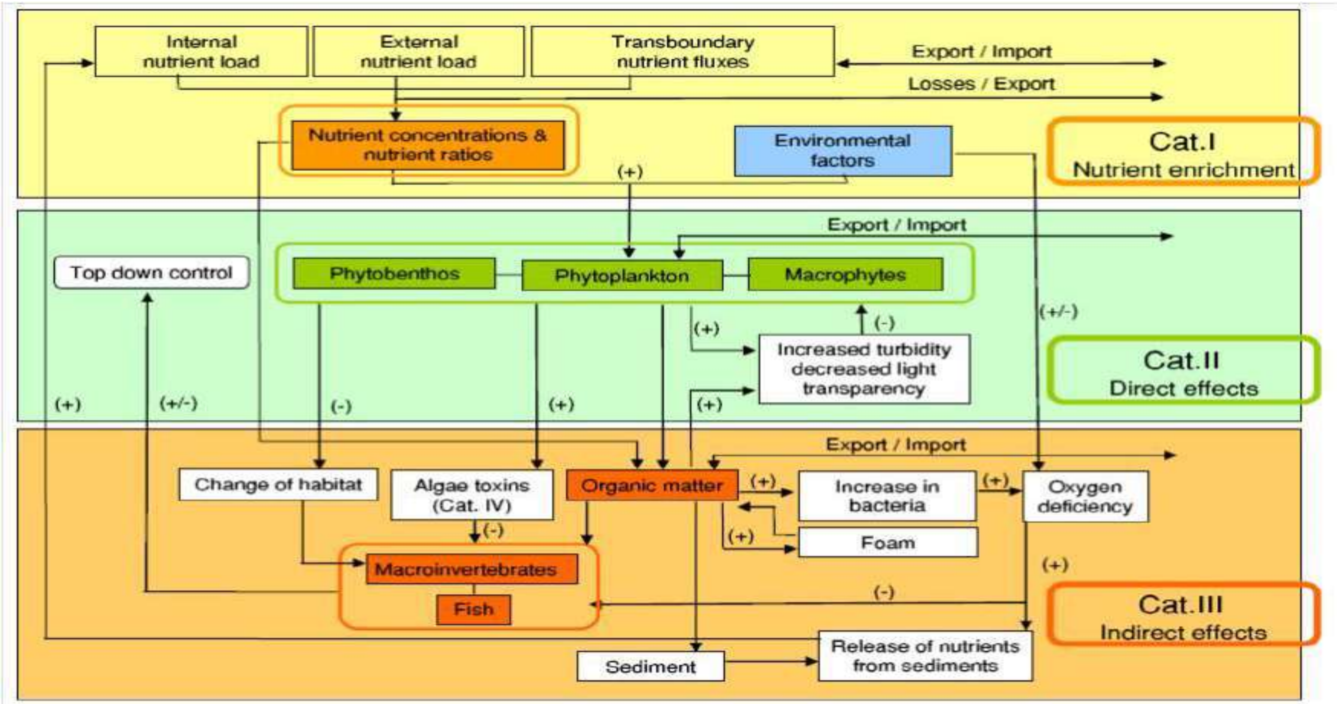
Long Term, High-Resolution and multi-parameter approach

Spatial and temporal scales involved
during phytoplankton blooms

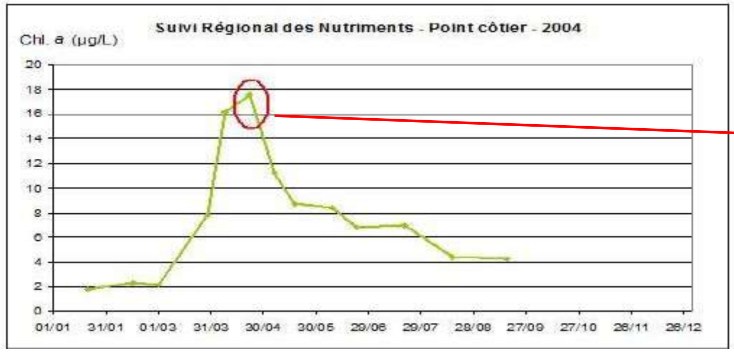


Source: Dickey, 2003

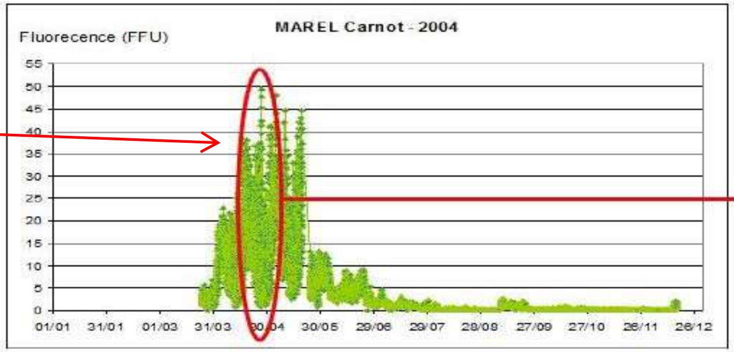
Interactions of the differents elements involved
in the eutrophication process

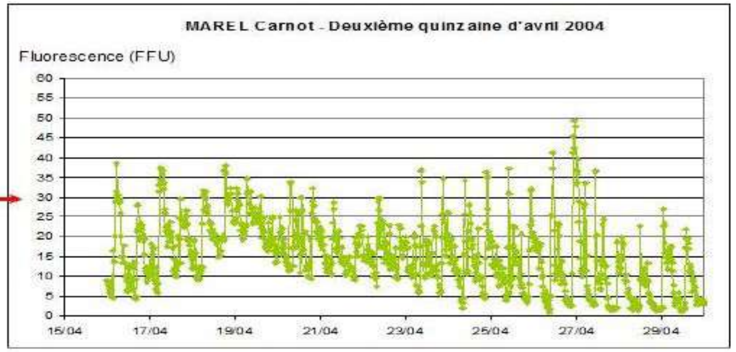


Source: Claussen U. (stagesproject.eu)



Low frequency

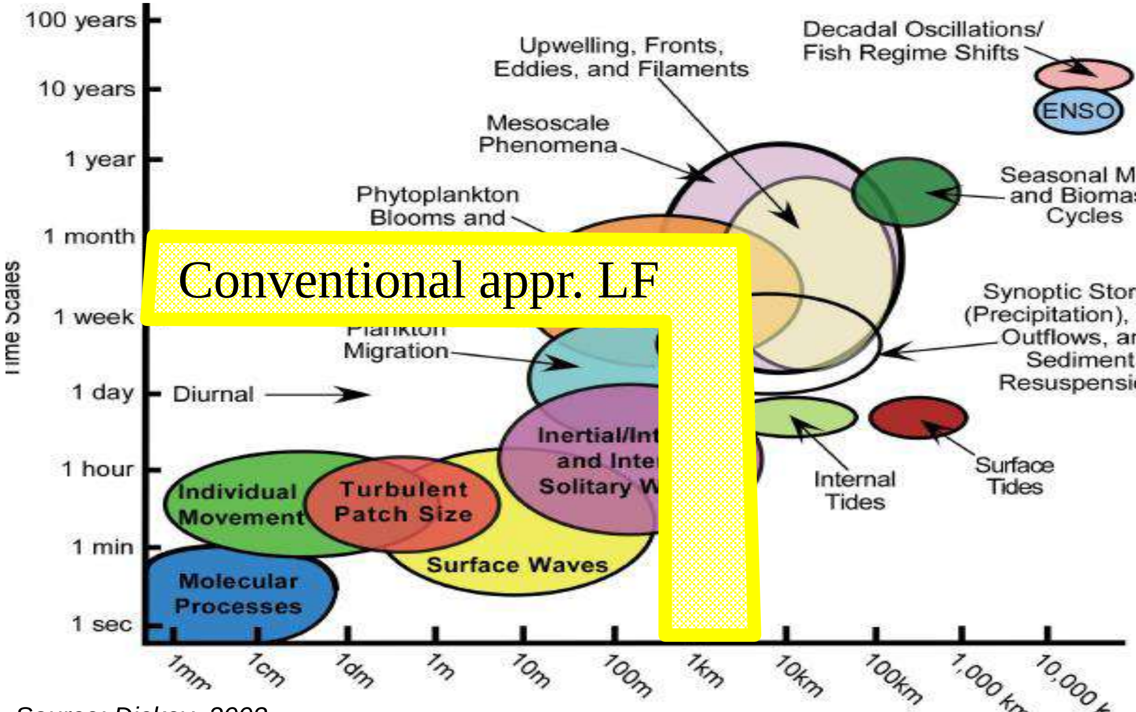




High frequency

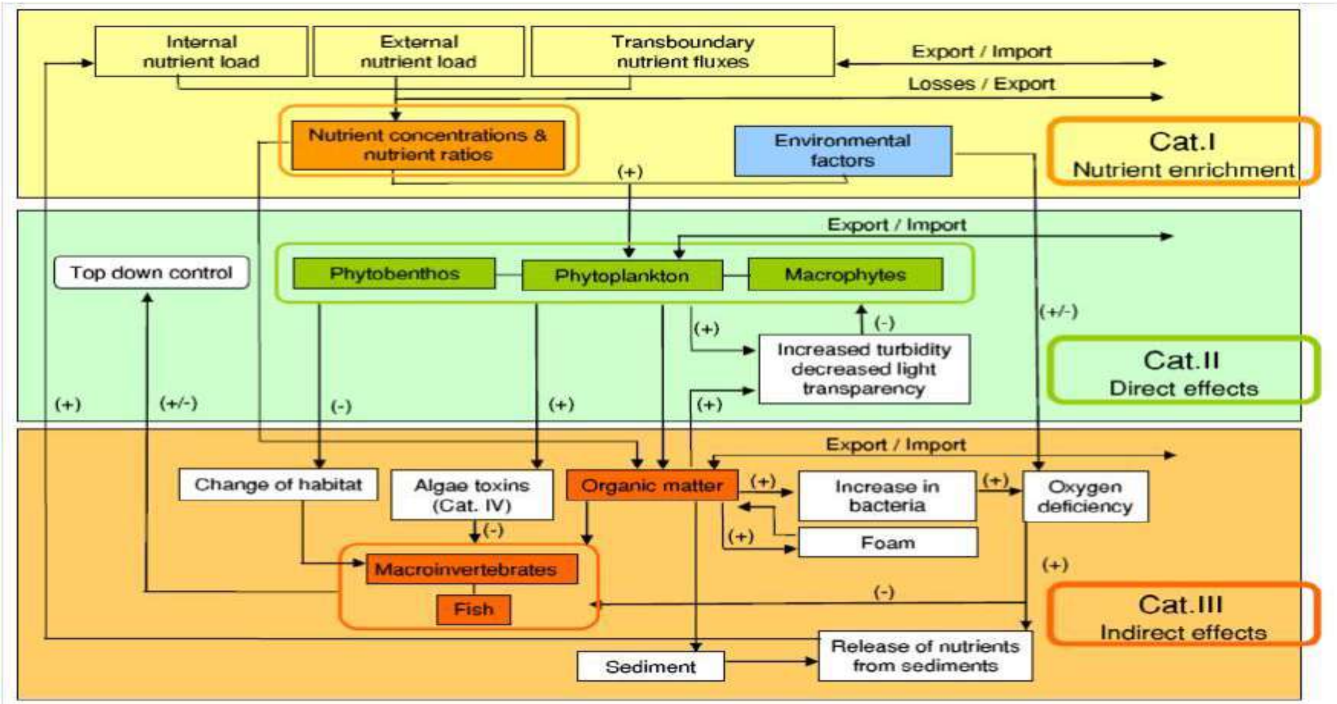
Long Term, High-Resolution and multi-parameter approach

Spatial and temporal scales involved during phytoplankton blooms

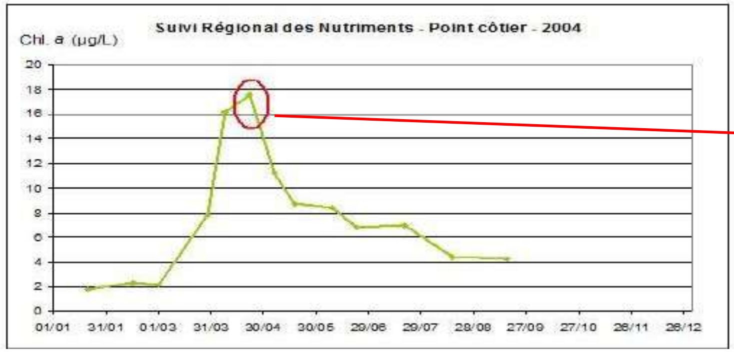


Source: Dickey, 2003

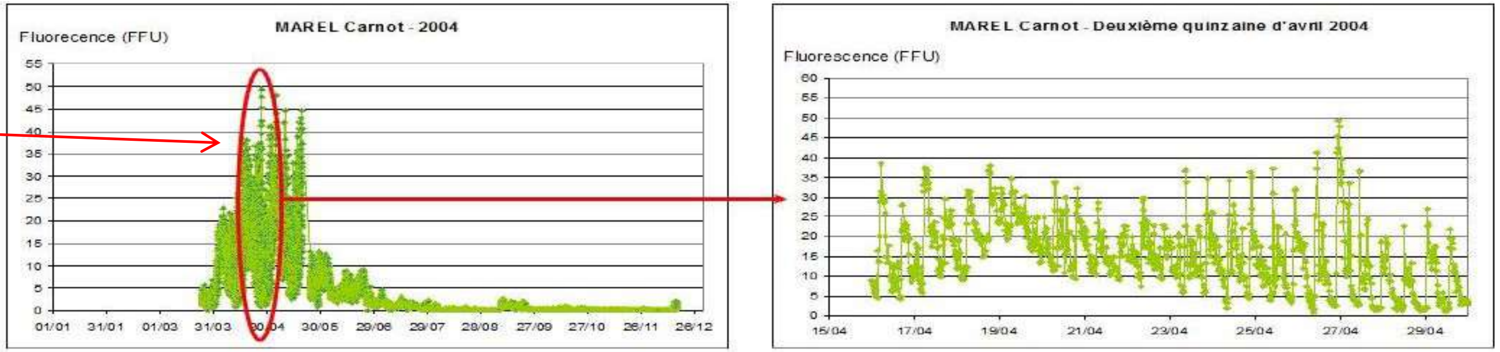
Interactions of the differents elements involved in the eutrophication process



Source: Claussen U. (stagesproject.eu)



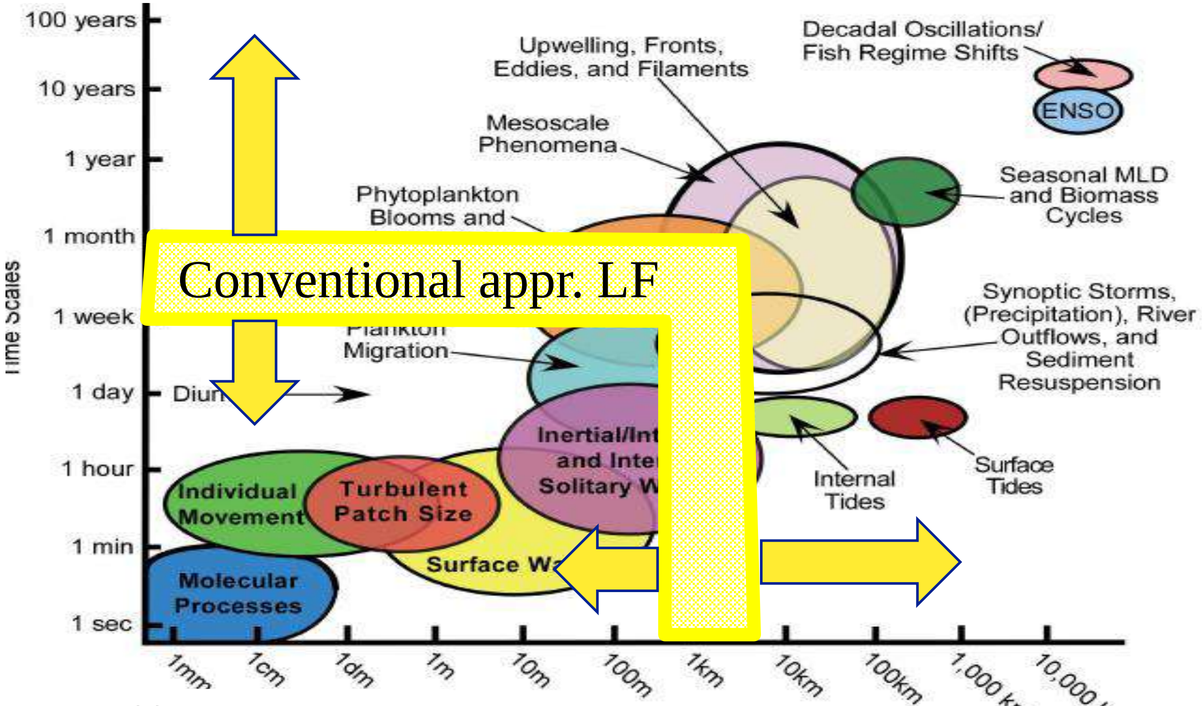
Low frequency



High frequency

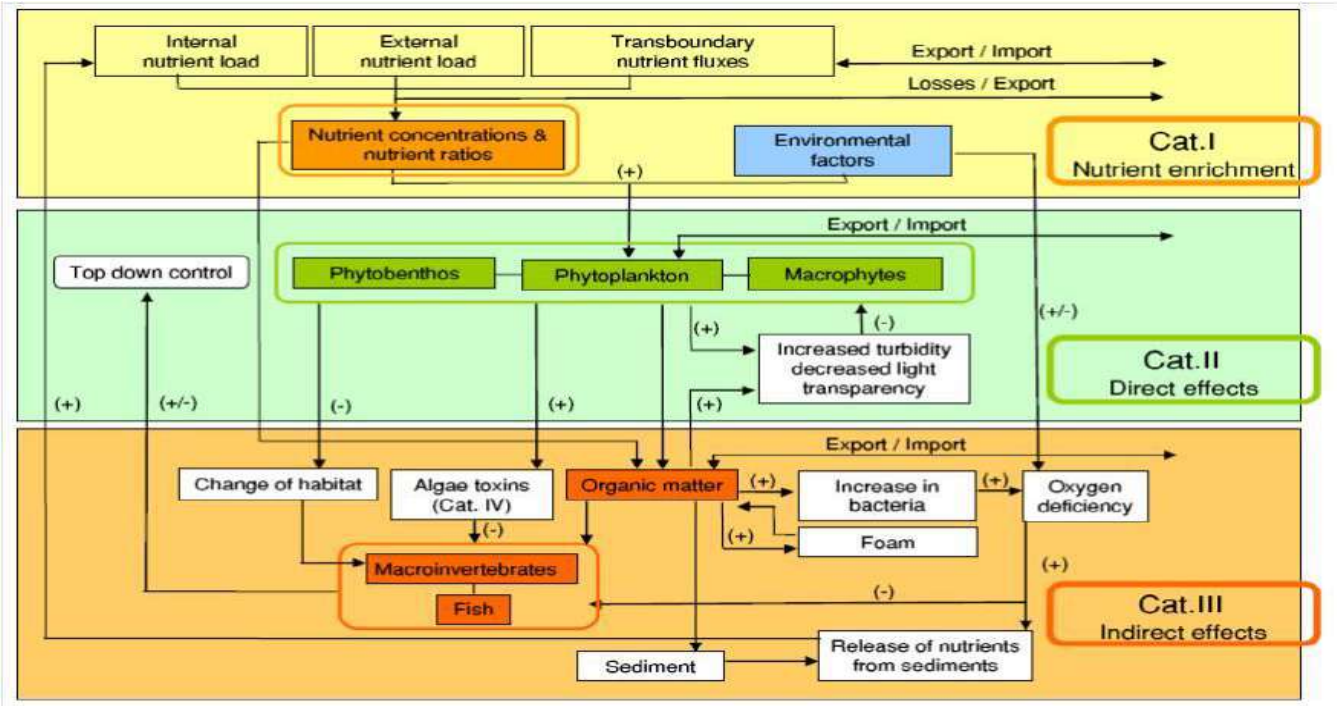
Long Term, High-Resolution and multi-parameter approach

Spatial and temporal scales involved during phytoplankton blooms

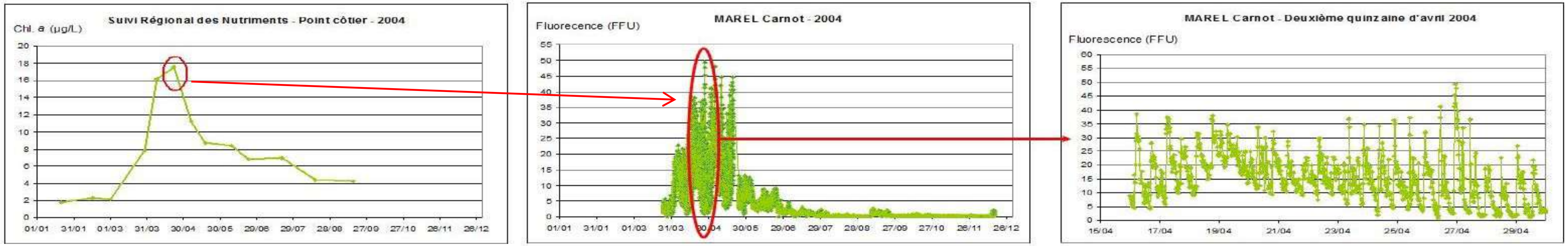


Source: Dickey, 2003

Interactions of the differents elements involved in the eutrophication process



Source: Claussen U. (stagesproject.eu)

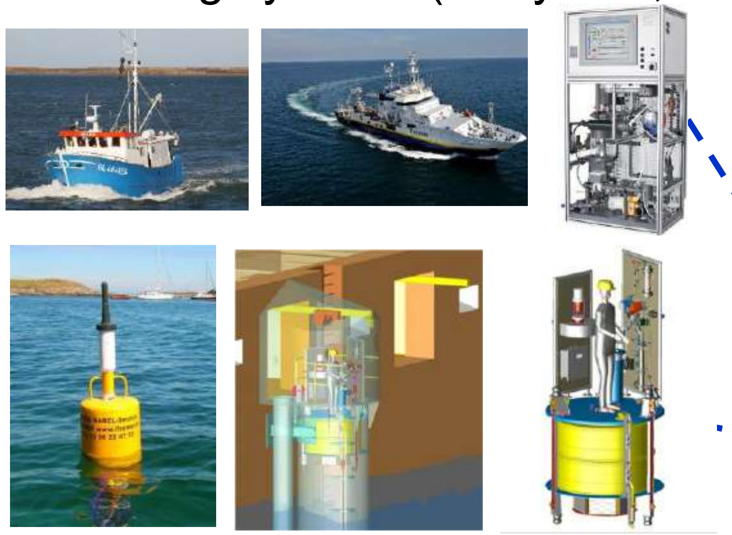


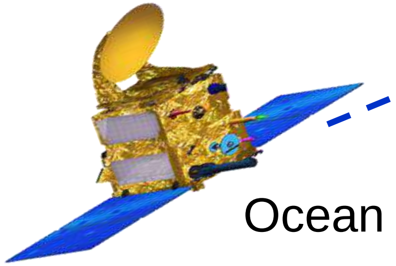
Low frequency

High frequency

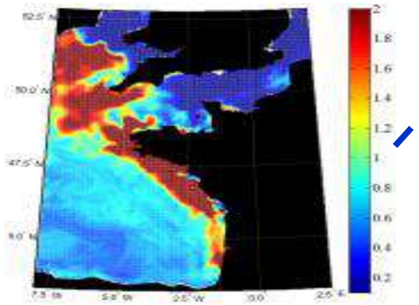
Integrated Observation

Data flow from Low to High Resolution monitoring systems (Ferry Box, buoys,…)





Ocean Color



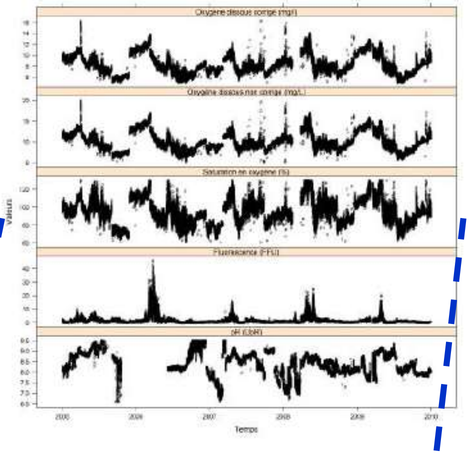
Modelling

Data Base

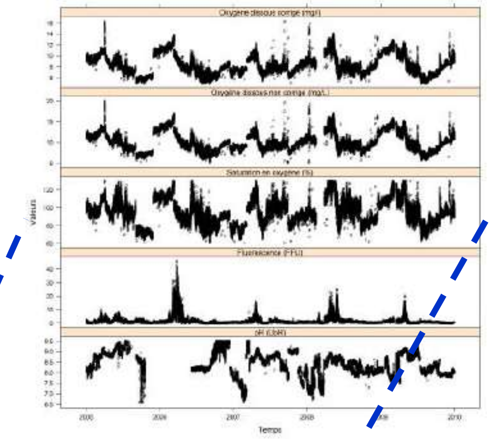
DATA TERRA

ODATIS

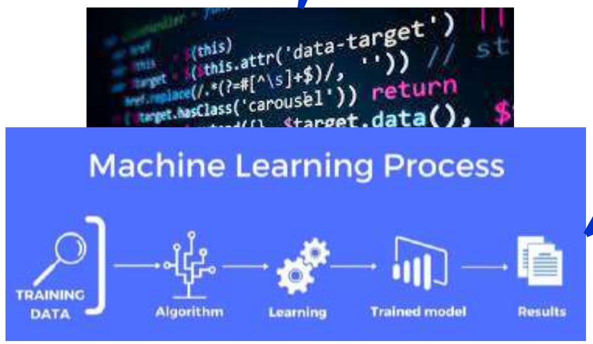
Other DB (EU)

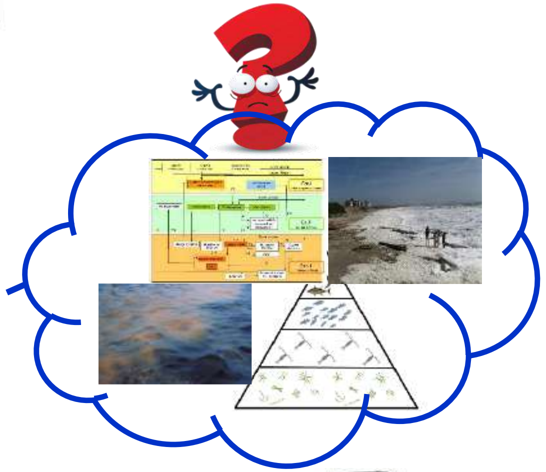


Raw Data



Processed Data (incl. QA/QC)





Huge amount of data!
Data Quality?
Missing data? Completion?
Regularization of time series?
Optimal Information from HF data?
Modelling? Forecasting?

CONTEXT

Bloom of Phaeocystis globosa in the English Channel

General context

- Geographical location

Channel / Strait of Dover / North Sea

Cross-border aspects (Belgium, U.K.)

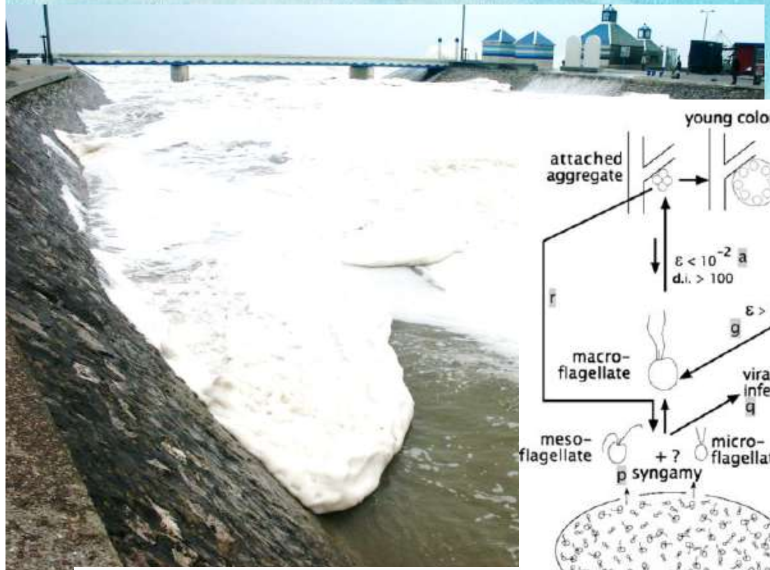
- Attendance, multiple activities, issues

- Pressures

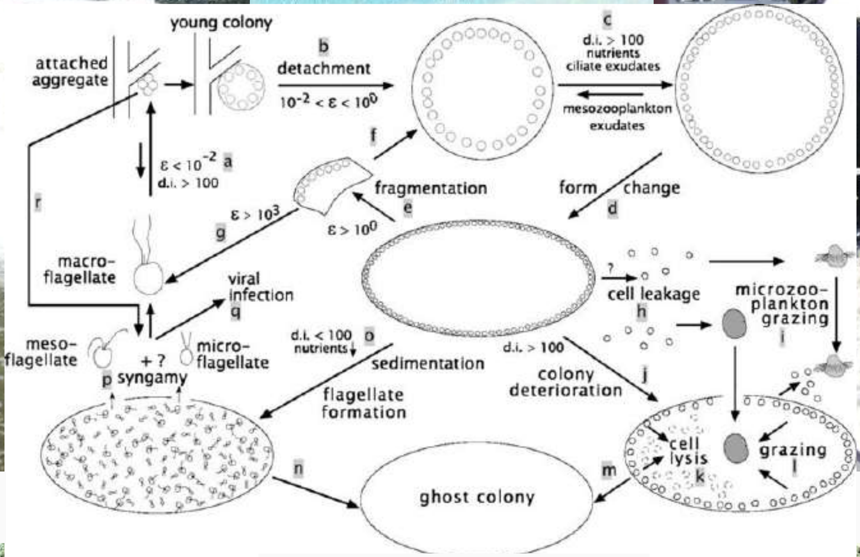
Nutrient inputs

- Recurring bloom of *Phaeocystis sp*

CROSS-BORDER SITE STRONGLY IMPACTED BY ANTHROPIC ACTIVITIES



River « Le Wimereux »



Complex Life Cycle

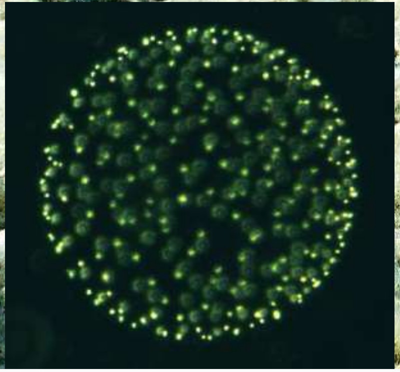
Boulogne-sur-Mer Harbor

HAB with high biomass

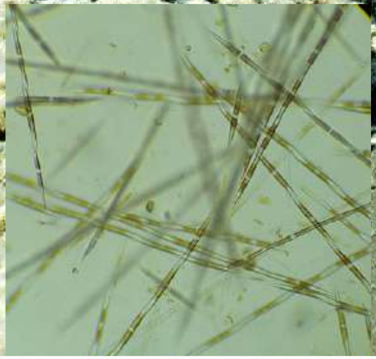
HAB with phycotoxine

Threats on pelagic and benthic compartments (biodiversity, food web…)

Human Health



Colony (Microscopy)



Pseudo-nitzschia (Harmful Algae)
Production of phycotoxine

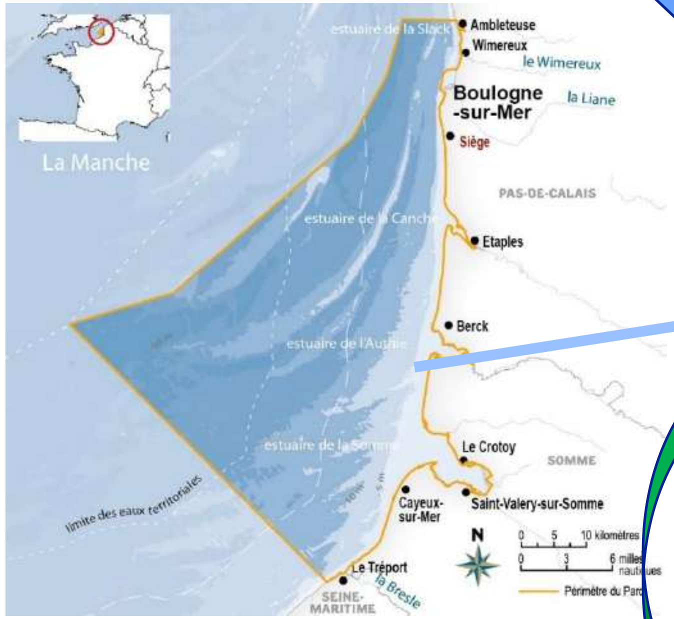
Main considered pressures in the eastern English Channel – North Sea Ecosystem

Global Change, Extreme Events

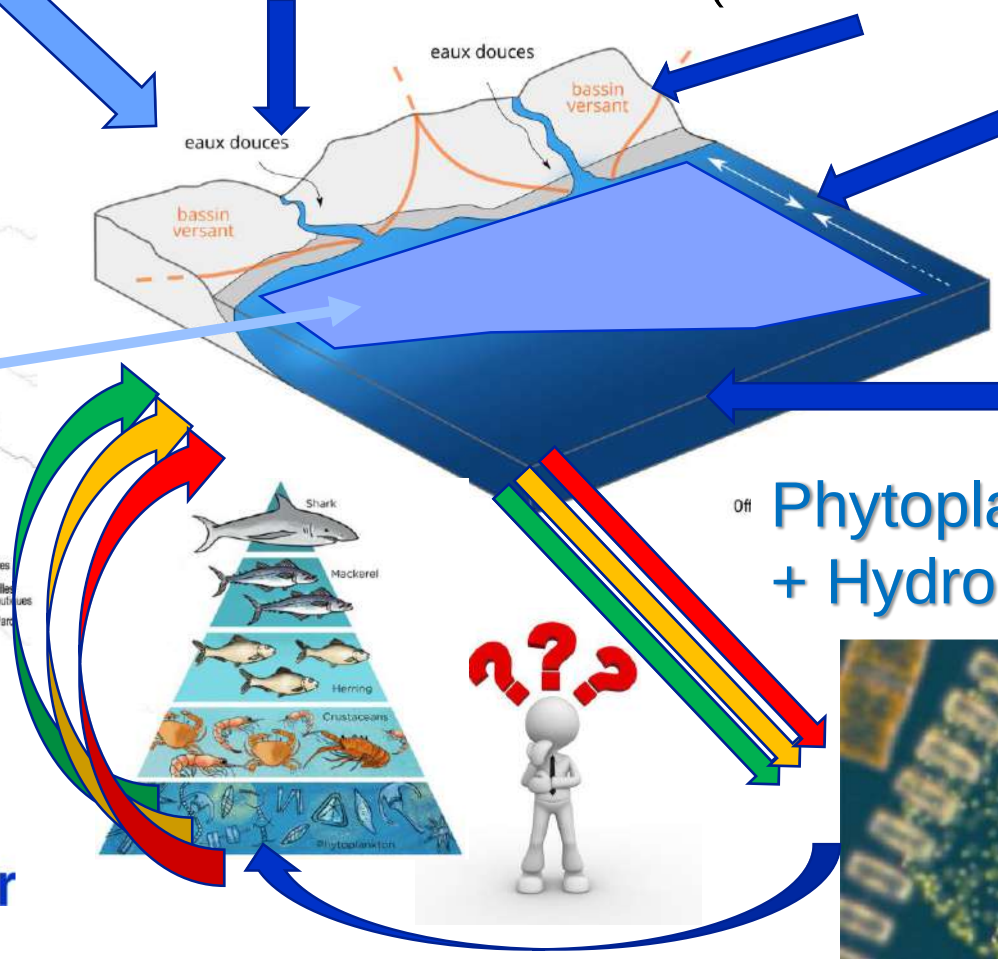
OM, nutrients inputs
Transboundary effects

Nuclear Power Plant
(Gravelines + Normandy)

Aquaculture
Fish farming

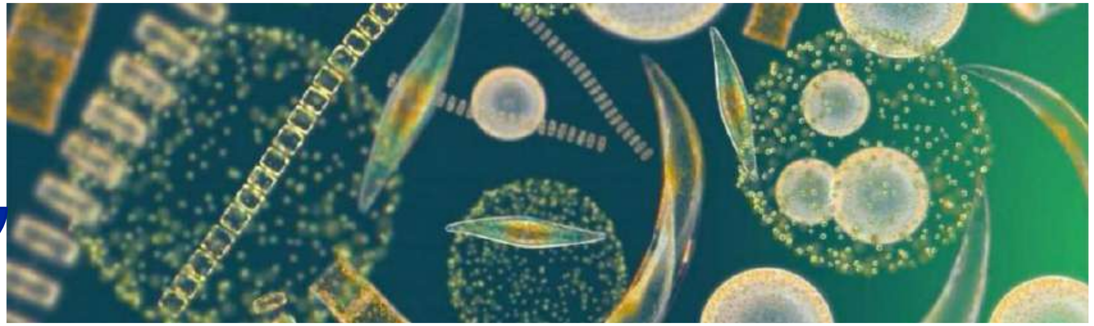


Marine Protected Area

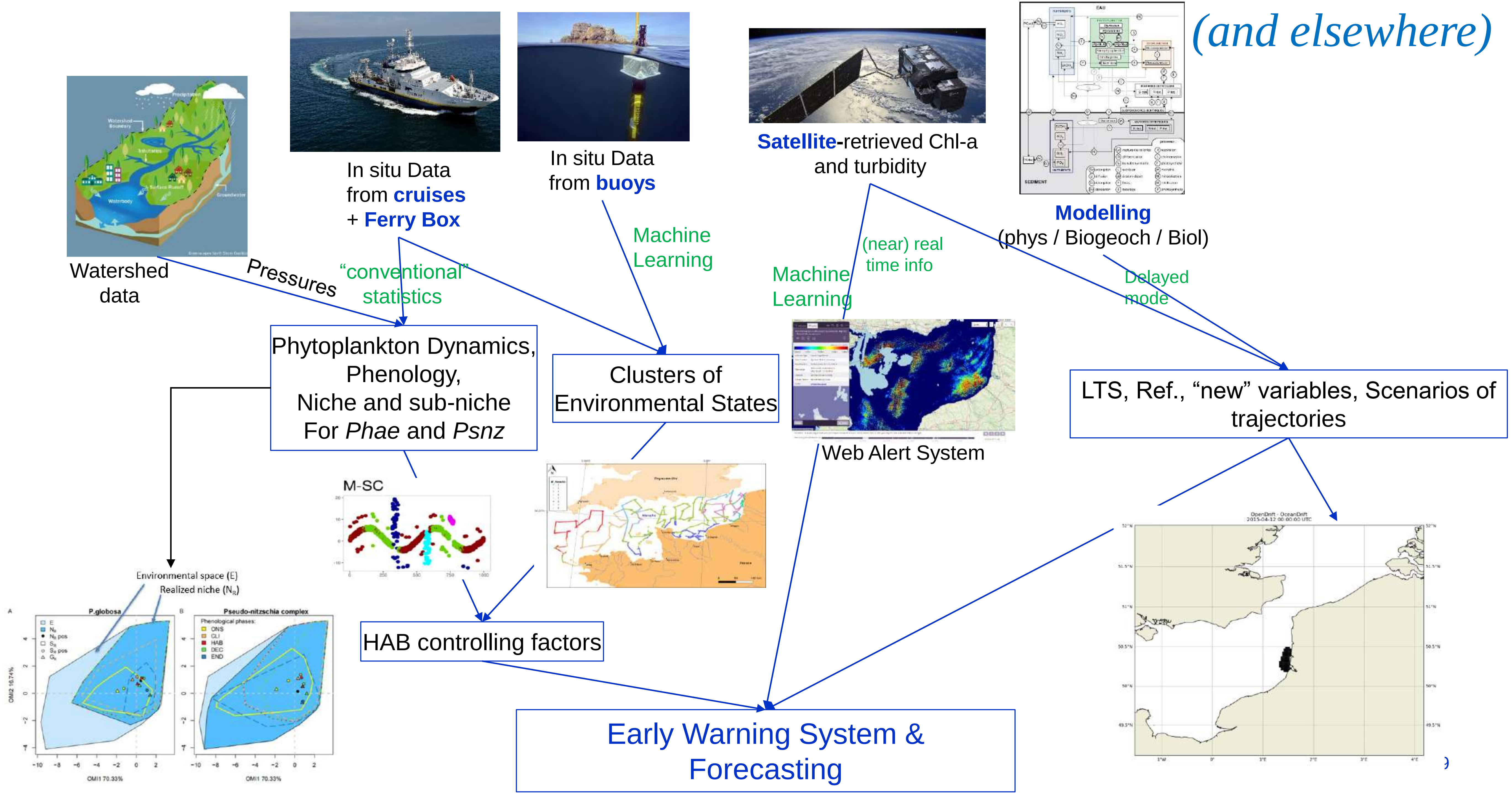


Offshore windfarm

Phytoplankton Biodiversity, Dynamics (incl. HAB)
+ Hydrology + Interactions (biot., abiot., scales)



Web Alert System and Forecasting of HAB in the English Channel (and elsewhere)

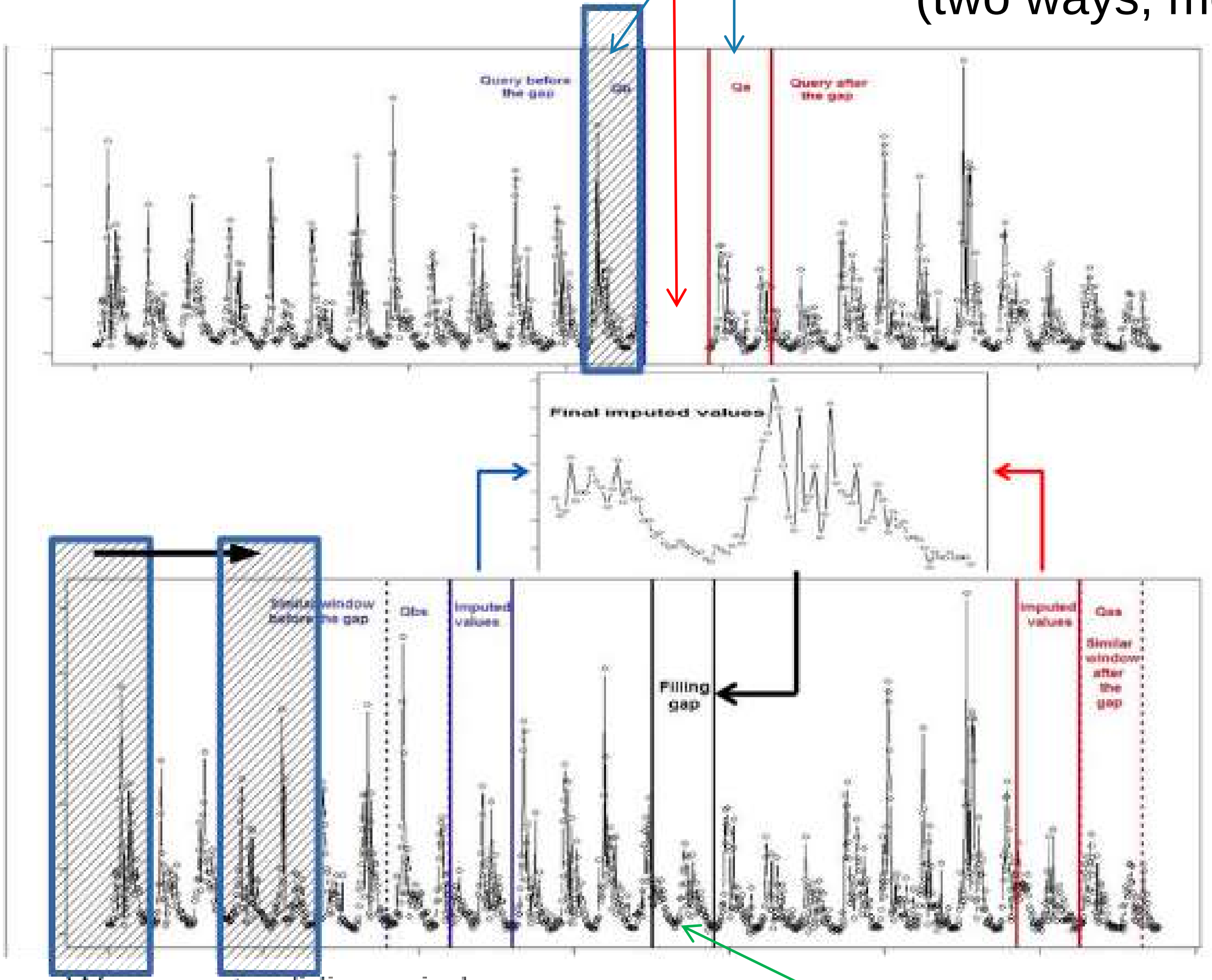


Data completion with Dynamic Time Warping (DTW)

How to fill the gaps ? Missing data: sensor failure, maintenance,…

Query before or after the gap

DTW query building
(two ways; mono/multivariate time series)



W_request : sliding window

Filling the gap

1. Feature extraction from Query Qa/Qb and W_Request.
2. Selection of n W_request that satisfied cosinus criterion.
3. Computation of DTW function on the W_request and selection of a unique Qbs.
4. Direct Imputation or mixing from Qbs and Qba.

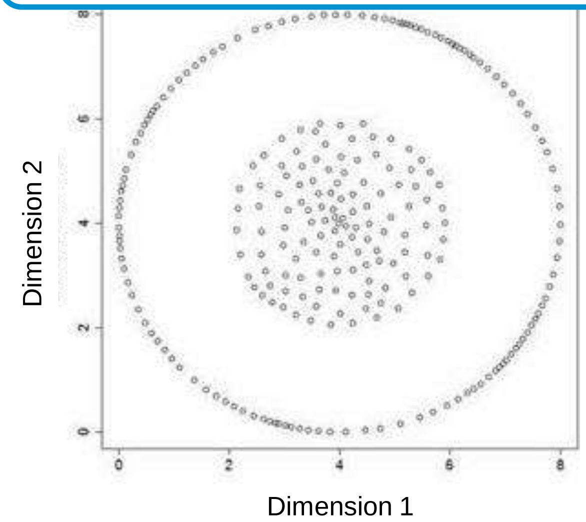
R packages DTW, DTWBI available

Phan H. et al., 2017. Pattern Recognition Letters

Multi-level spectral clustering

Classification spectrale NJW

(Ng et al., 2001)



1

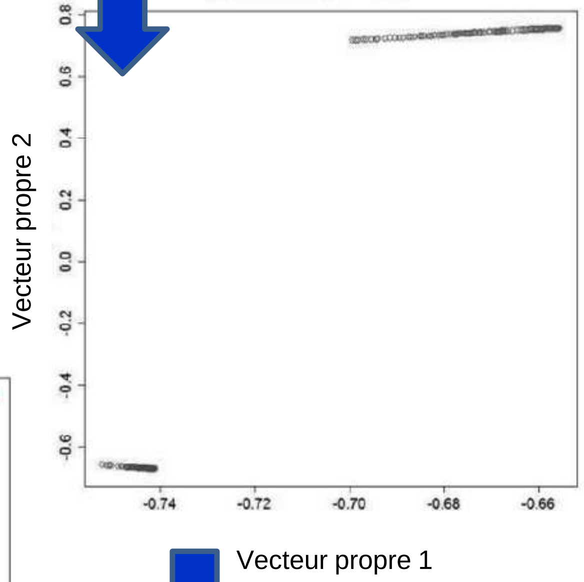
Calcul matrice de similarité

similarity matrix

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

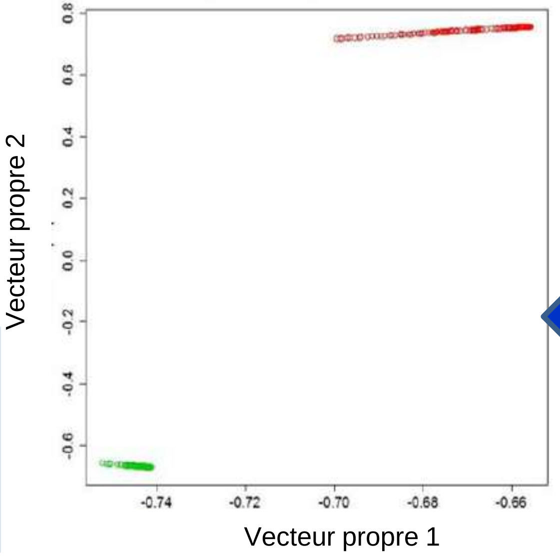
2

Projection des données dans l'espace spectral (espace des K vecteurs propres)



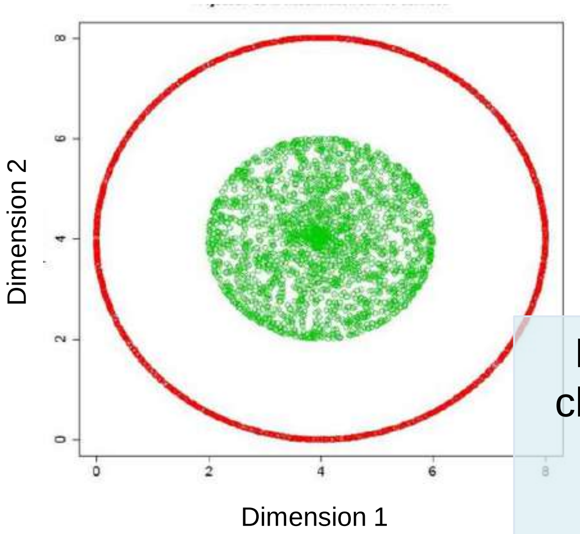
3

Classification K-means



4

Projection de la classification dans l'espace initial (échantillonné)



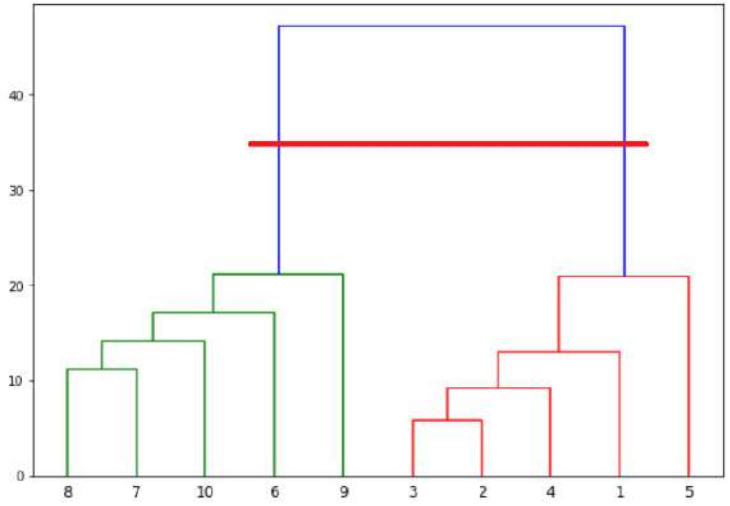
Grassi et al., 2019. IEEE Oceans 2019, Marseille

Multi-level spectral clustering

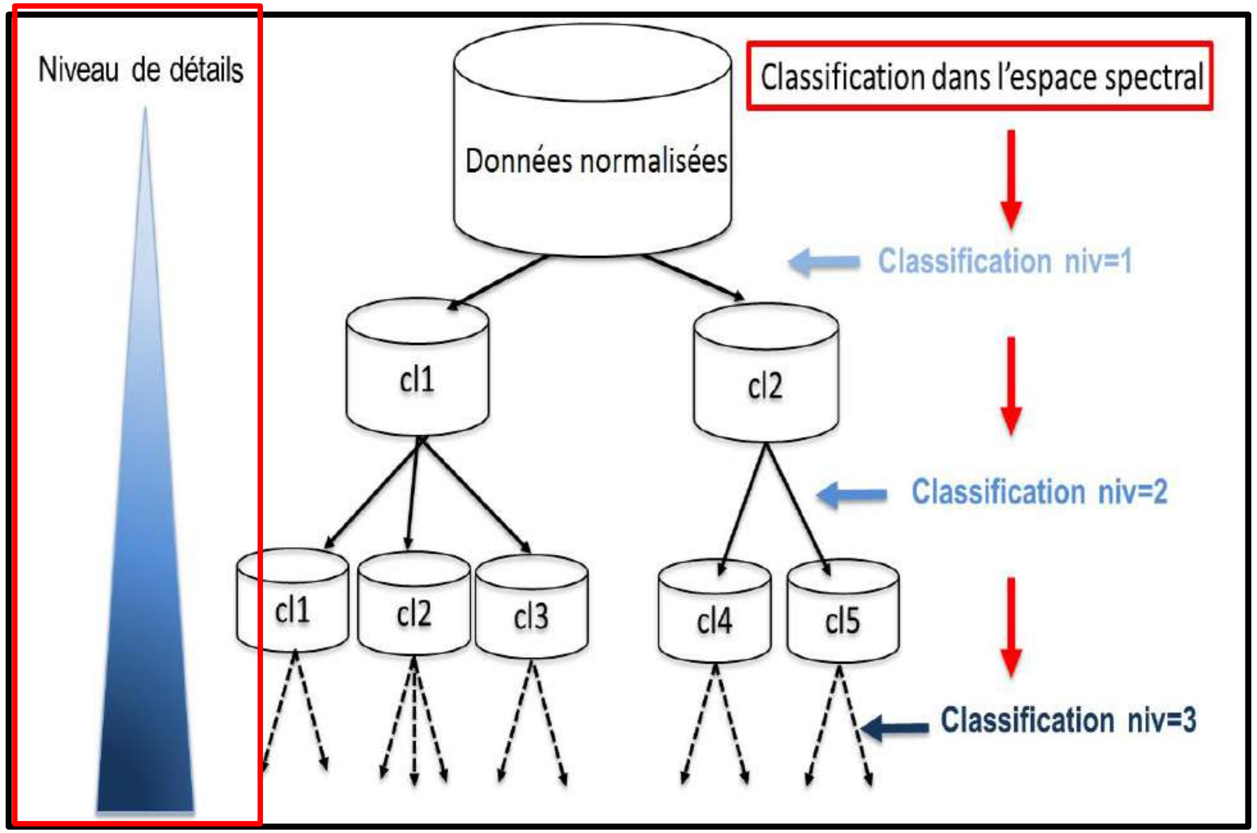
Du général … vers l'identification d'événements extrêmes

M-SC → Classification spectrale + Hiérarchique

Exemple Arbre
Hierarchique Clustering



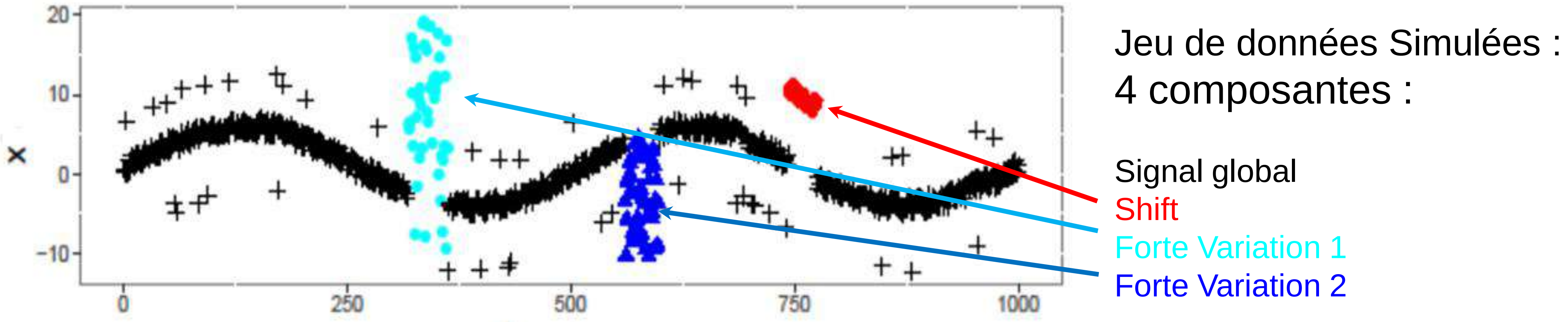
→ Augmentation du niveau de détail

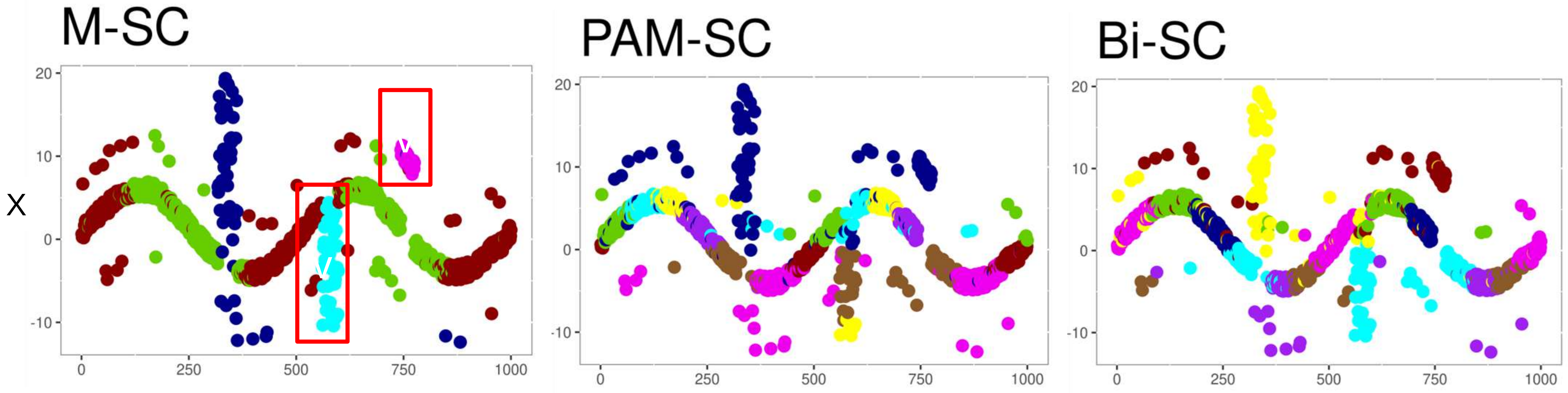


Multi-level Spectral Clustering

Exemple : Données simulées

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)





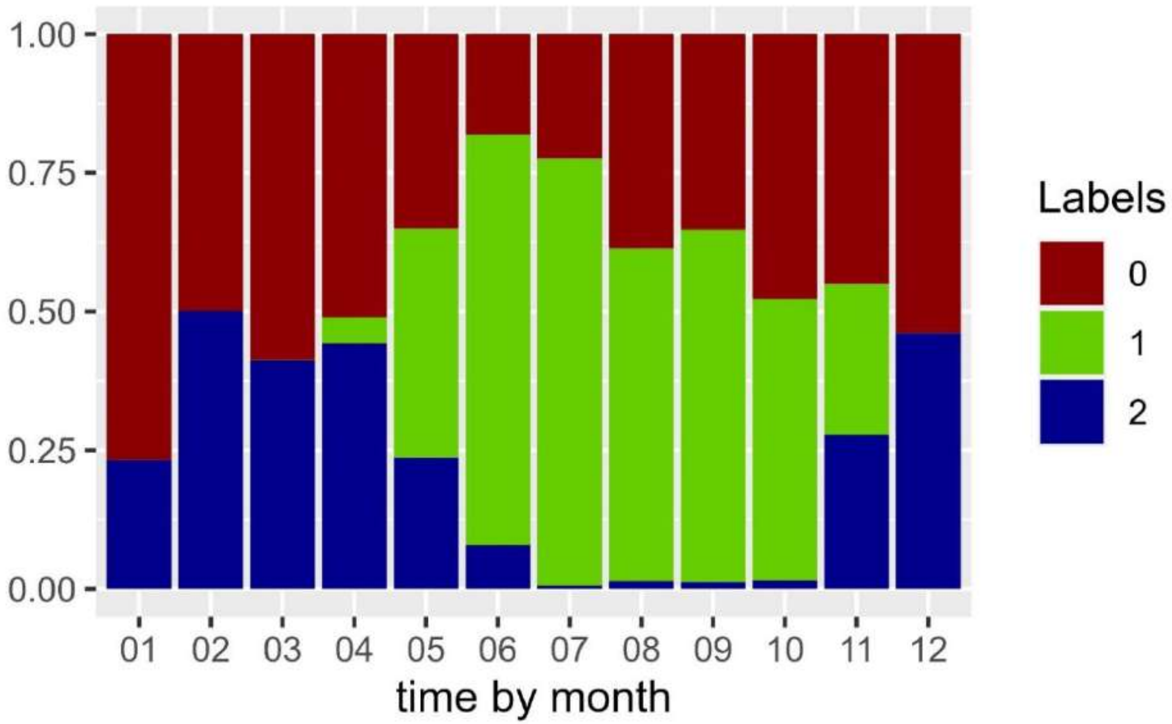
→ **M-SC : définition d'états extrêmes**
→ **SC-PAM, Bi-SC : Confusion avec le signal global**

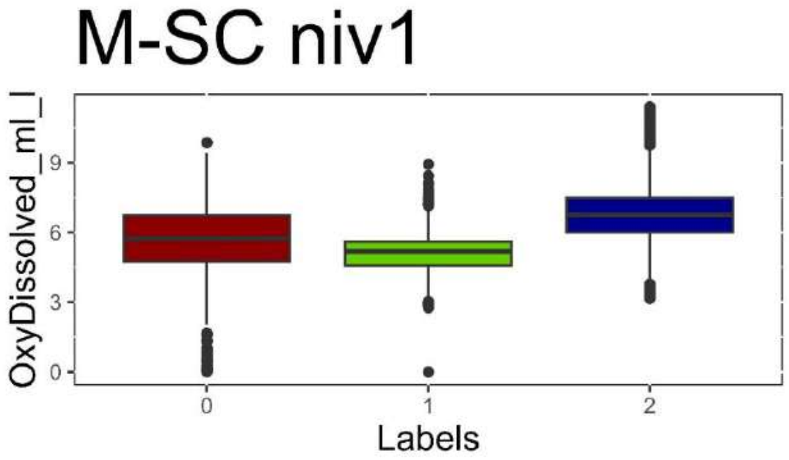
Multi-level Spectral Clustering

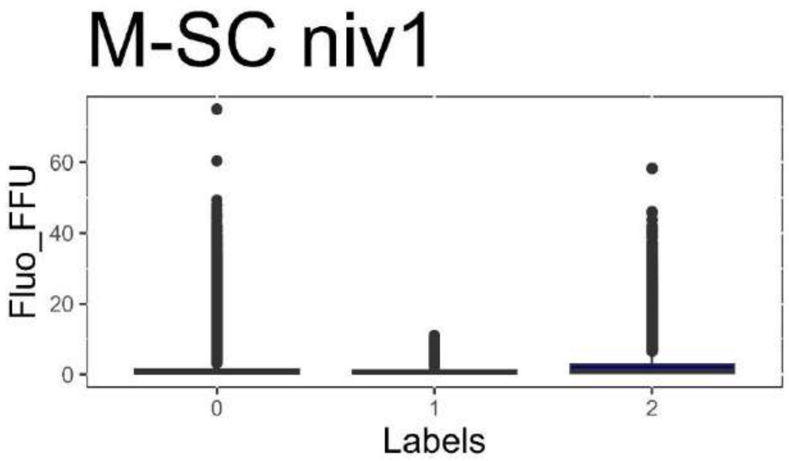
Characterization of phytoplankton biomass dynamics by defining multi-criteria environmental states

- MSC Level 1: Two periods were identified, one being more productive than the other

LEVEL 1







Journal of Marine Science and Engineering

MDPI

Article

Comparative Study of Clustering Approaches Applied to Spatial or Temporal Pattern Discovery

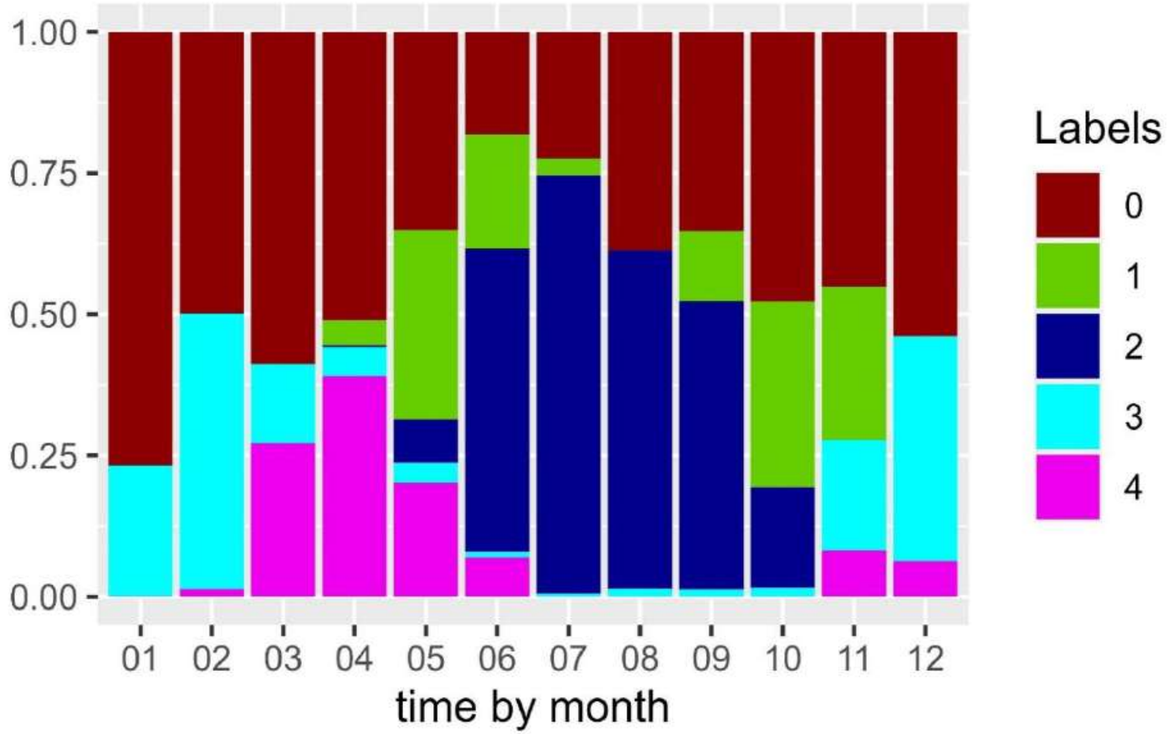
Kelly Grassi [1,2,3], Émilie Poisson-Caillault [2], André Bigand [2] and Alain Lefebvre [3,*]

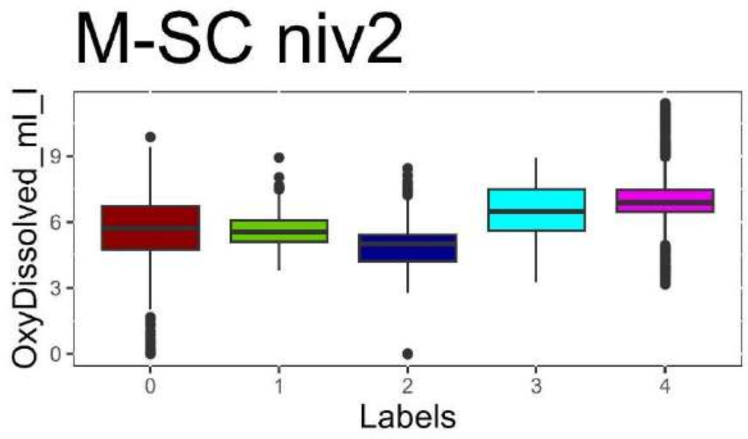
Multi-level Spectral Clustering

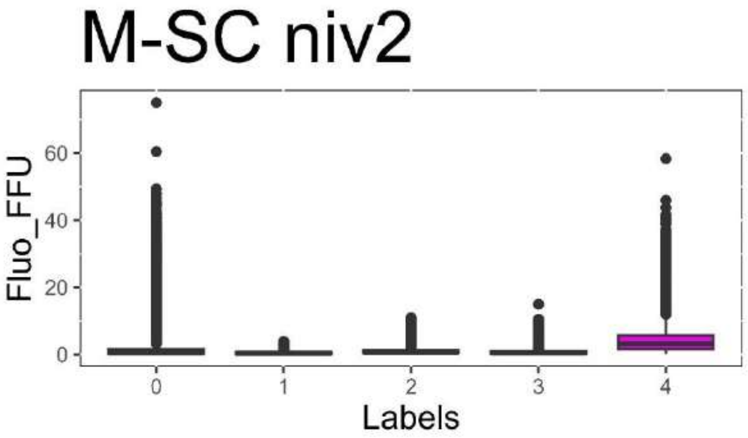
Characterization of phytoplankton biomass dynamics by defining multi-criteria environmental states

- MSC Level 2: Each of these two main periods (productive and non-productive) is divided into sub-periods corresponding to key environmental states: pre-bloom, bloom and post-bloom.

LEVEL 2





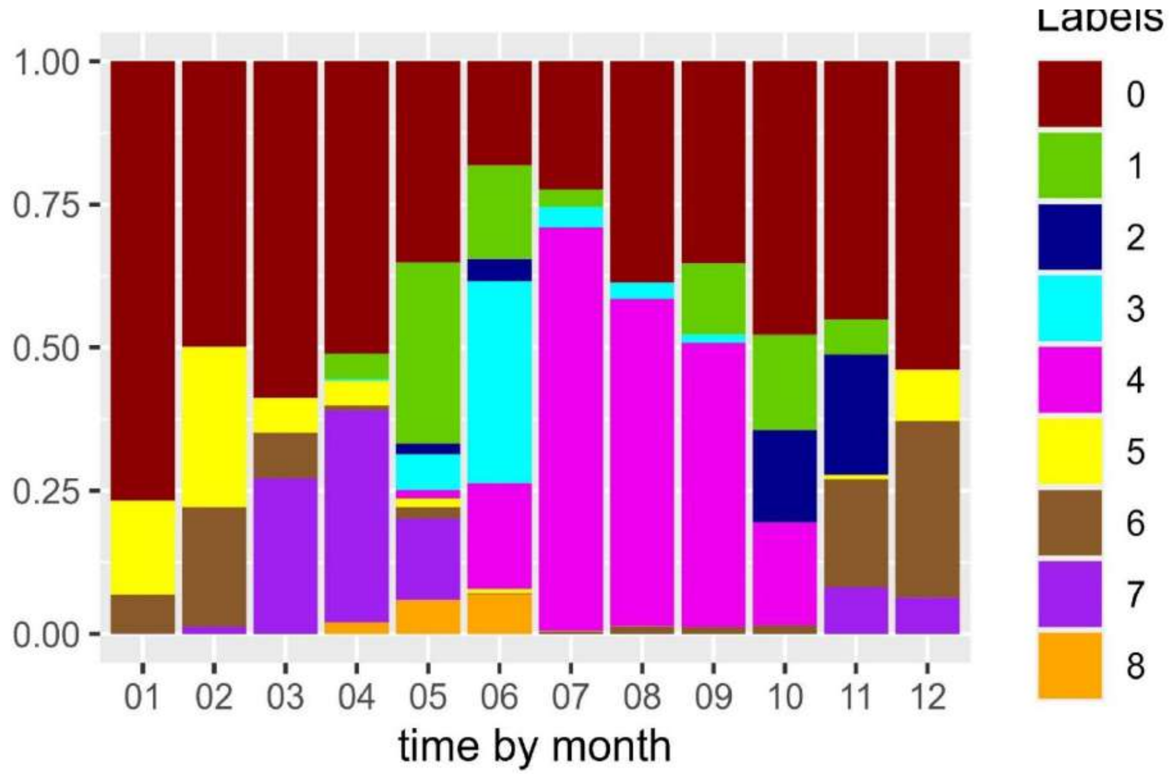


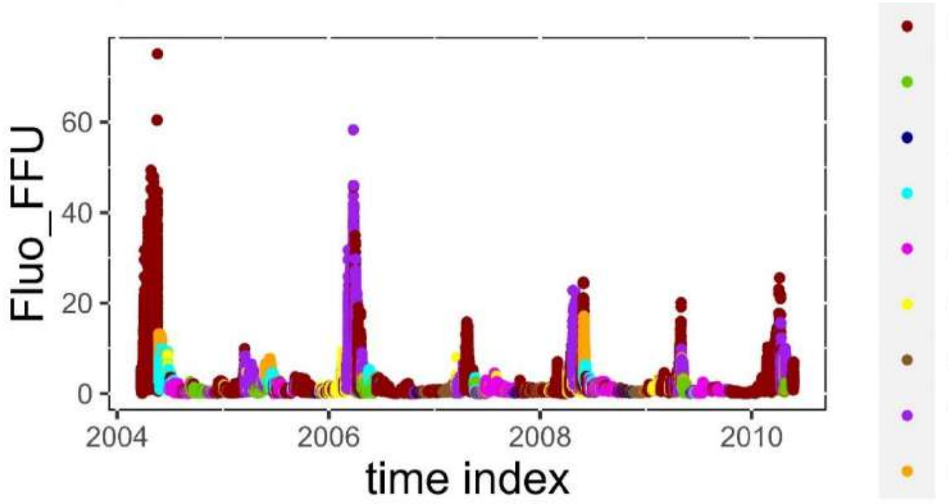
Multi-level Spectral Clustering

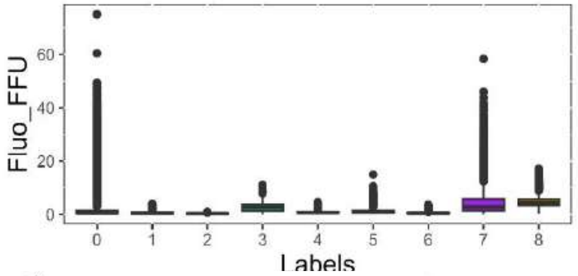
Characterization of phytoplankton biomass dynamics by defining multi-criteria environmental states

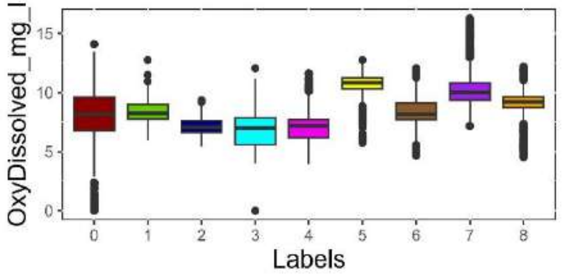
- MSC Level 3: 8 environmental states with different dynamics and characteristics in terms of controlling factors.
- Detect the start of a phytoplankton bloom when nutrients are added.

LEVEL 3

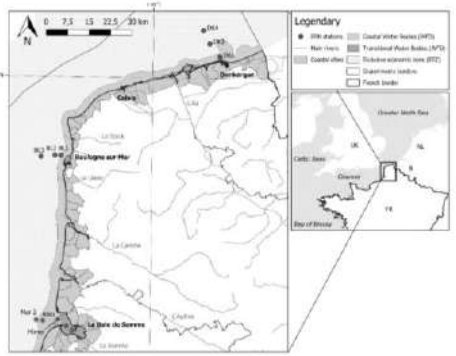






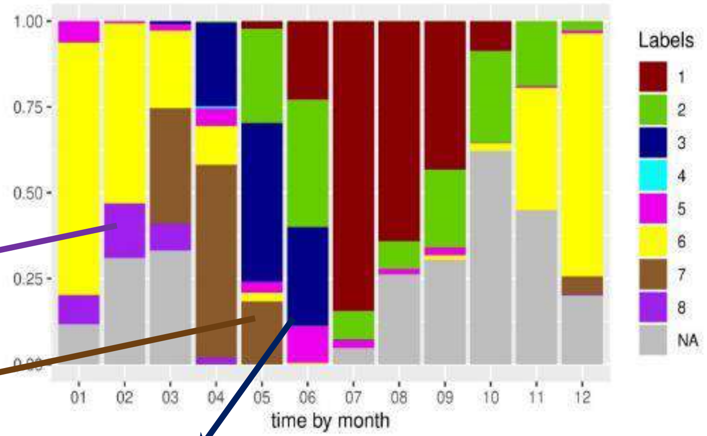


Multi-level Spectral Clustering



Bloom states: Classes 8, 7 and 3

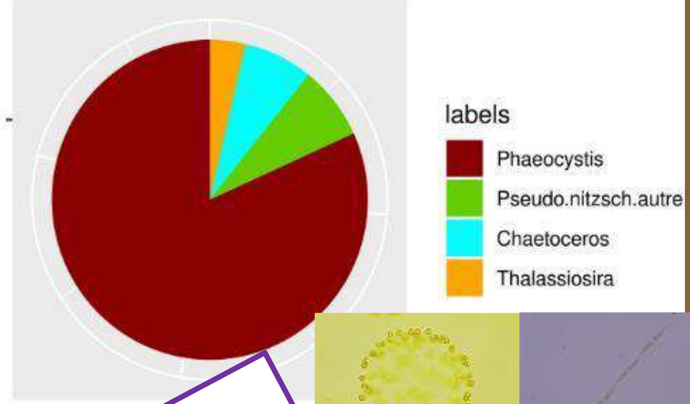
States correlated
with different phytoplankton community,
Including **Harmful Algae**

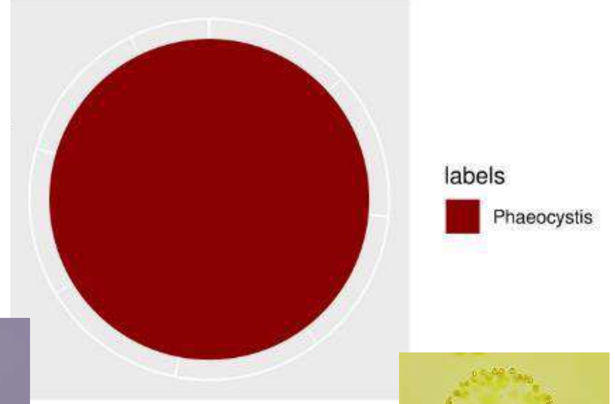


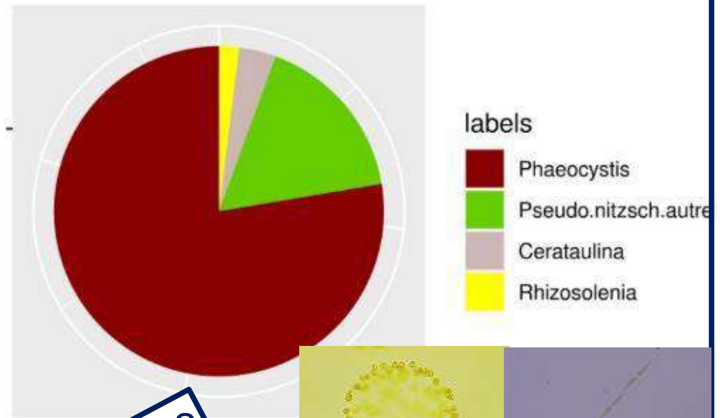
8 - Winter transition towards bloom state

7- Bloom

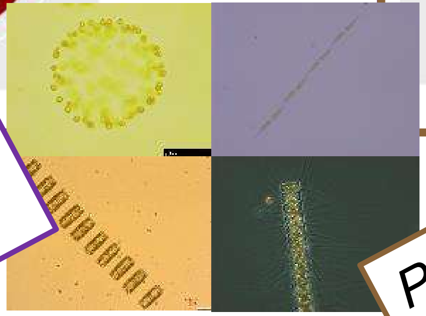
3- End of bloom





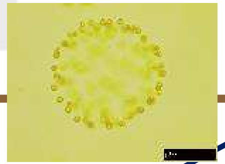


Phaeocystis 8
+ Pseudo Nz
+ autres



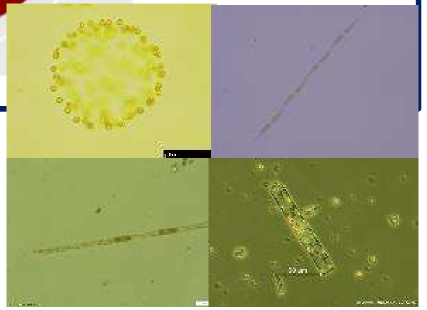
Toxic sp.

Phaeocystis 7



Harmful sp.

Phaeocystis 3
+ Pseudo Nz

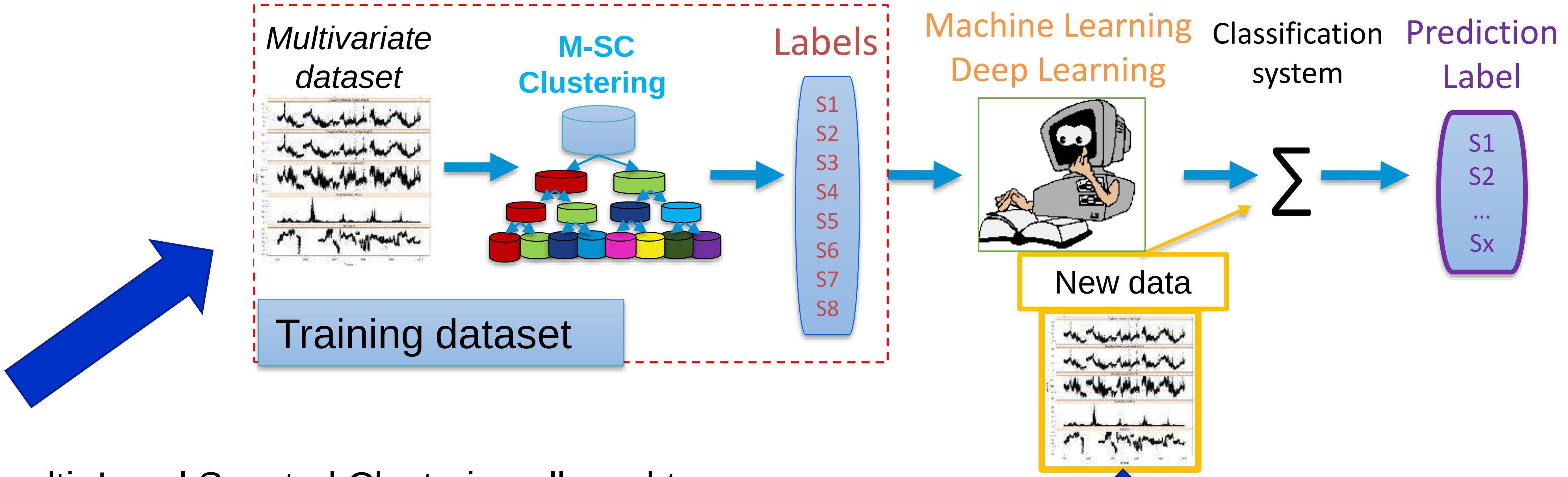


Toxic sp.

REPHY – SRN Data: Lefebvre & Devreker, 2023 - ESSD

CONCLUSION

R. Halawi Ghosn (PhD 2021-2024)



The multi- Level Spectral Clustering allowed to :

- Define **states** in multi-parameters time series
- Detect, identify and **characterize this states**
- **Extract label for** frequent, **rare or extreme events**

Based on labels, the system is able to **learn** and **build a model**.

New incoming data will be processed and affected to a given cluster.

Actual environmental state (= cluster) is identified ⬄ **HAB Forecasting**

PERSPECTIVE

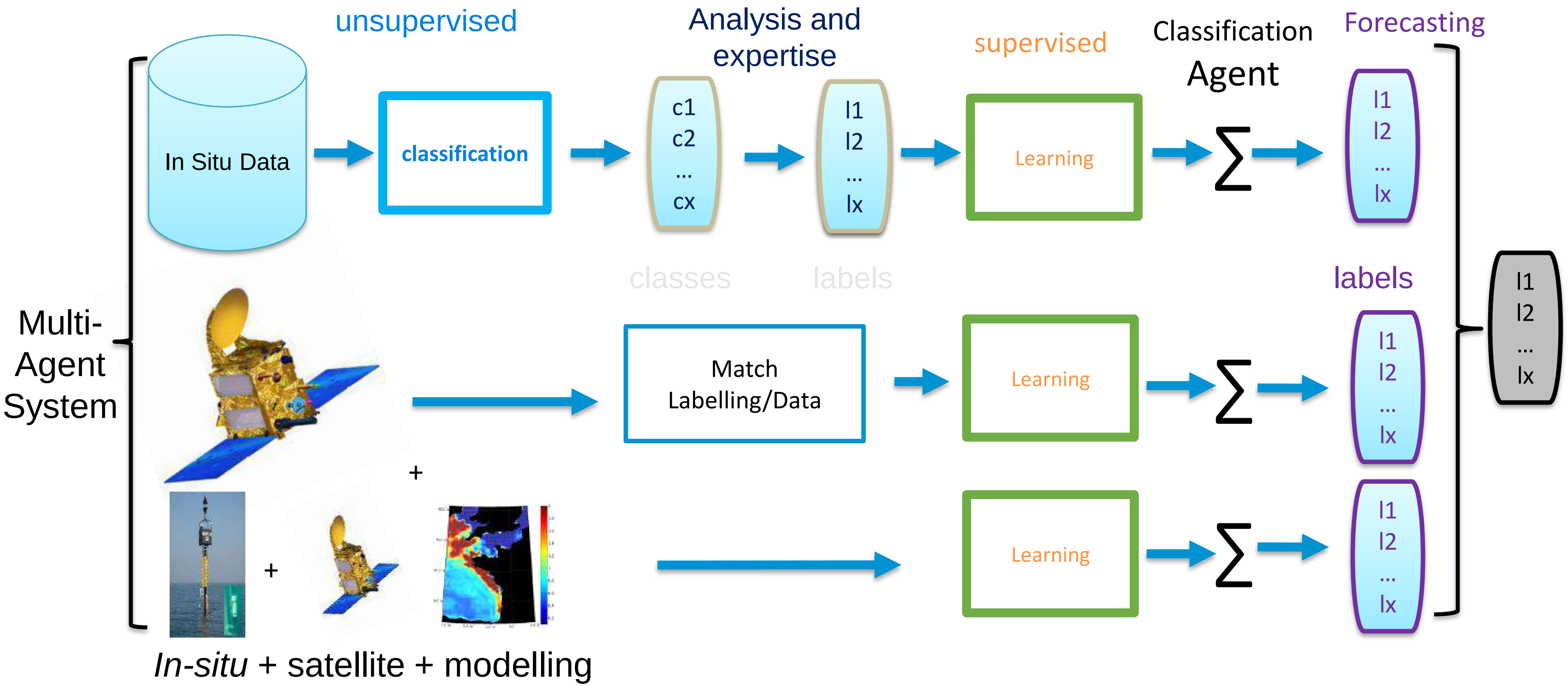
Towards an expert multi-agent system including Data Completion

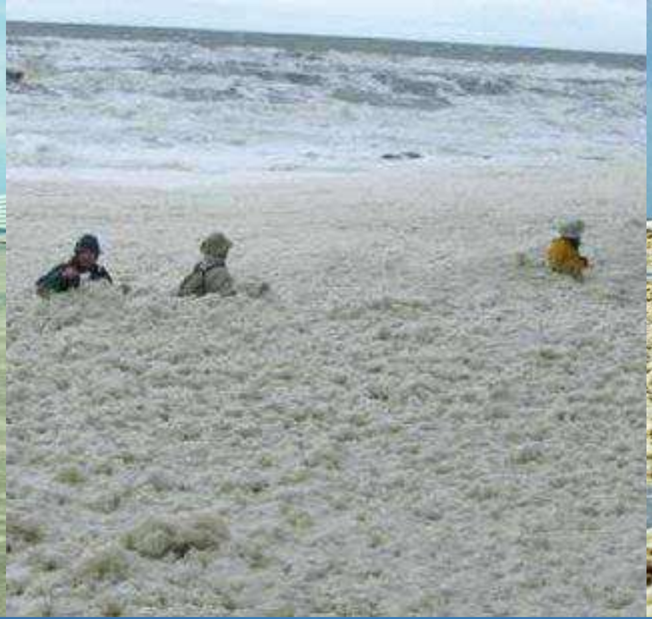
With identification from general pattern to extreme events

Based on multiparameter data series

Including an early warning system,

Allowing adaptative sampling strategy, forecasting, support of public policy





MERCI POUR VOTRE ATTENTION

Raed Halawi Ghosn PhD Thesis is funded under the Grant Agreement Number FP7-20186 by the Office Français pour la Biodiversité (OFB) and IFREMER

OFB
OFFICE FRANÇAIS DE LA BIODIVERSITÉ

ulco Université Littoral Côte d'Opale

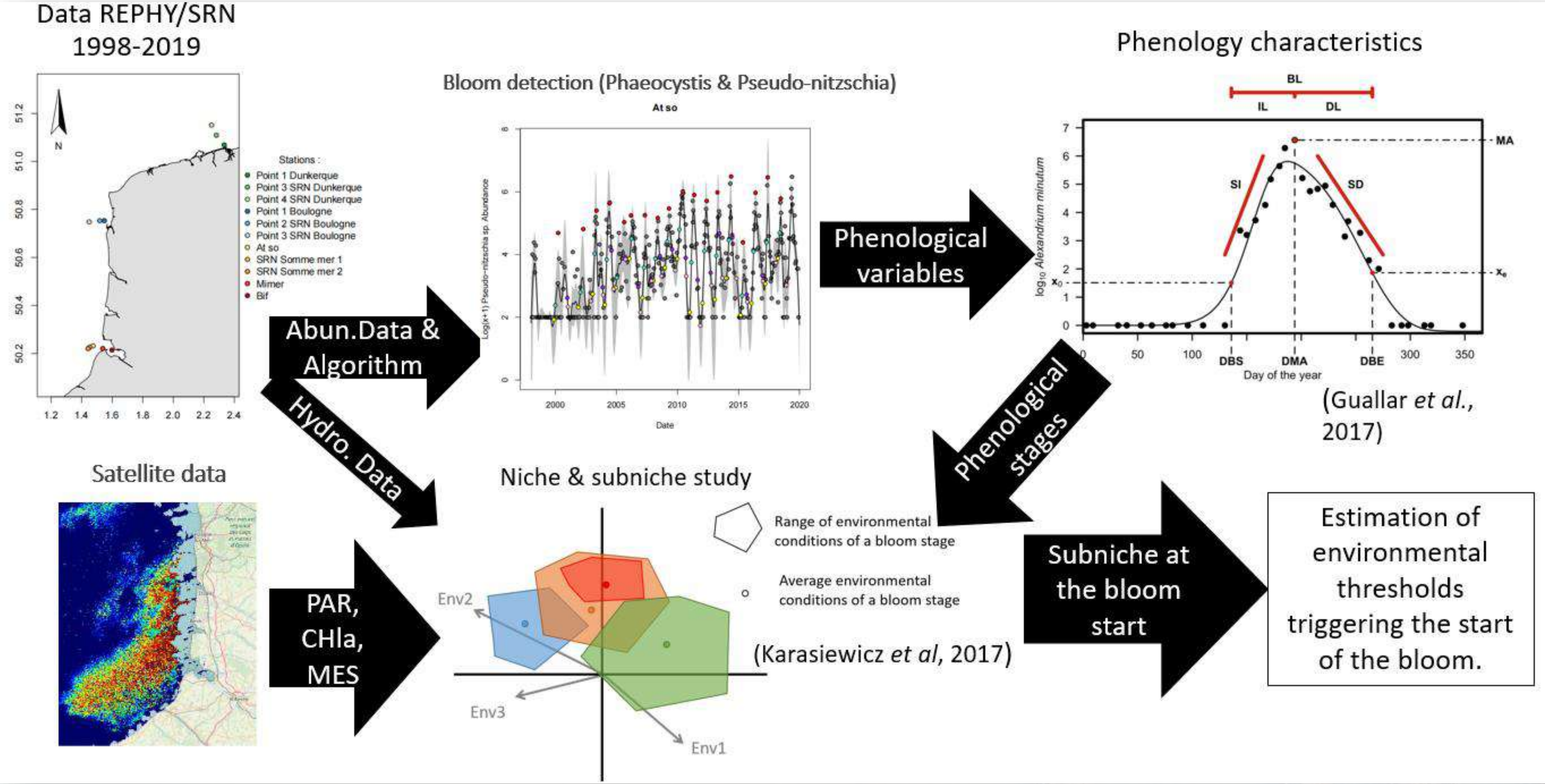
LISIC Laboratoire d'Informatique Signal & Image de la Côte d'Opale

Citation: Karasiewicz, S.; Lefebvre, A. Environmental Impact on Harmful Species *Pseudo-nitzschia* spp. and *Phaeocystis globosa* Phenology and Niche. *J. Mar. Sci. Eng.* **2022**, *10*, 174. https://doi.org/10.3390/jmse10020174

Environmental Impact on Harmful Species *Pseudo-nitzschia* spp. and *Phaeocystis globosa* Phenology and Niche

Stéphane Karasiewicz * and Alain Lefebvre *

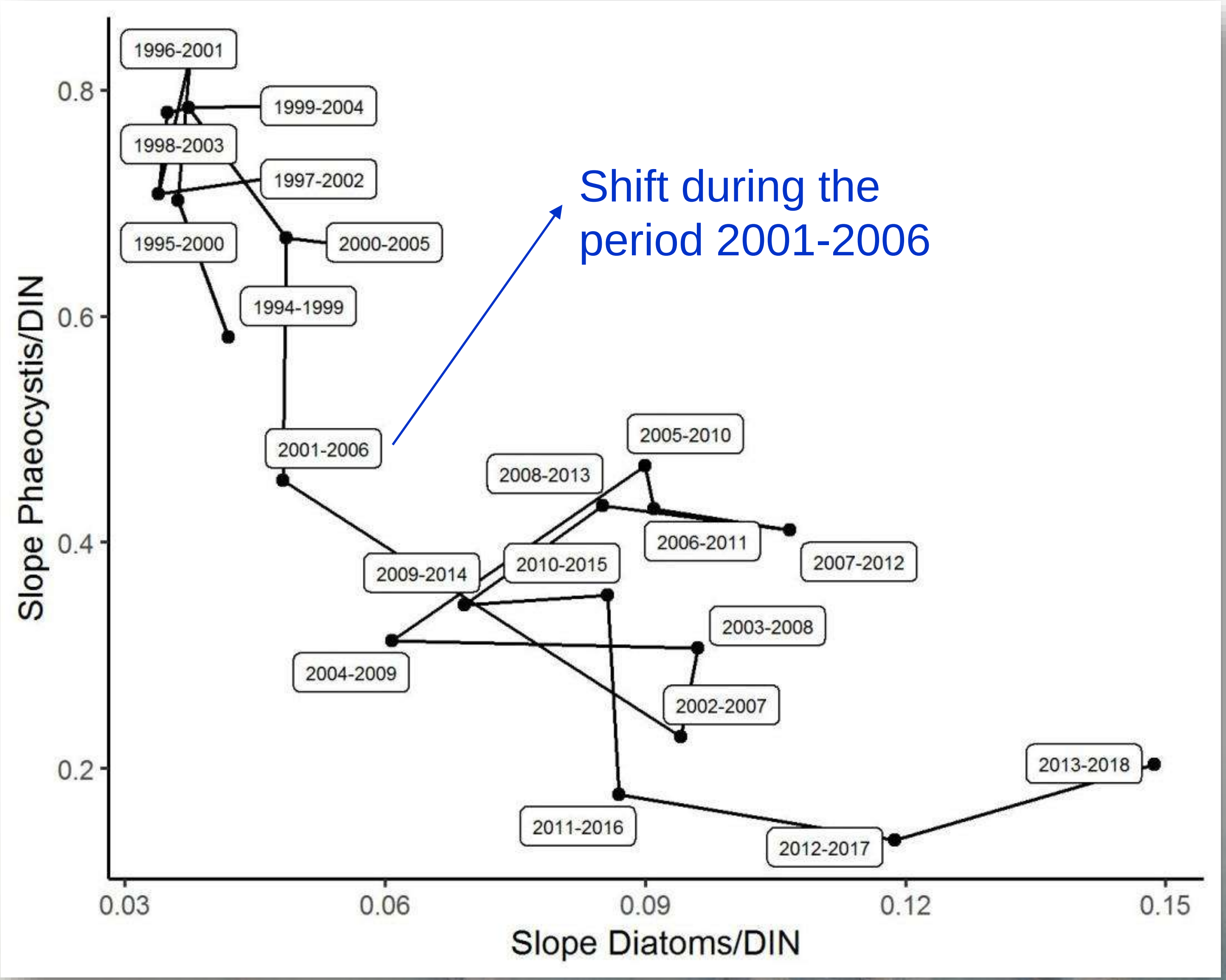
Coastal *in situ* observations



Harmful Algal Blooms – Observing, Understand and Predict

Bloom of Phaeocystis globosa in the English Channel

Changes in the balance of the *Phaeocystis* / Diatoms ratio



With long-term **coastal observations**:

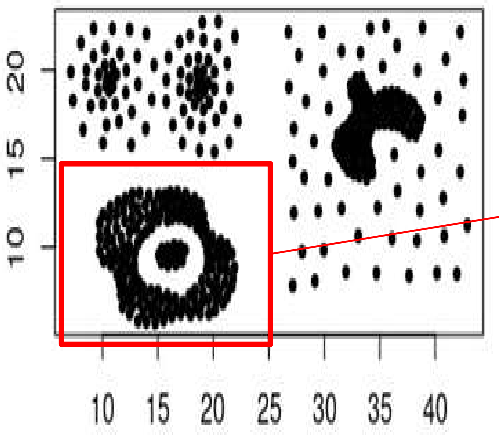
- Ability to **detect and prevent HABs**
- Ability to deconvolute **global and local changes**
- Including changes from low to high trophic levels

Multi-level Spectral Clustering

Exemple : Base de référence

Jeu de données «Compound »
Données UCI
N=399, k=6, D=2
(source C.T. Zahn, Compound, 1971)

X2



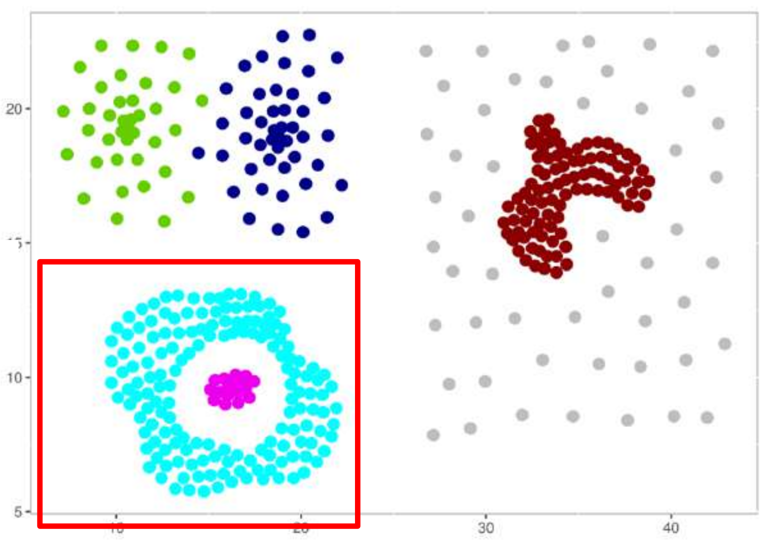
X1

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

Classes imbriquées

M-SC Niveau 3

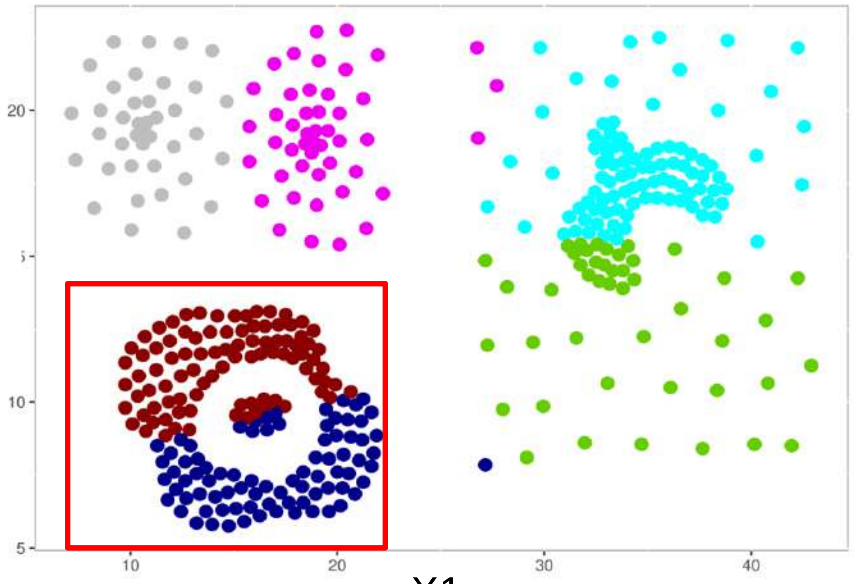
X2



X1

KM K=6

X2



X1

→ M-SC: isole les classes imbriquées

→ KM : tendance à sur-segmenter


Ifremer


Ifremer