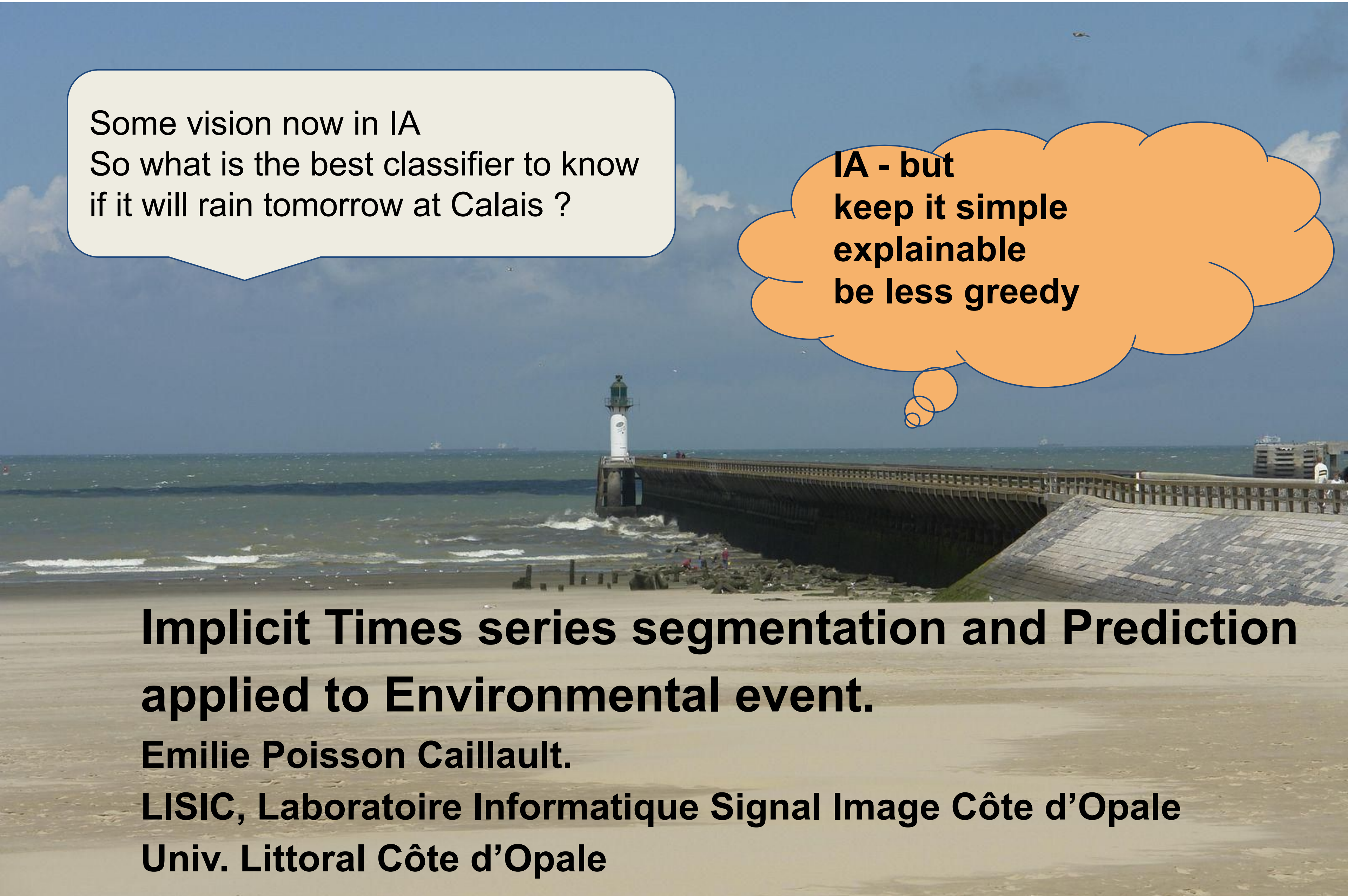
Journée ADéM - Lille 9.12.2024
Apprentissage à partir de Données d'Écologie Marine



Implicit Times series segmentation and Prediction applied to Environmental event.

Emilie Poisson Caillault.

LISIC, Laboratoire Informatique Signal Image Côte d’Opale

Univ. Littoral Côte d’Opale

Some vision now in IA
So what is the best classifier to know if it will rain tomorrow at Calais ?

Do you see the English coast ?
No, so it will be sunshine
Yes, Take your rain coat tomorrow.

IA - but
keep it simple
explainable
be less greedy

LISIC - Focus on *IA applied to Ecology*

LISIC 4 Teams with IA Machine Learning core https://www-lisic.univ-littoral.fr/

IC - Intégration des connaissances	Explained IA, Ontology
OSMOSE - Optimisation et modèles	Fitness landscape, heuristics, evolutionary algorithms
IMAP - Image and Apprentissage	Image Synthesis, hyperspectral descriptor, Machine Learning
SPECIFI Syst de Perceptions	NMF-networks, GNSS, Information fusion

2010…Automatic detection and monitoring of pollution by analysing of smoke plumes. (SPECIFI)

2009…Prediction of rare and common events for monitoring water quality (algal blooms). (IMAP)

Automatic recognition of marine waste using hyperspectral imaging. *LOG*

Analysis of particle filters containing microplastics using hyperspectral microscopic vision. *LOG*

Biogeochemical analysis of coastal environments from multispectral satellite images.

Identification of fish stocks by otoliths. *IFREMER*

Automatic identification of fish traits of life from 3D otoliths shape *IFREMER*

Characterization of the rate of infestation of marine fauna by Anisakis worms. *ANSES*

Automatic recognition of R. Tridactyla diet regim from their dejection (Otolits). GON/LOG/*IFREMER*

Now- Short-Term monsoon Forecasting from opportunity sensors *IRD*

Short personal card

2005 : PhD, Univ. Nantes at LS2N (IRCCYN Lab).

Architecture and Training of a hybrid Neuro-Markovian System for On-Line Handwriting Recognition

Keywords : SD-TDNN, MS-TDNN, global discriminant training, MLE-MMI, Mask/Filter in convolution layer.

2006 : Assistant Professor - Univ Littoral in data science and machine learning

2014-16 : **IFREMER delegation**

2020 : HDR **Contributions to the classification and segmentation of Time series by statistical unsupervised or guided learning**

Keywords : time series, similarity, DTW-criteria, DTW-imputation, spectral clustering, multi level approach

2023 : University Professor, Univ. Littoral Côte d'Opale

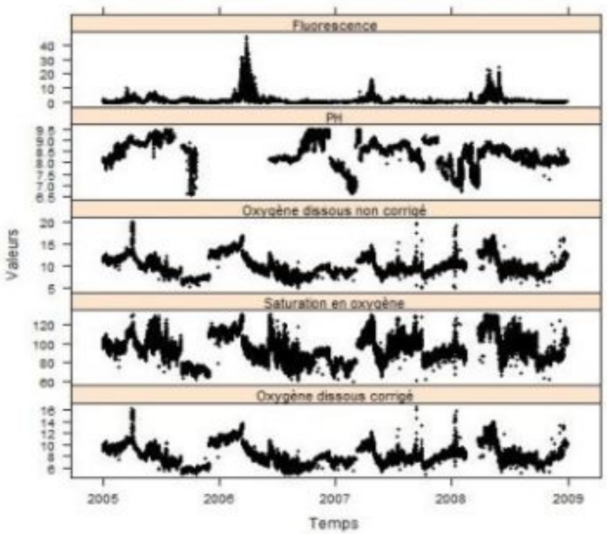
Project in progress

CPER IDEAL, CPER CornelIA
LISIC/IFREMER PhD supervision
ANSES/LISIC/FromNord - projet STImule Anisakis
GON/LOG/IFREMER/LISIC - R. Tridactyla/Otolits
IRD/LISIC - Rainsmore

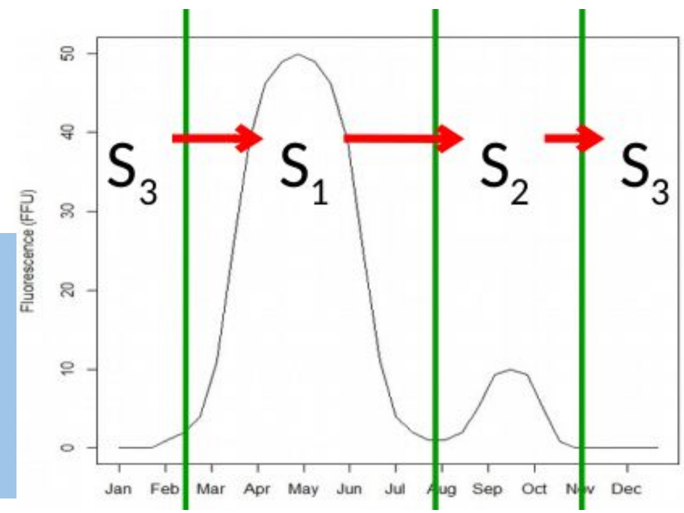
Functions

ULCO Computer sciences Dep. Director
ULCO CAC and CVFU member
ULCO VAE jury member
MAIA - Teaching coordinator
SFR - CS member

Implicit Times series segmentation
and Prediction
applied to Environmental event.



Discovering
Modelling
of events
(S: states)



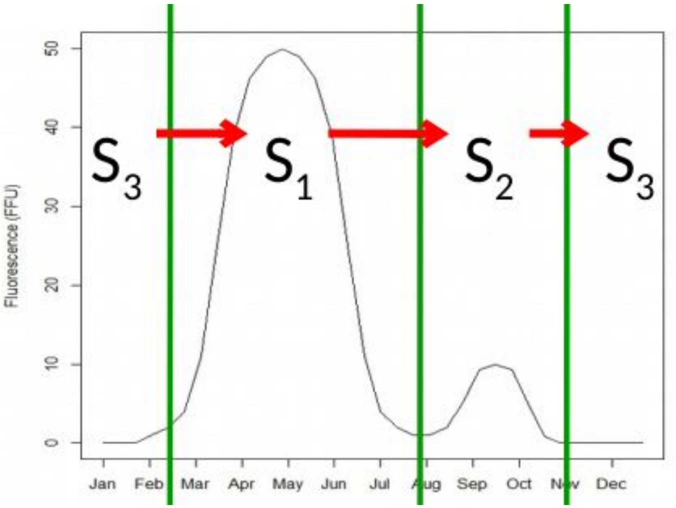
1- unsupervised context
state revealing

GMM, SC-HMM, MSC

2 - supervised context
state condition learning
state dynamics learning

RF, ConvLSTM, DTWUMI, TimeGAN

1- Time Series Segmentation by clustering



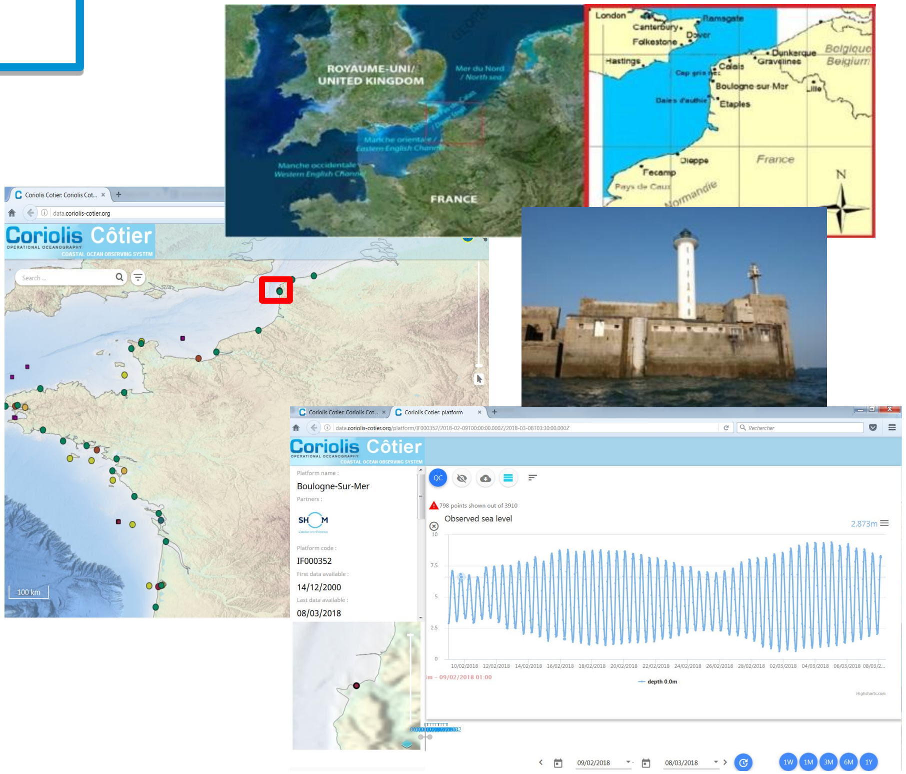
Main case of study: MAREL Carnot

Dataset
MAREL Carnot

Instrumented station (HF)
EOV measures <-> Phytoplancton
sampling every 20 minutes / 12 hours

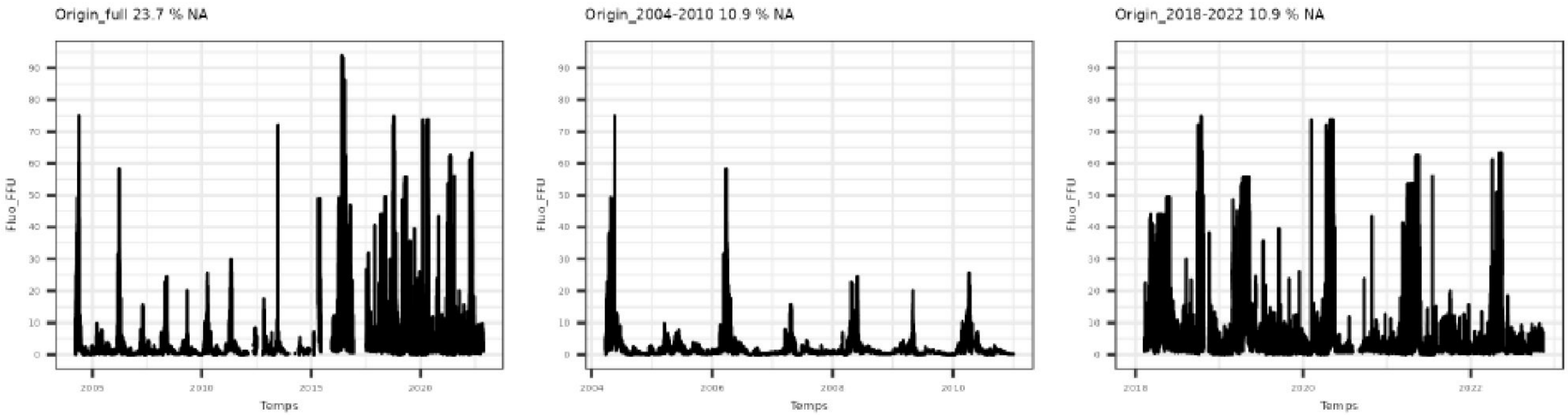
Raw data
2004-2024
#NA> 60 %

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

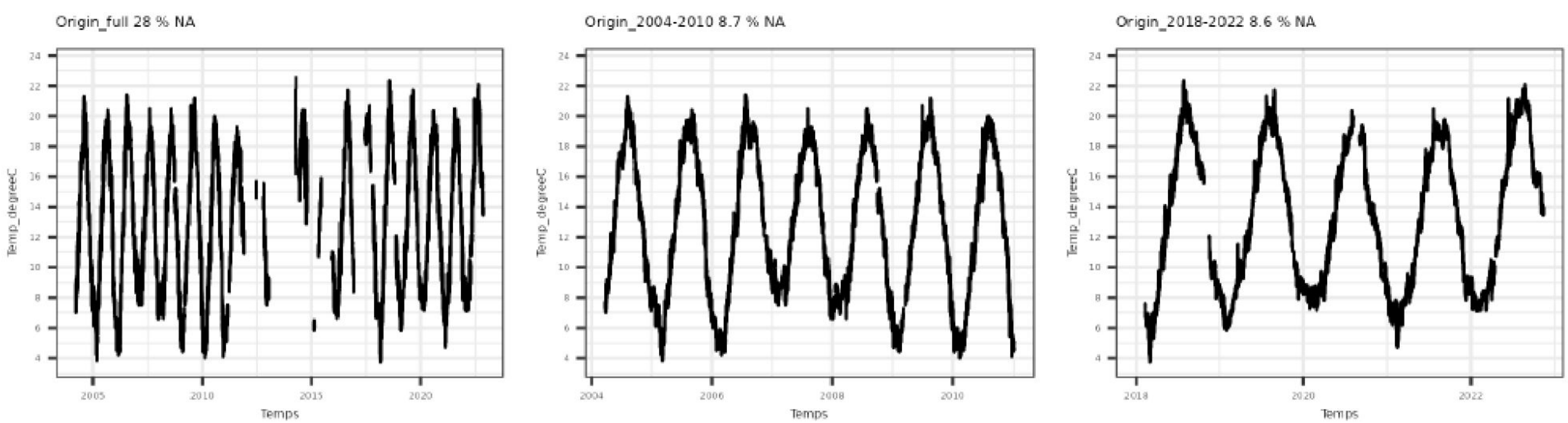


http://data.coriolis-cotier.org/

MAREL Carnot : ‘clean’ Time series

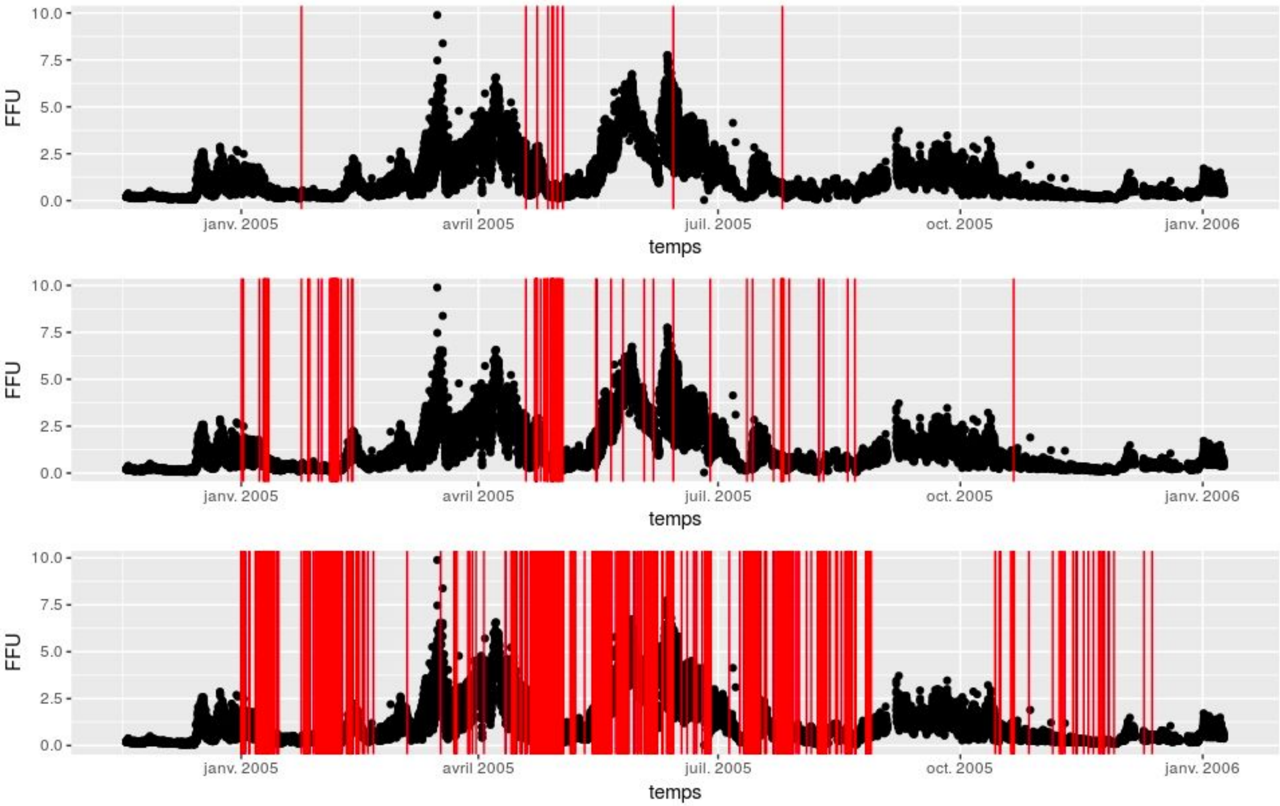


↑ Fluorescence ↓ Température



Raed Halawi Ghosn, Guillaume Charria, Armel Bonnat, Michel Repecaud, Jean-Valery Facq, et al.. MAREL Carnot data and metadata from the Coriolis data center. *Earth System Science Data*, 2023, 15 (9), pp.4205-4218. ⟨10.5194/essd-15-4205-2023⟩. ⟨hal-04229441⟩

Event or region detection



Approaches

Univariate :

- Breakpoints, PIP, trend

- Explicit segmentation
- Implicit segmentation

Multivariate :

- Explicit segmentation
  Scattering moments
- Implicit segmentation

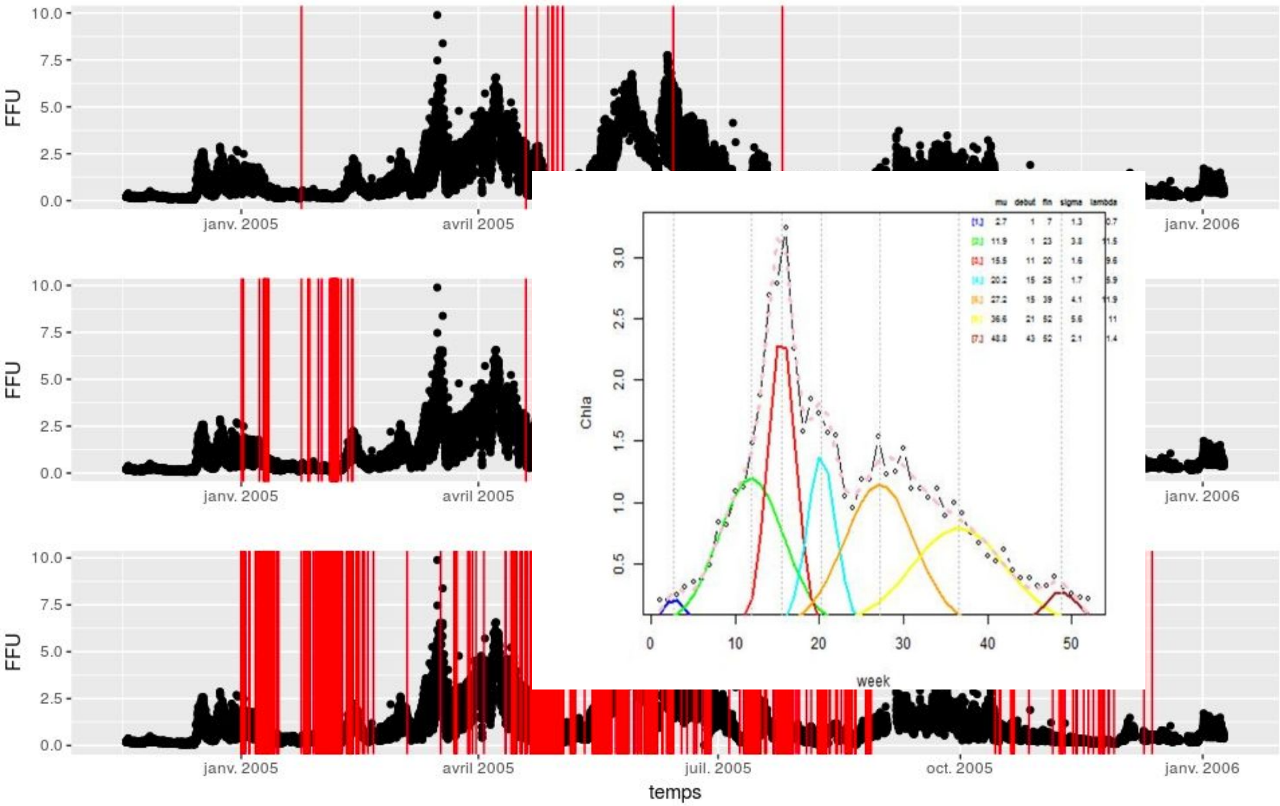
cut process:

- Suitable for trend analysis
- Imposes clustering/matching before labelling

.... Tedious for the expert

.... Costly in terms of calculation

Event or region detection



Approaches

Univariate :
- Breakpoints, PIP, trend
- **Explicit segmentation**
- Implicit segmentation

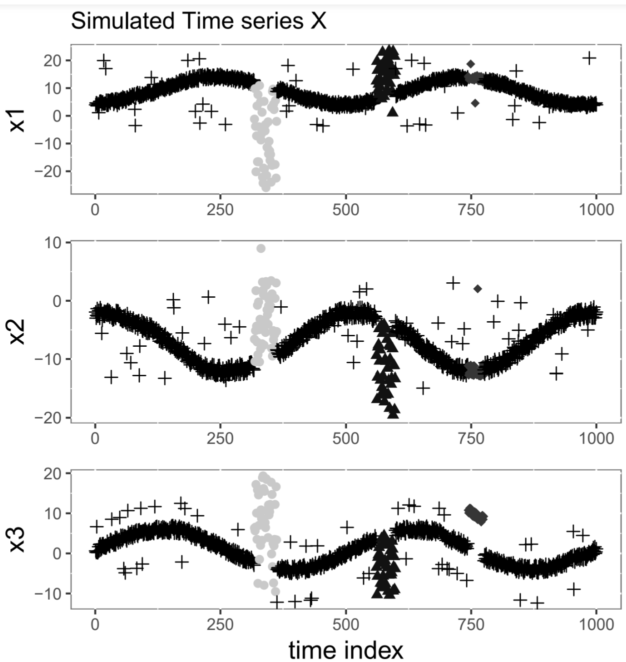
Multivariate :
- **Explicit segmentation**
  Scattering moments
- Implicit segmentation

Article : Towards Chl-a Bloom Understanding by EM-based Unsupervised Event Detection. Emilie Poisson CAILLAULT and Alain LEFEBVRE. Full accepted paper. OCEANS 2017 MTS/IEEE, Aberdeen, Scotland, 06/2017

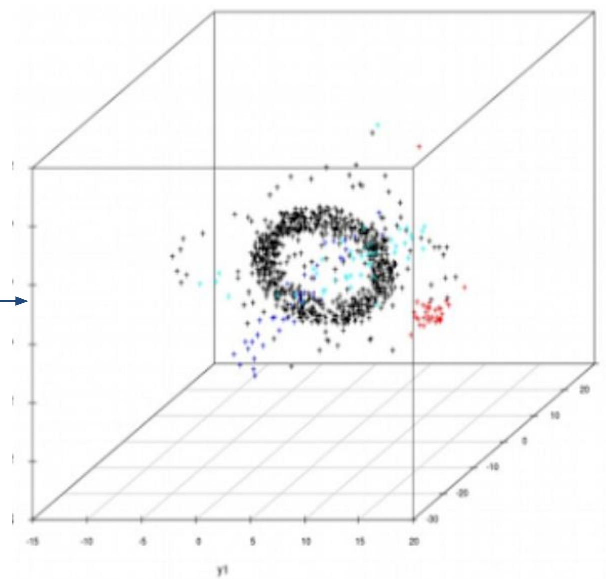
Detection of mixture of patterns
Requires a priori
- Forms of event (gaussian?)
- Series statistics

Implicit segmentation by clustering approach

1- Compute similarities between Observation features → W
2 - Apply Partitioning algorithm in this Observation space
3 - Analyse obtained dynamics and sometimes correct it.



(x1,x2,x3,t)



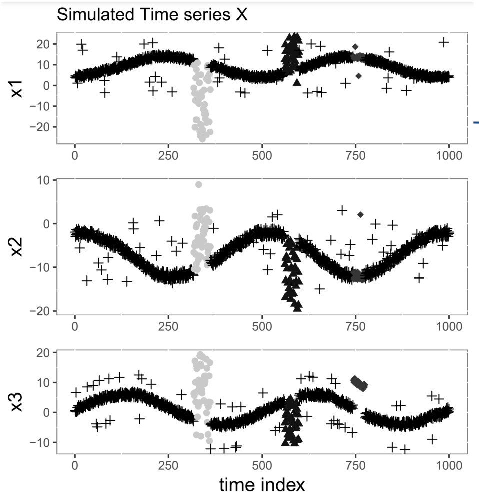
(x1,x2,x3)

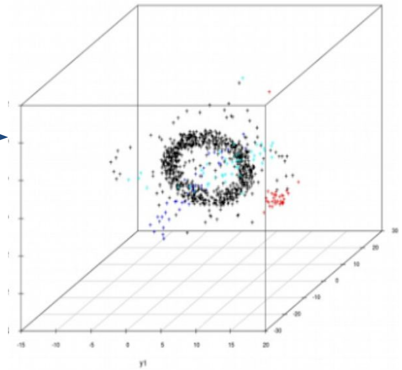
Clustering

(x1,x2,x3) + label

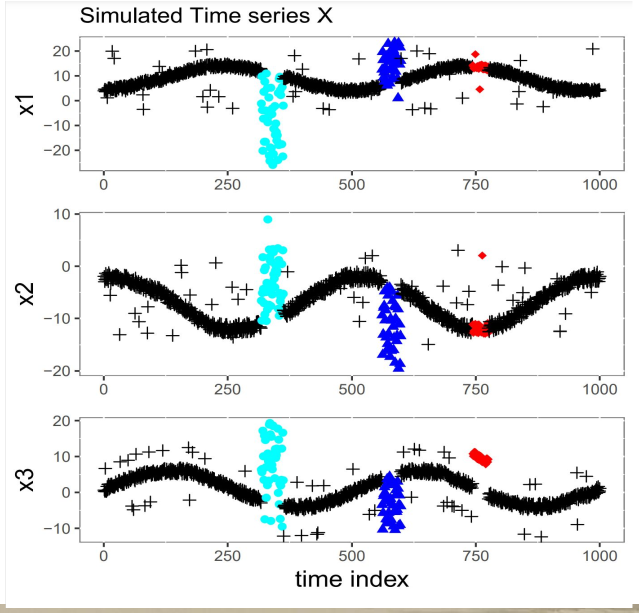
Implicit segmentation by clustering approach

1- Compute similarities between Observation features → W
2 - Apply Partitioning algorithm in this Observation space
3 - Analyse obtained dynamics and sometimes correct it.



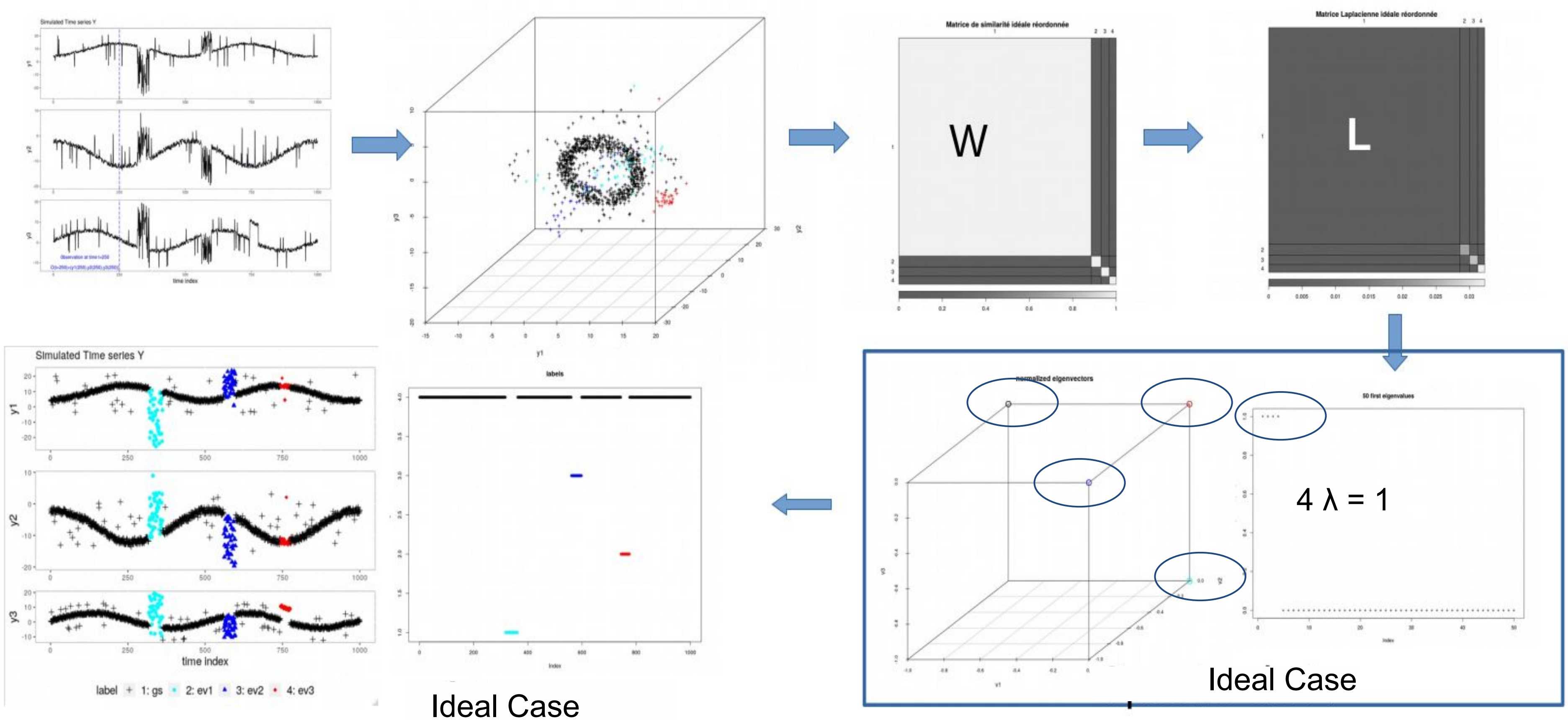


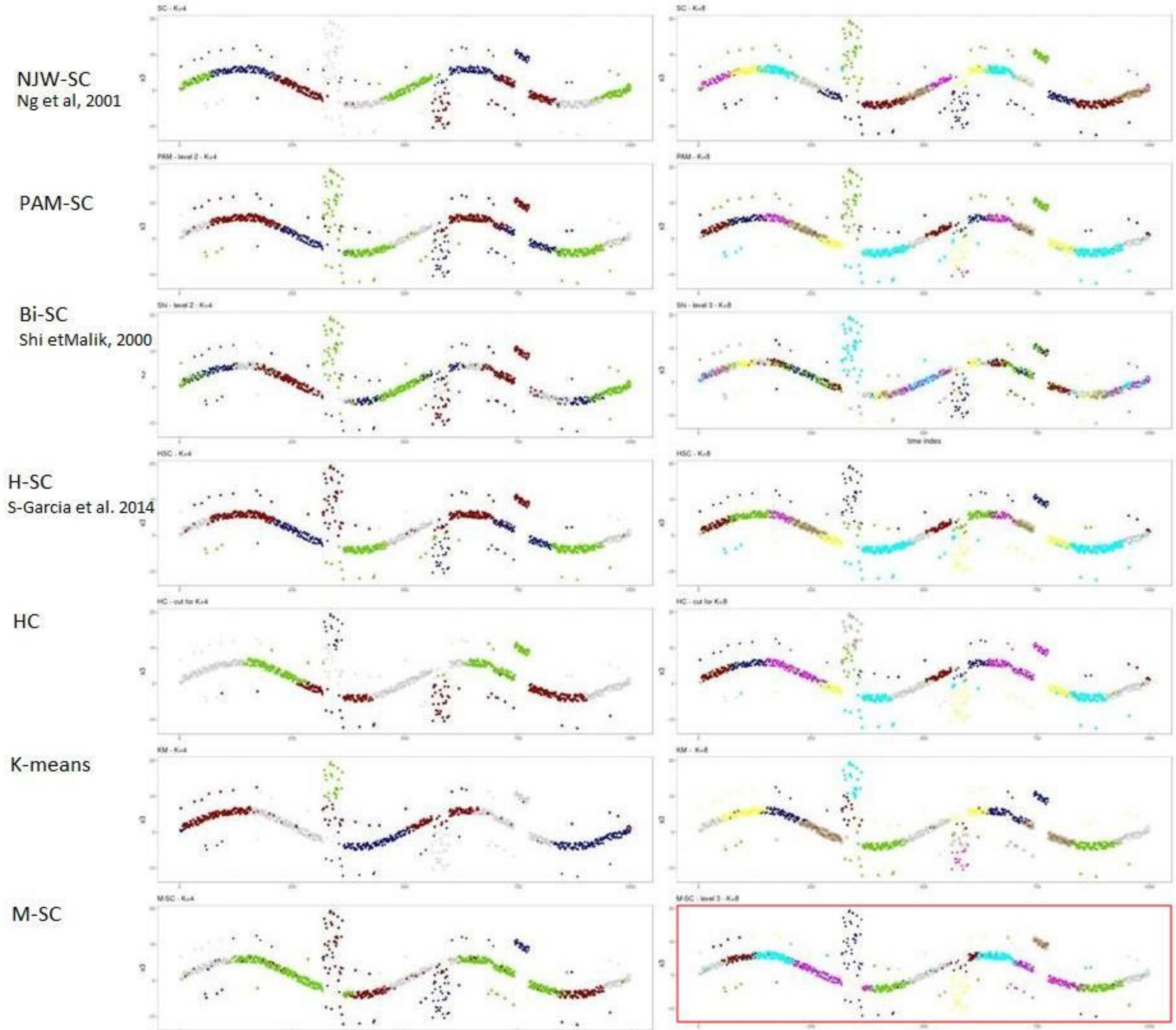
Clustering



Implicit segmentation by spectral clustering approach

1- Compute similarities between Observation features -> W
2- Compute Laplacian matrix from W
3- Extract eigenvectors V and eigenvalues -> detect K principal values
4 - Partitionning data in the normed K-first vector eigenspace U (PAM)

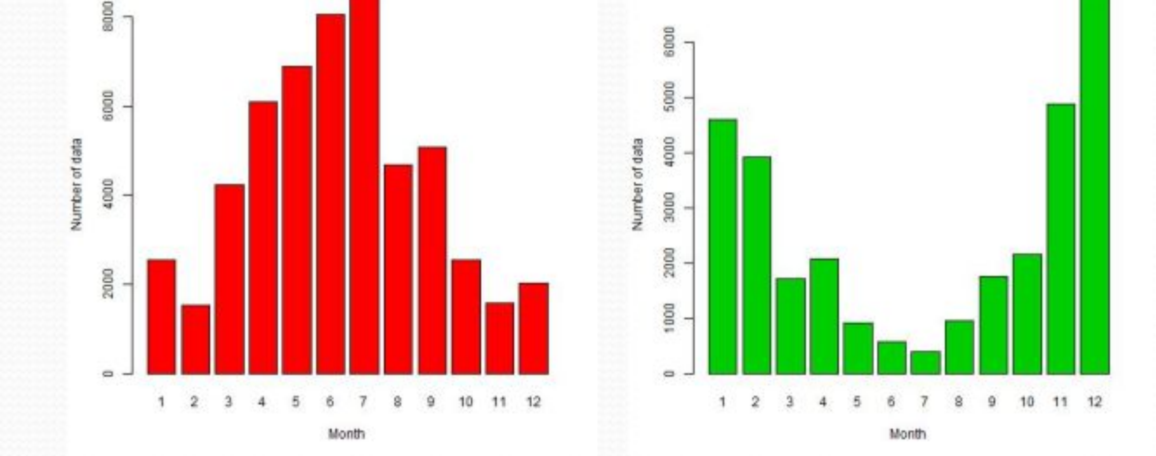



NJW-SC
Ng et al, 2001
PAM-SC
Bi-SC
Shi etMalik, 2000
H-SC
S-Garcia et al. 2014
HC
K-means
M-SC

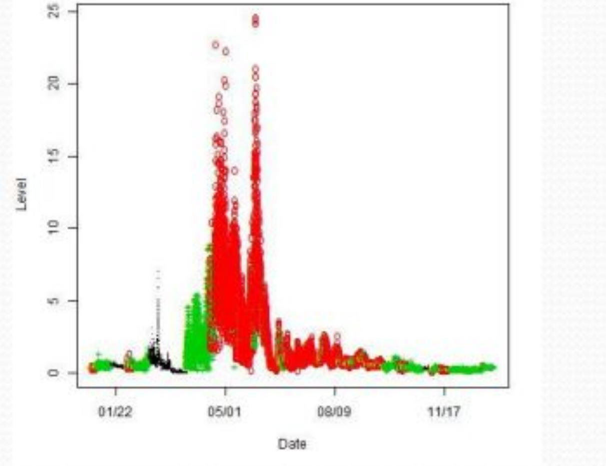
Long-term series segmentation: Marel Carnot application.

SC with K=2: Identification of low-productive period vs productive period.

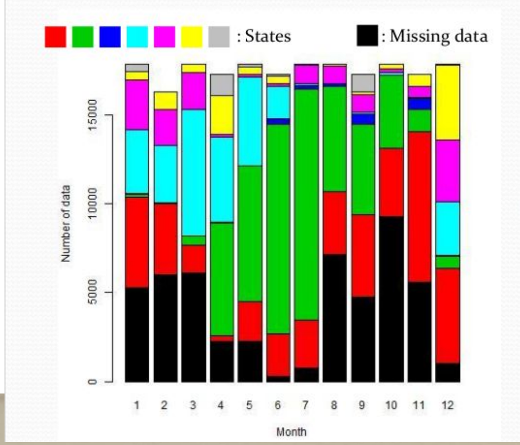
2005-2009 state distribution

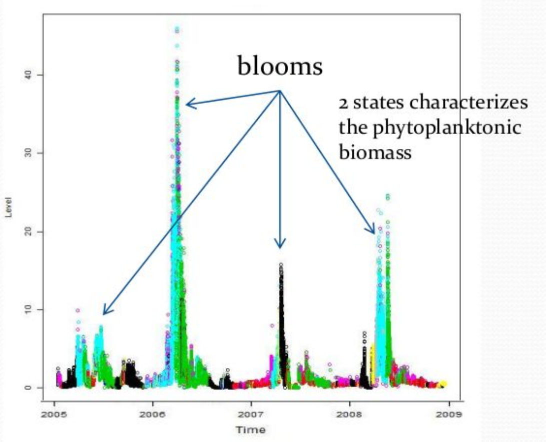


2008- Fluorescence



SC with K=7: Identification of blooms, pre/post-blooms, rare events



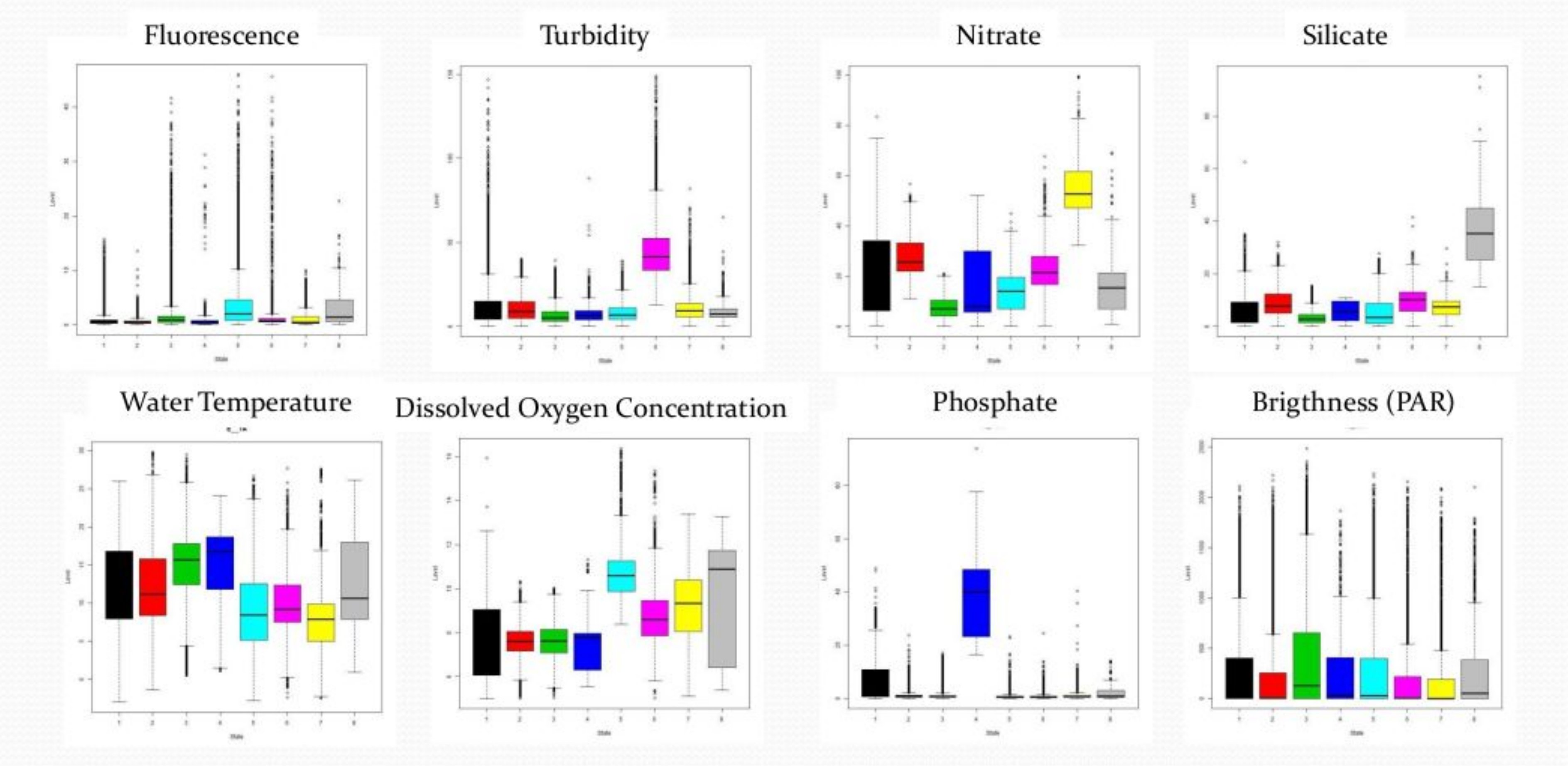


phD Rousseeuw, 2013

enc. Alain Lefebvre et E. Poisson

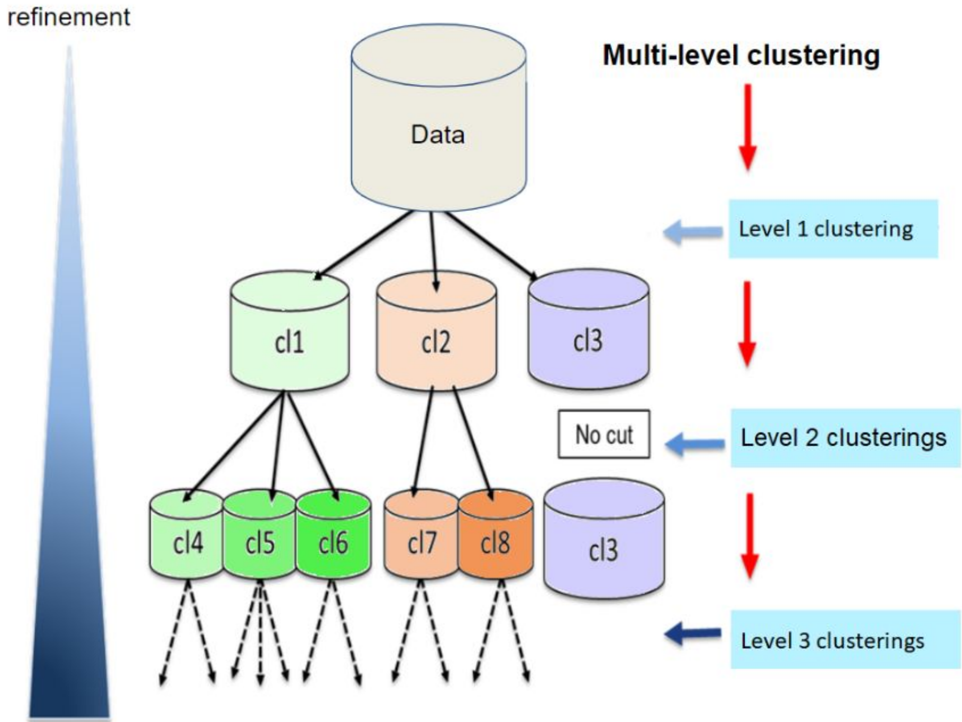
10.1109/JSTARS.2014.2341219

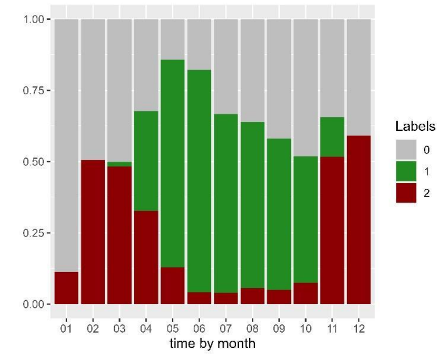
Long-term series segmentation: Marel Carnot application

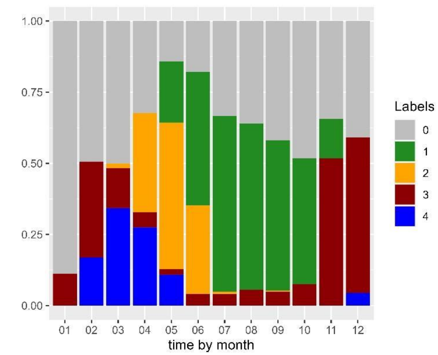


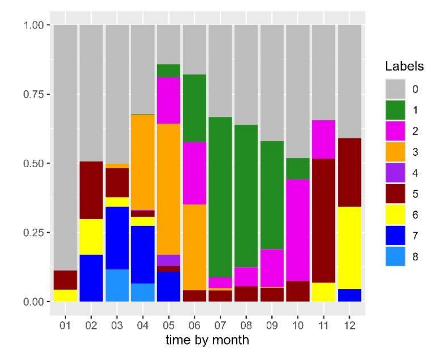
interface in R-package : uHMM and sClust

Multi-Level implicit segmentation: Marel Carnot application.









Grassi phD, 2020
dir. Alain Lefebvre et E. Poisson
10.1109/OCEANSE.2019.8867261

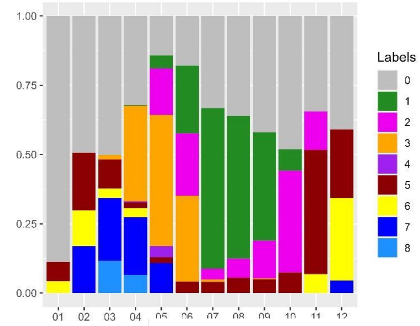
Fuzzy version, controlled cut : Cabotte et al. 2022 [10.5220/0011550800003332](10.5220/0011550800003332)

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

ULCO UNIVERSITÉ DU LITTORAL CÔTE D'OPALE

Explained Diagrams

Phenology and environmental conditions



Summer non-productive

Transition state

Potentially Toxic Bloom

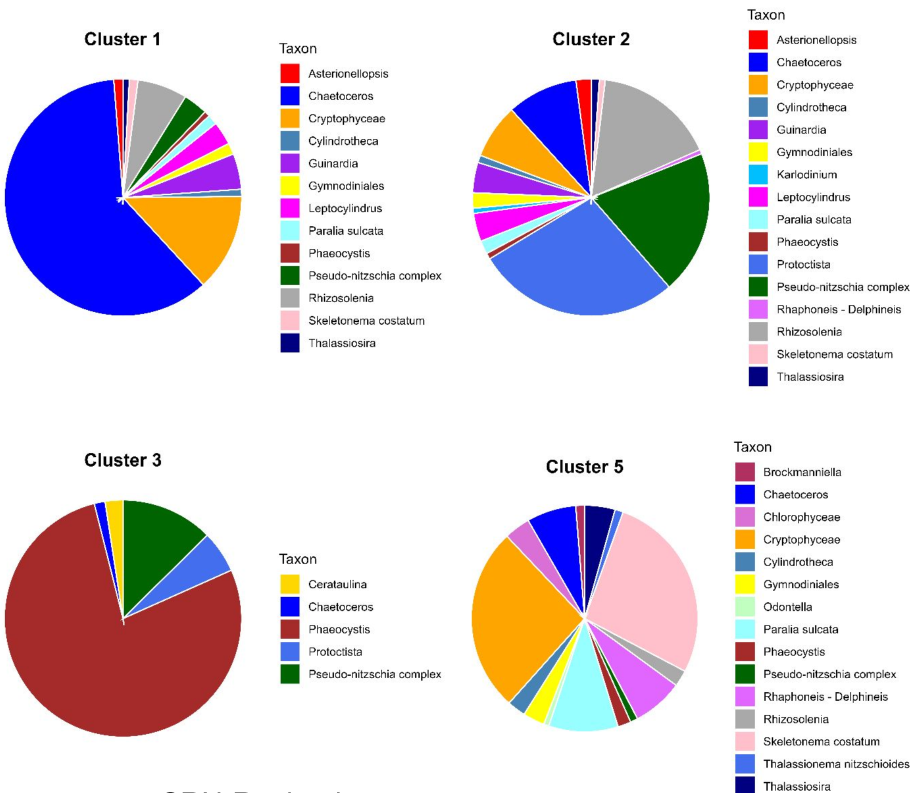
Rare event

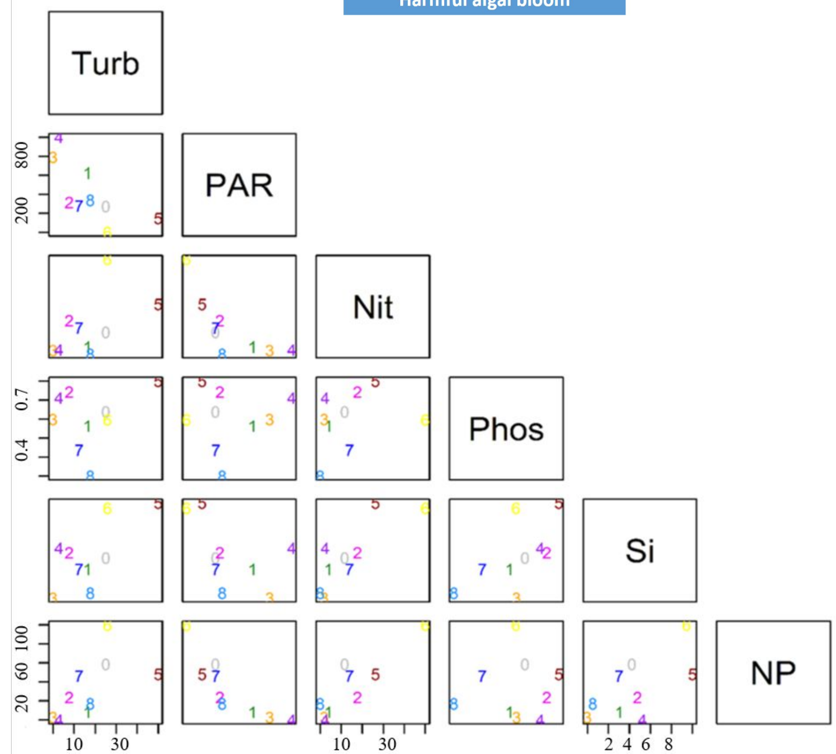
Winter non-productive period

Extreme event

Harmful algal bloom

Harmful algal bloom



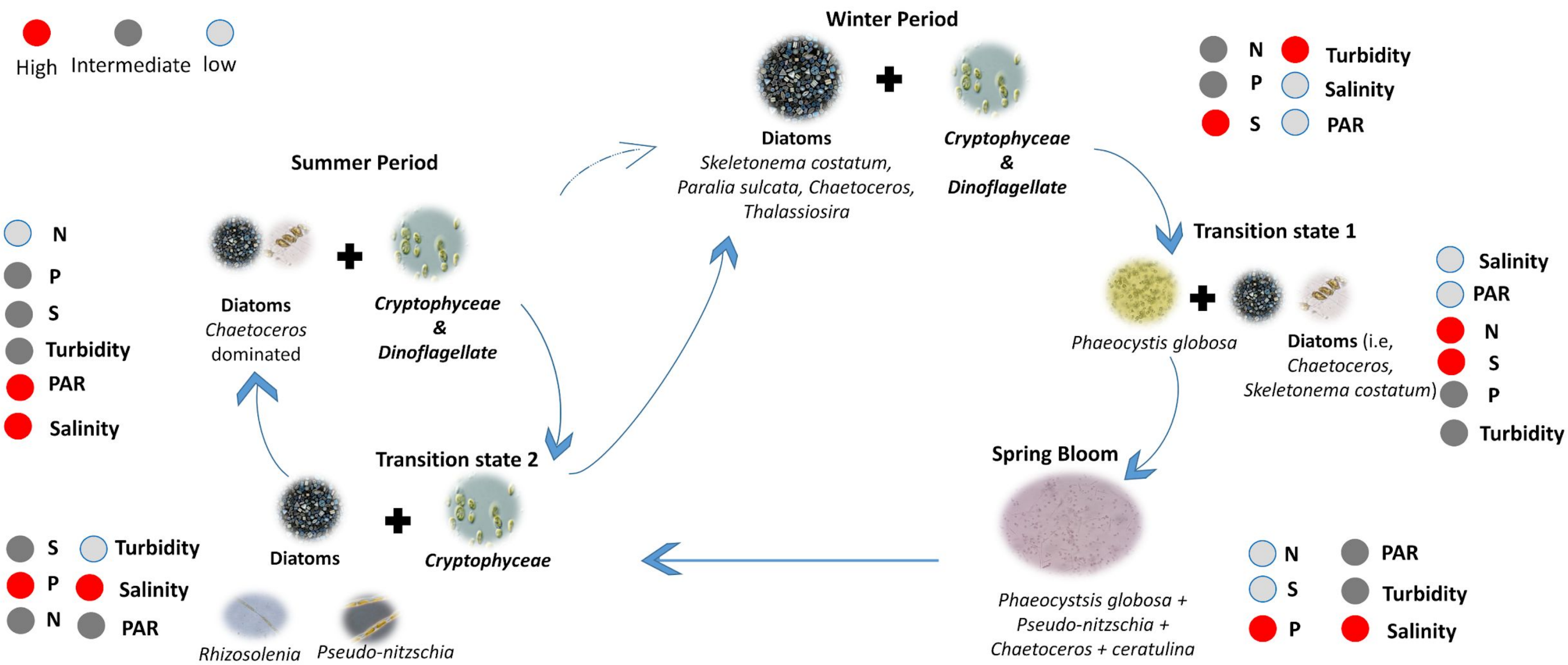


\+ SRN-Rephy data

phD : R. Halawi Ghosn, K. Grassi

Explained Diagrams

Phenology - environmental conditions - phytoplankton diversity



phD : R. Halawi Ghosn

Phytoplankton dynamics in EEC

Multi-Level implicit segmentation: short Time Series application.

CGFS 2018

FerryBox Data

sampling: 1 minute

TABLEAU I – Caractéristiques des variables FerryBox impliquées dans la classification : minima (min), maxima (max), moyenne, médiane et quartiles (Q1-Q3).

Paramètre (unités)	Moyenne-medianne	min-max	Q1-Q3	# % NA
Température (°C)	17,30 - 17,56	13,67 - 18,84	16,97 – 17,86	9,90
Salinité (PSU)	34,67 - 35,01	26,26 – 35,38	34,67 – 35,17	9,90
Turbidité (NTU)	5,67 – 5,20	0,41 – 515,03	1,80 – 8,94	9,90
Oxygène dissous ()	259,3 – 257,5	237,7 – 302,0	252,8 – 263,8	9,90
AOA Algues Vertes ($\mu g\, l^{-1}$)	0,29 – 0,33	0,00 – 1,78	0,00 – 0,57	9,90
AOA Algues Bleues ($\mu g\, l^{-1}$)	0,002- 0,00	0,00 – 5,88	0,00- 0,00	9,90
AOA Algues Brunes ($\mu g\, l^{-1}$)	0,96 – 0,63	0,00 – 9,81	0,09 – 1,30	9,90
AOA Cryptophytes ($\mu g\, l^{-1}$)	0,83 – 0,82	0,00 – 20,18	0,57 – 0,99	9,90

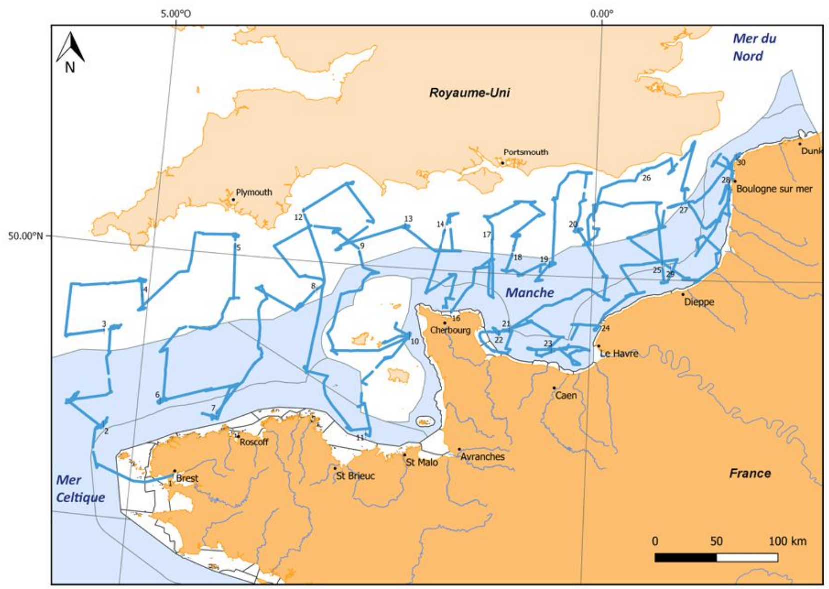


FIGURE 1.1 – Trajet de la campagne CGFS 2018 représenté par les points de mesure du FerryBox du N/O Thalassa. Les numéros le long du parcours représentent les jours successifs de campagne. (carte issue du rapport DCSMM LEFEBVRE et David DEVREKER 2019)

Multi-Level implicite segmentation : short Time Series application.

CGFS 2018

FerryBox
Data

sampling: 1
minute

Automatic
clustering by
MSC
vs
SHOM labeling

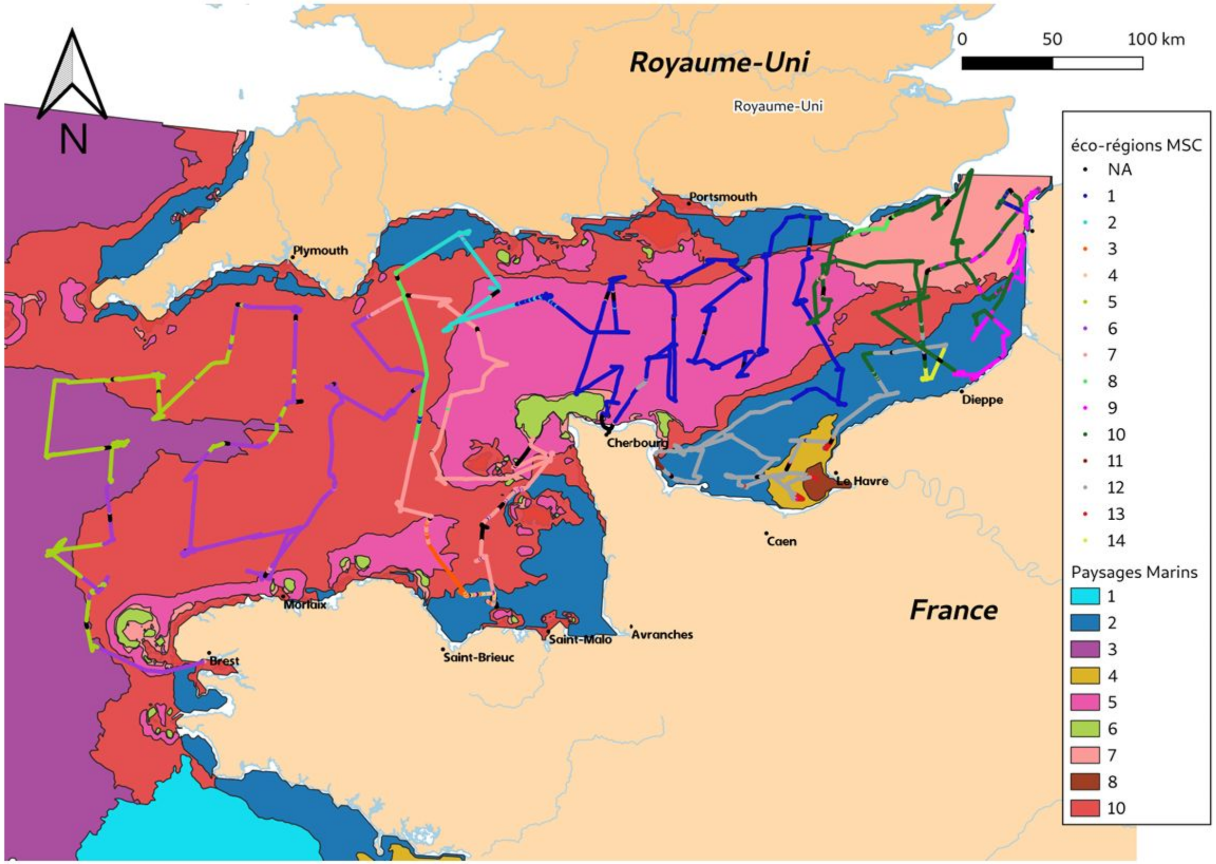


FIGURE 1.3 – Les 14 éco-régions définies par Classification Spectrale Multi-niveaux (M-SC) transposées sur le trajet de la campagne GCFS 2018 et superposées aux 10 paysages marins definis par le descripteur 7 (Changements Hydrographiques) du SHOM.

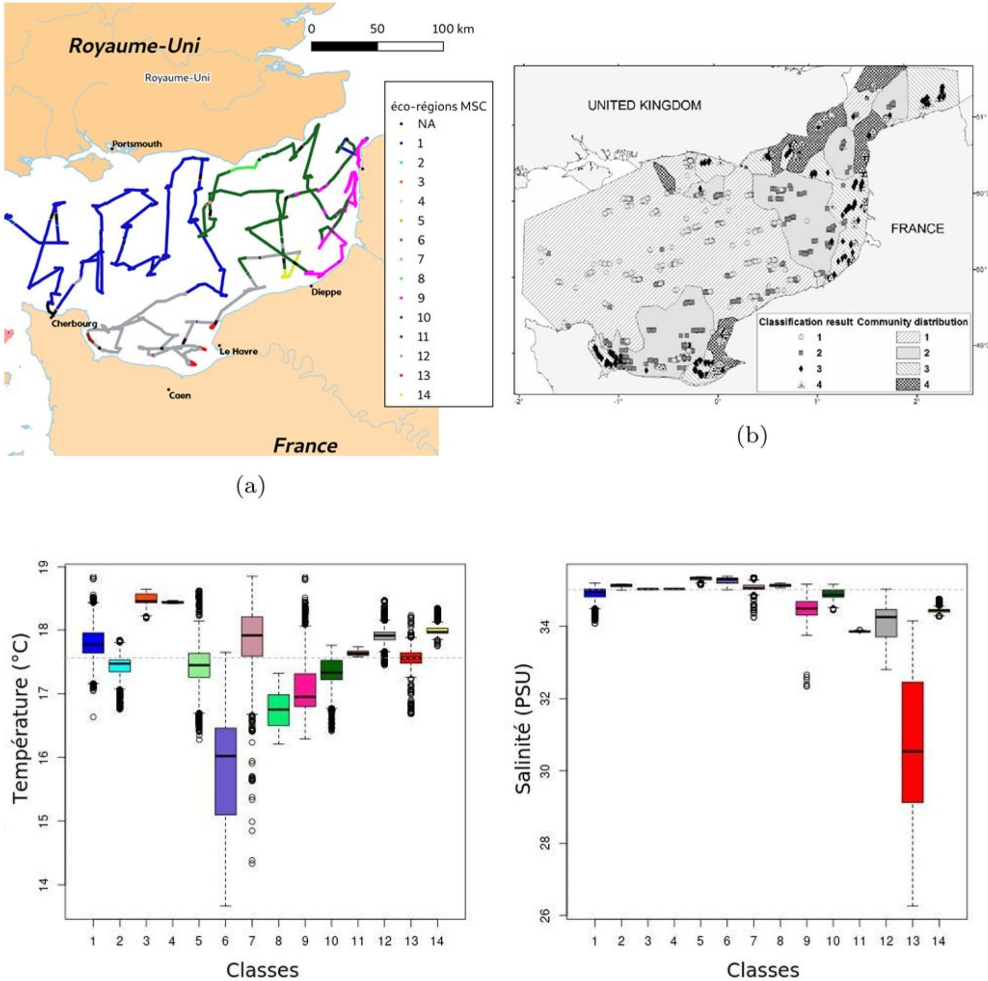
Multi-Level implicite segmentation: short Time Series application.

CGFS 2018

FerryBox Data

sampling: 1 minute

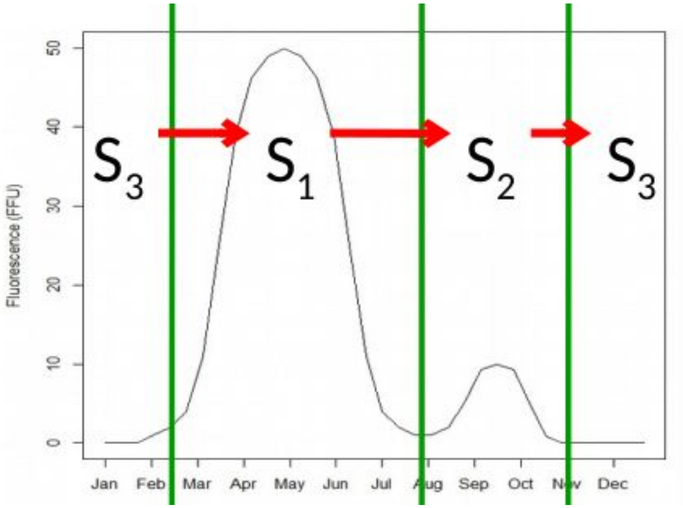
Automatic clustering vs Fish community assembly



(a) (b) (c) (d)

FIGURE 1.7 – Classification au niveau 4 des données de la campagne CGFS 2018 en Manche Est sur la période de septembre à octobre. (a) répartition spatiale des classes ; (b) répartition des 4 assemblages de communautés benthiques. ; Boîte de Tukey par classe (c) de la température (°C) et (d) la salinité (PSU).

2- Forecasting - short/middle/long prediction

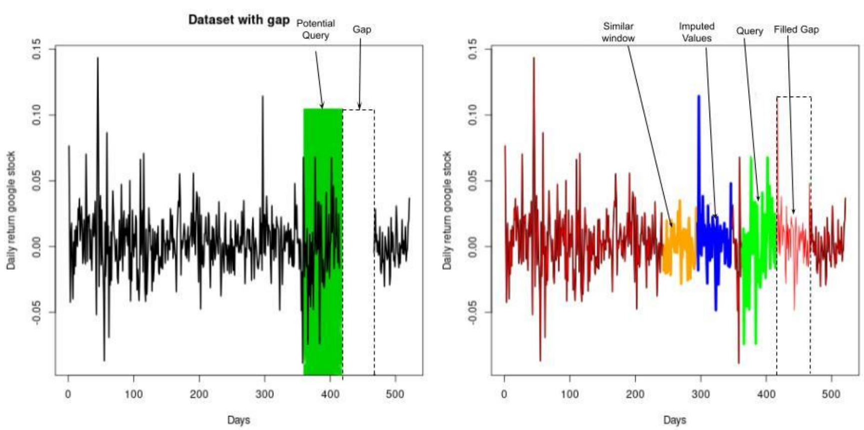


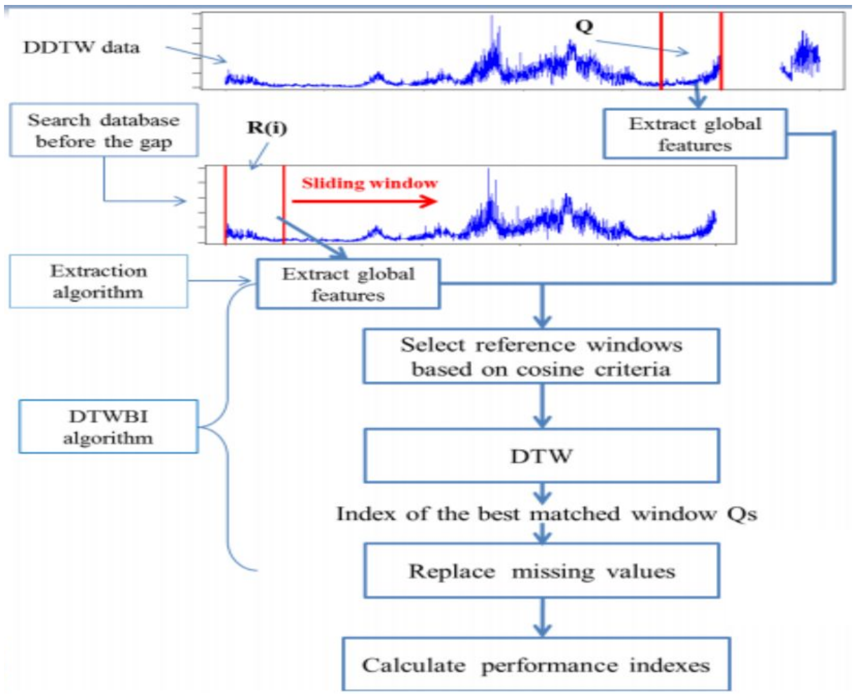
? states or conditions

now
tomorrow
week-forecasting
scenario

Short-prediction using Pattern retrieval

Approaches : DTBWI, DTWUMI, FSMUMI (packages R)
input : N-previous window of each EOV
output : N-Future values of these EOV





Elastic matching for classification and modelisation of incomplete time series

dir. E. Poisson et A. Bigand

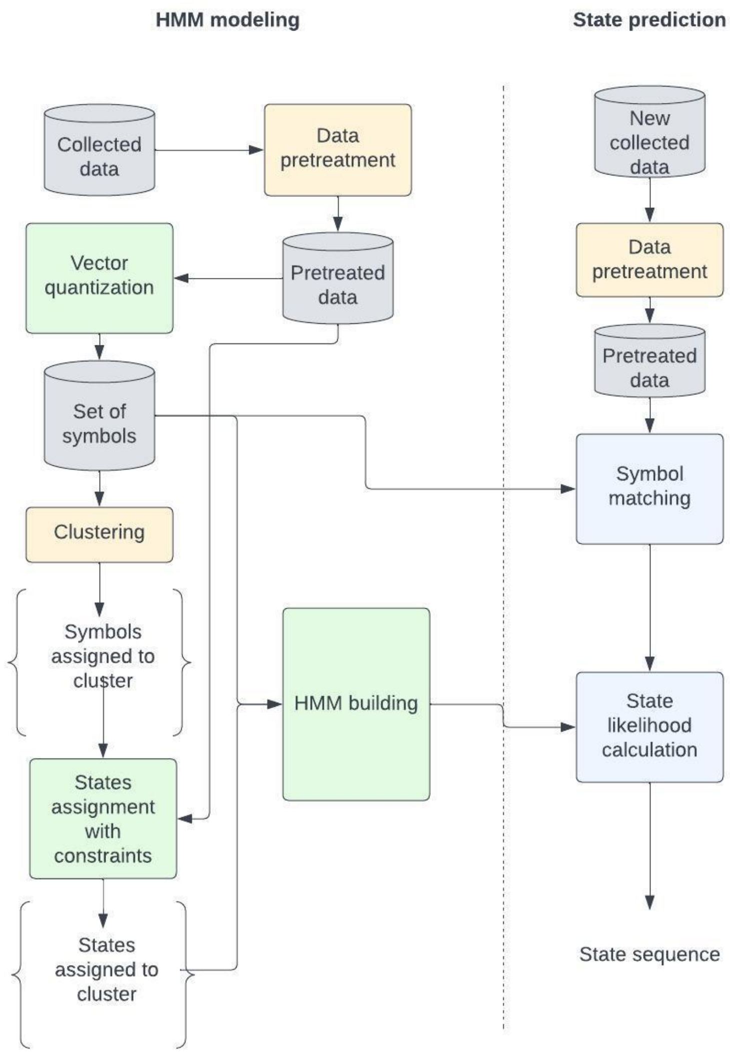
https://doi.org/10.1016/j.patrec.2017.08.019

Drawbacks : no previous pattern, NA data
Others directions : ConvLSTM, TimeGAN

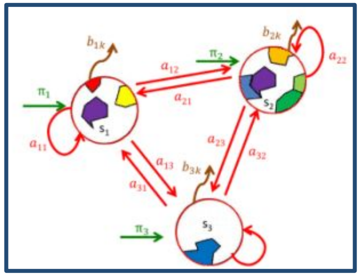
Short-state prediction using uHMM models

HMM modeling

1. Vector quantization
2. Clustering
3. State assignment
4. HMM building



Hidden markov model



State prediction

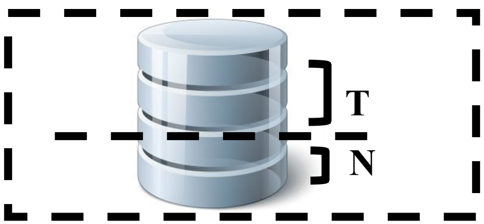
1. Symbol matching
2. State likelihood calculation

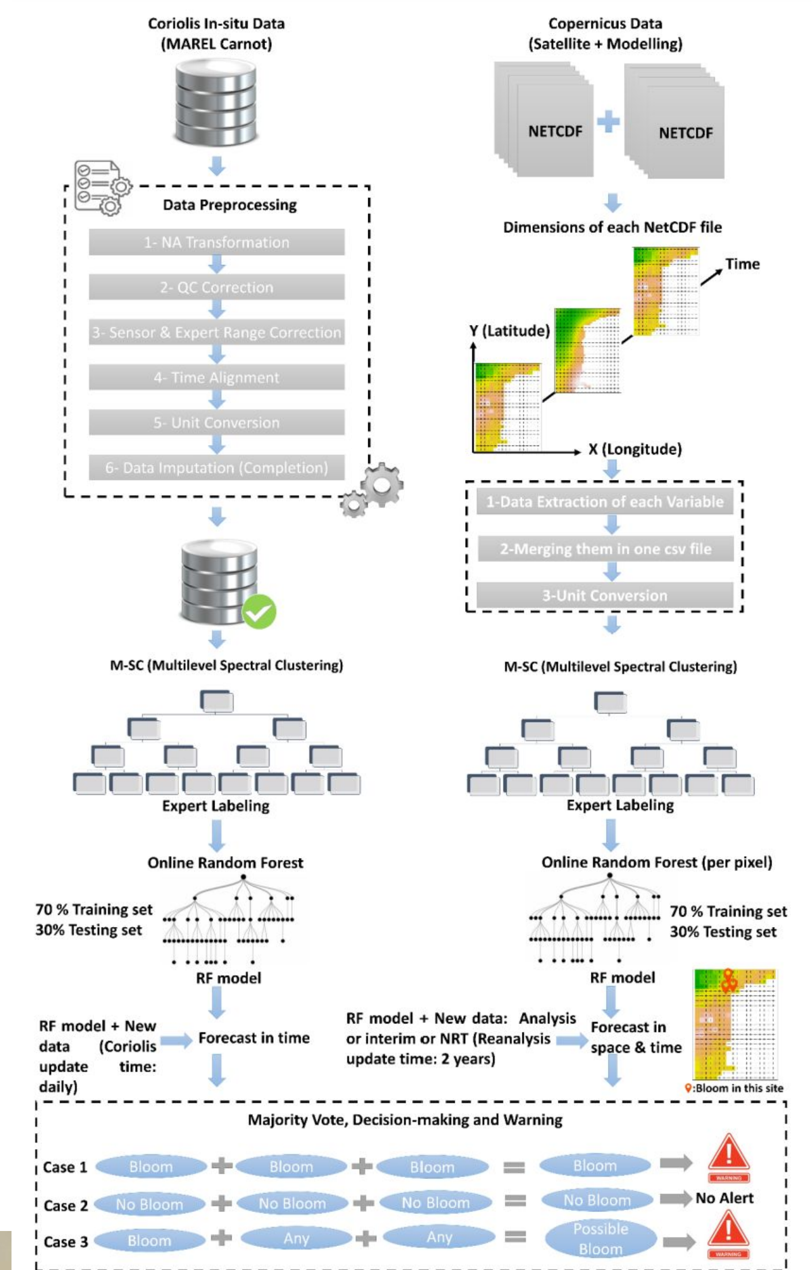
Kevin Rousseuw, phD 2023 uHMM
Aziz Allouche, phD in progress. nov 2023-..
Time dynamics Constraint in segmentation and prediction

Short-term and scenario: state prediction

Incremental Random Forest

- input : T- past instant windows
  - 2, 4, 7, 10, 30 days
  - up to 365 days for modelling and satellite data
- Output : N: forecasting future labels
  - (2, 4, 7, 10 days)
- Evaluation metrics:
  - accuracy, No information rate NIR, Kappa, recall, F1 score



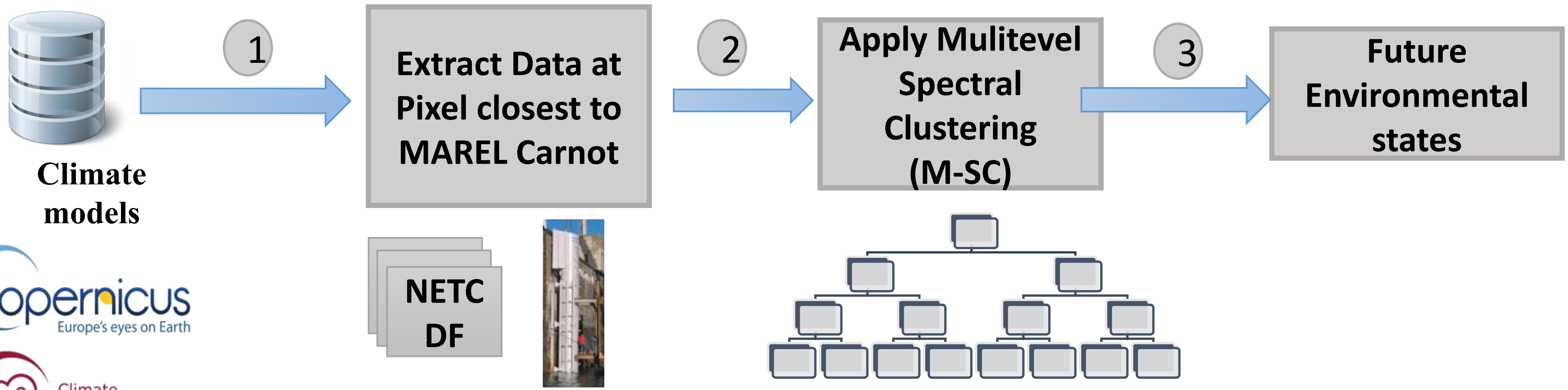


Raed Halawi Ghosni
phD, 2024

dir. Alain Lefebvre et E. Poisson

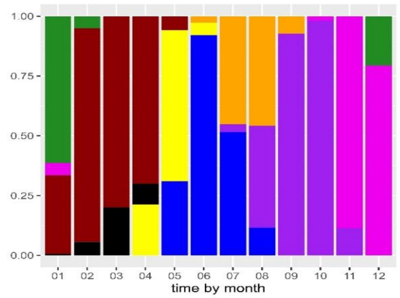
Short-term and scenario: state prediction

- Long-term
- Gain insight on possible future environmental states, particularly HABs
- Climate model/ scenario RCP (4.5): greenhouse gases emissions peak around 2040

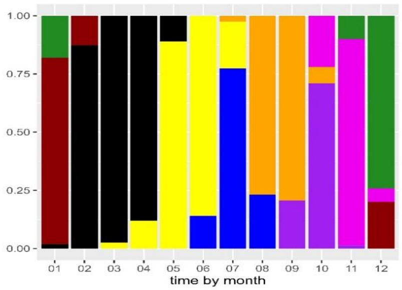


Daily variables: Chl-a, DO, Temperature, Salinity
Years Chosen:
2018-2022: validation point
2038-2042: peak of green house gases emission
2095-2099: end of century

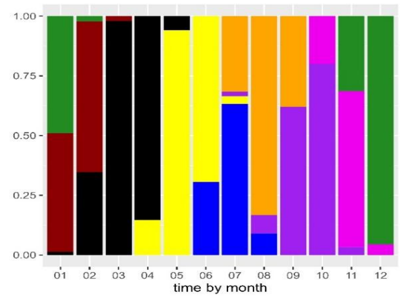
2018-2022



2038-2042



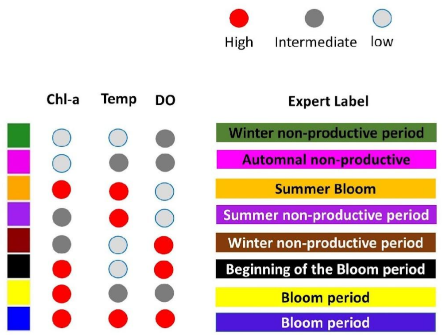
2095-2099



Beginning of the Bloom → **Uncommon Winter Bloom**

the Bloom → **Higher Frequency & longer duration**

Non-productive periods → **Changes in emission probabilities**



Machine learning:
emilie.poisson@univ.littoral.fr (PU, 61)
pierre-alexandre.hebert@univ-littoral.fr (MCF, 61)

Data, eutrophisation, phytoplankton:
alain.lefebvre@univ-littoral.fr (CR, 67)
guillaume.wacquet@univ-littoral.fr (CR, 61)