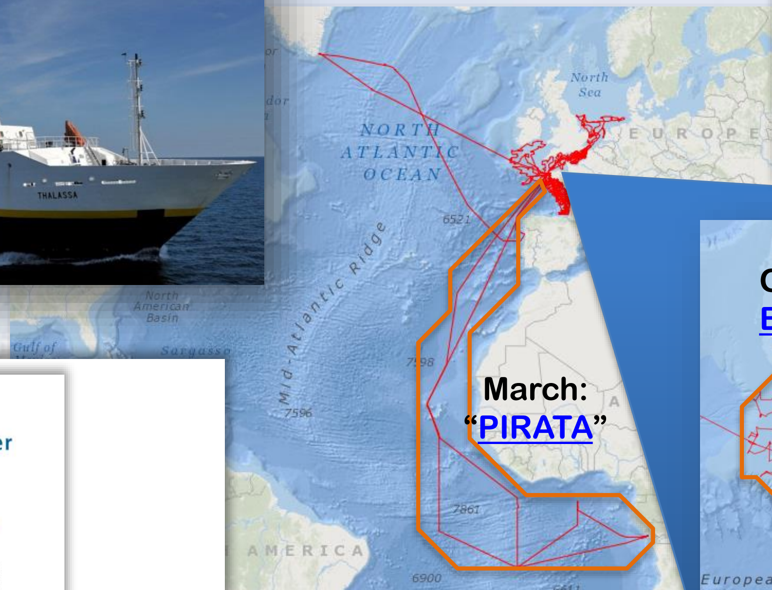
Identification of spatial hydro-biological structures by spectral clustering. Towards implementation of machine learning for Ferry Box data processing.

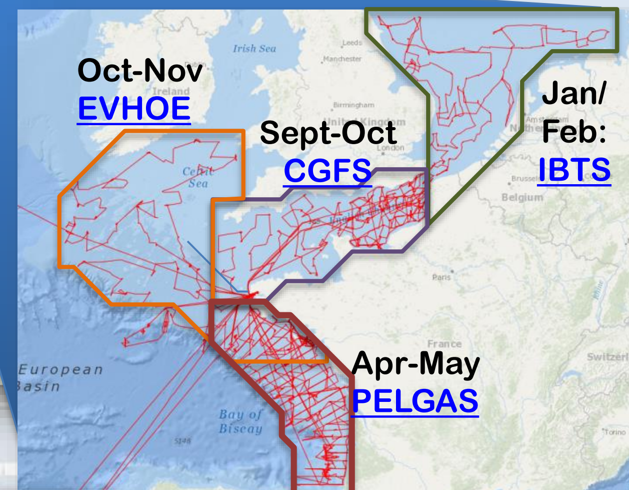
Lefebvre A[1]., Grassi K.[1,2,3], Crenan B.[4], Poisson-caillault E.[3]

[1] Ifremer, Boulogne sur Mer, France
[2] WeatherForce, Toulouse, France
[3] LISIC, EA 4491, Université du Littoral Côte d'Opale, Calais, France
[4] Ifremer, NSE, Brest, France

RV Thalassa's Ferrybox

Manufacturer:	-4H-Jena, "Ferrybox I"
Thermosalinograph	SBE45 + SBE21
Oxygen:	Anderaa 4835
Turbidity	Seapoint
Fluorescence	BBE-AOA
pH:	Meinberg MV3010 (for tech. purpose)






Ifremer
N/O Thalassa
Guide d'utilisation et de gestion de la Ferrybox

IFREMER's RV Thalassa FB Guideline available here:
https://doi.org/10.13155/59685
143 pages… But… Only !

Thalassa's path in 2018: all European cruises are renewed year after year since 20 years at the same period on the same transect, and it will bring a huge FB data set in the coming years

In June 2019, Thalassa will sail in Mediteranean sea (MOOSE cruise)

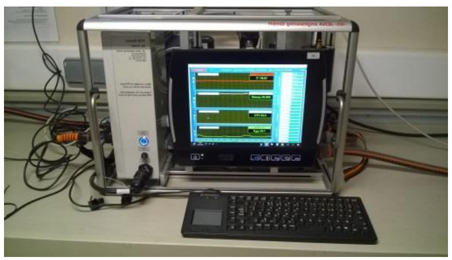
2018 statistics:
Thalassa sailed **~245 days**
The Ferrybox was running **229 days**
Operational ratio: **93%** !

STRATEGY AND OBJECTIVES

Ferry Box Database
In-situ (HF)

- High Frequency (HF) automated measurement system

→ Large dataset, missing data …



Pocket Ferry Box RV Europe (Mediterranean sea)

Ferry Box RV Thalassa (Channel, Atlantic,…)

Coriolis Database

Data processing

Stakeholder decision

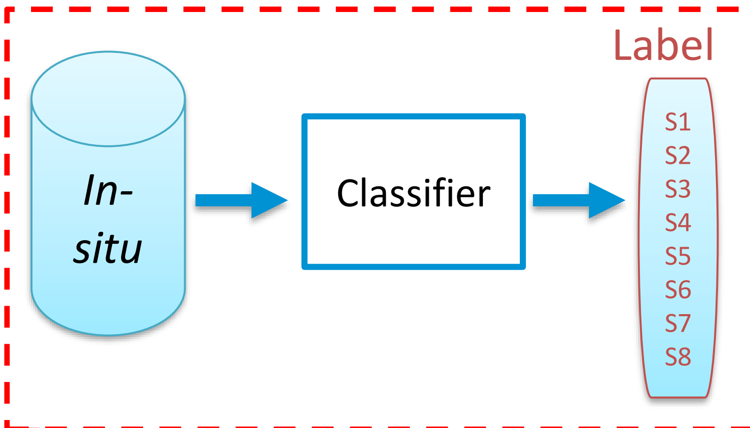
Key scientific questions

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

- How to optimize **data processing** and results interpretation?
- How **to detect and characterize environmental states** in multi-parameters time series?
- How to identify **frequent, rare or extreme events** and their **dynamics in time series**?

Towards implementation of **machine learning** for Ferry Box data processing

YES but
first step = **Labelling** via **Classification**



Examples of labels:
Bloom, incl. HAB
Oxygen deficiency
Nutrient impulse
…

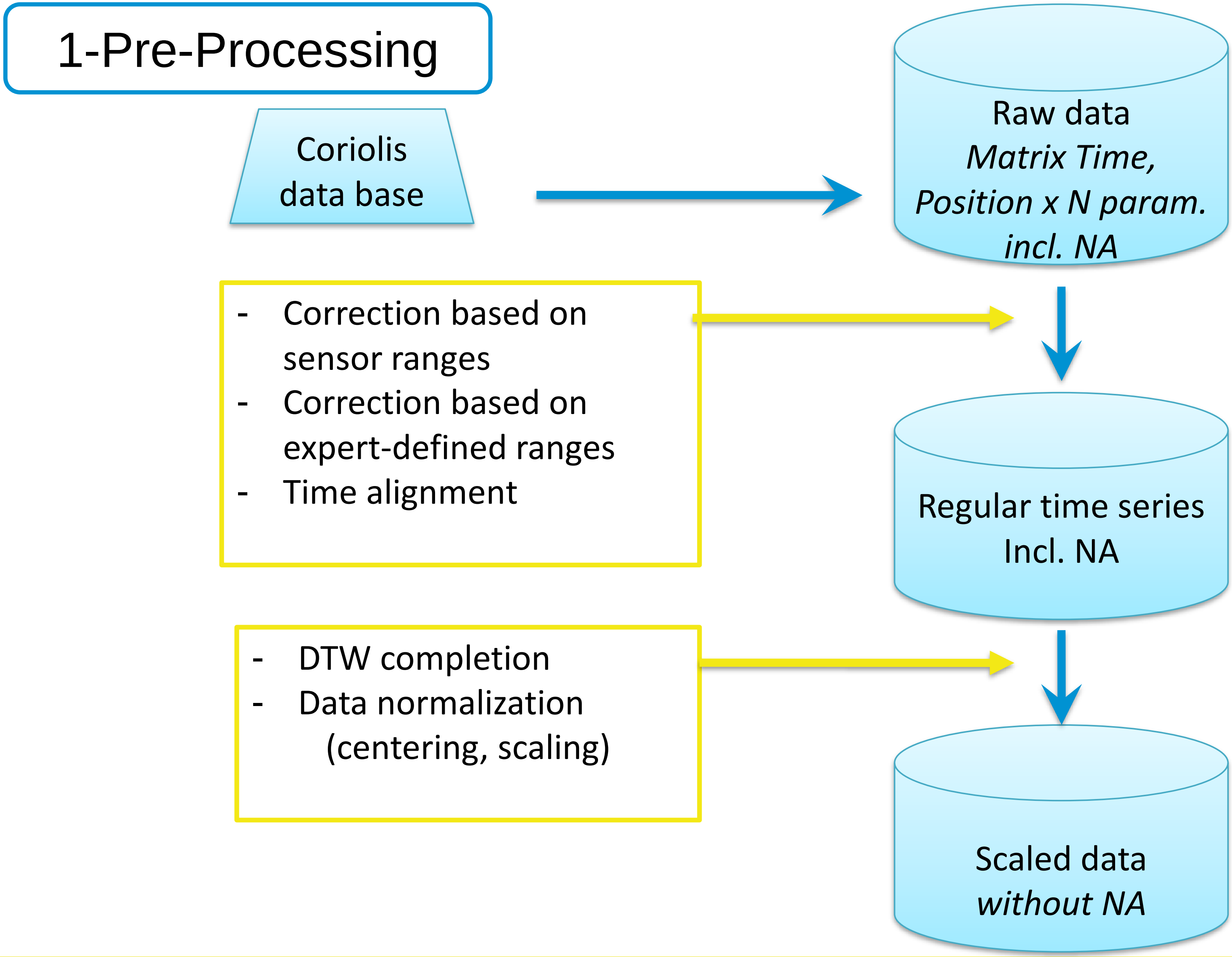
To learn

The spectral classification allowed to :

- Define **environmental states** in multi-parameter time series
- Detect, identify in time and space and **characterize states dynamics**
- **Extract <u>label</u> for frequent, rare or extreme events**

PROTOCOL

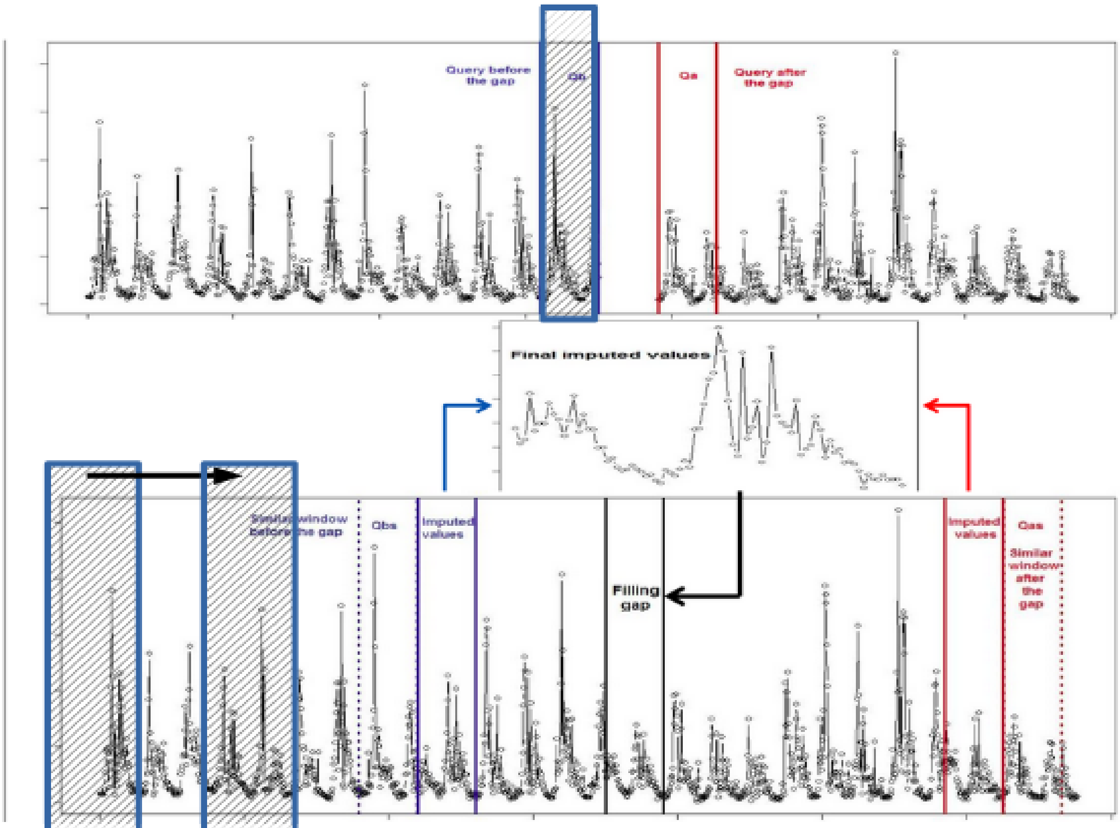
Ifremer



DTWBI Algorithm

Missing interval imputation (ICCE'2016, OCEANS 2017).

- DTW query building (two ways : mono/multivariate series)

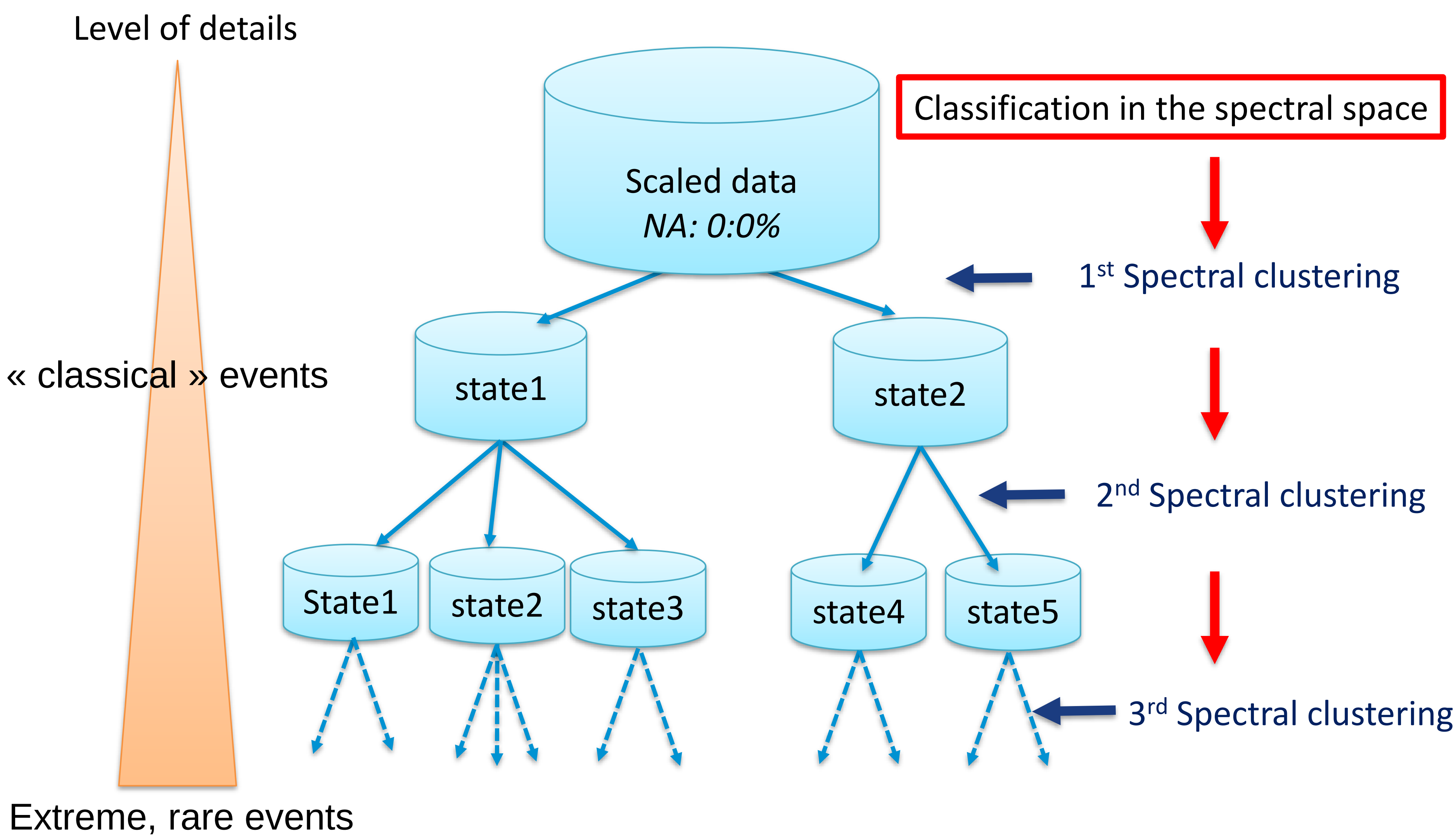


W_request : sliding window

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.

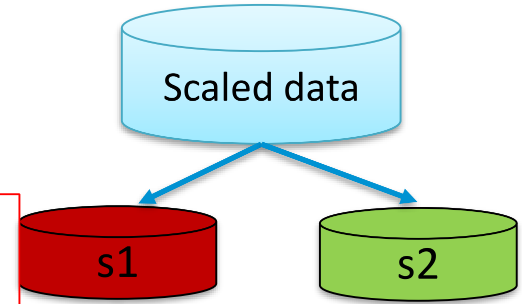
**Phan H., et al 2017: Pattern Recognition Letters*

2- Processing : Multi-level Spectral clustering



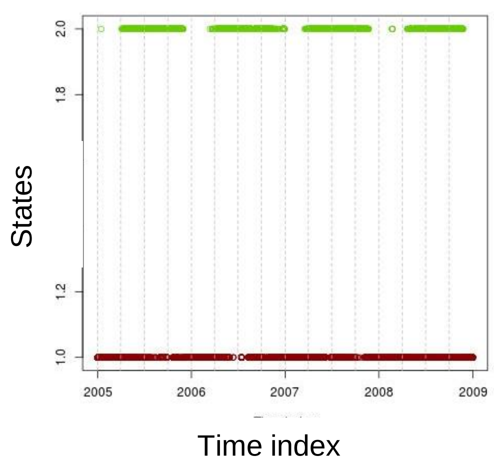
STATES DYNAMIC AND MAIN CONTRIBUTING PARAMETERS

1st Spectral clustering

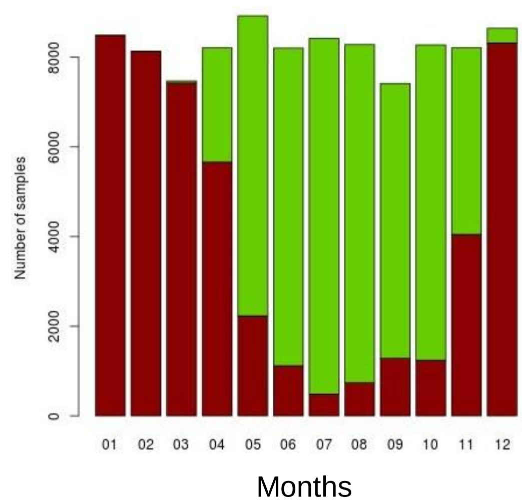


2 states: General Seasonal dynamics:
winter/automn vs. spring/summer

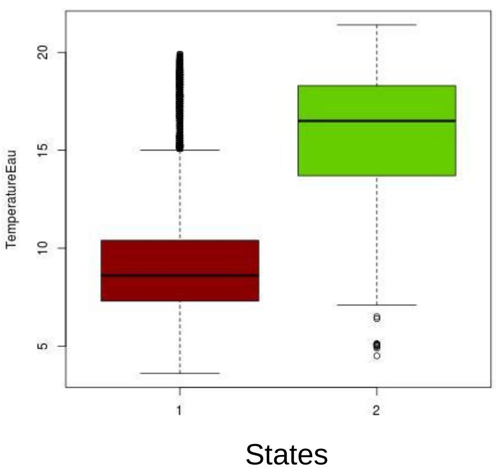
Dynamics



Frequency of states by months



Temperature

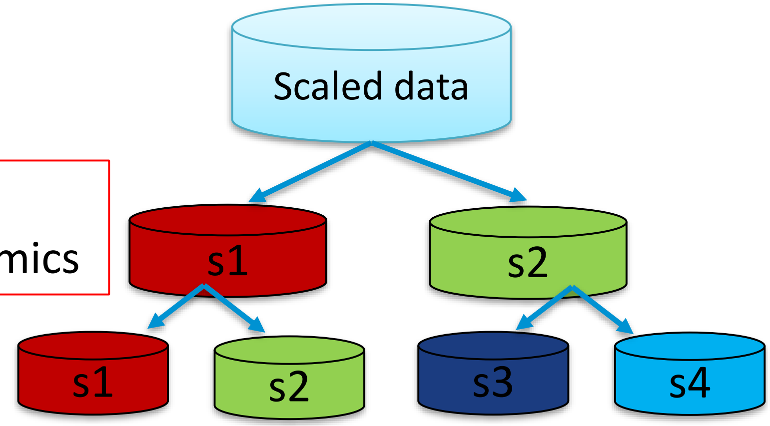


Correlation of each parameter for a given cluster :

	Salinity	Turbidity	Temperature	Dissolved Oxygen	Nitrate	Phosphate	Silicate	PAR	Sea Level
S1	-0.35	0.30	**-0.73**	0.52	0.38	0.21	0.38	-0.21	0.014
S2	0.35	-0.30	**0.73**	-0.52	-0.38	-0.21	-0.38	0.21	-0.014

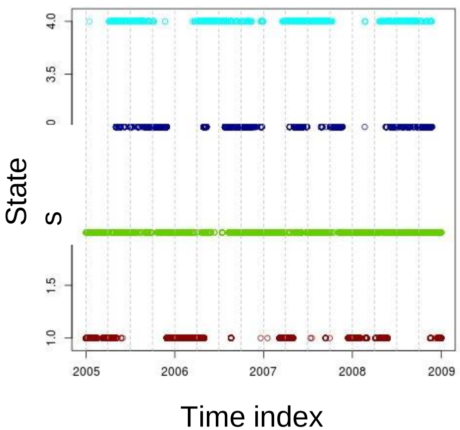
STATES DYNAMIC AND MAIN CONTRIBUTING PARAMETERS

2nd Spectral clustering

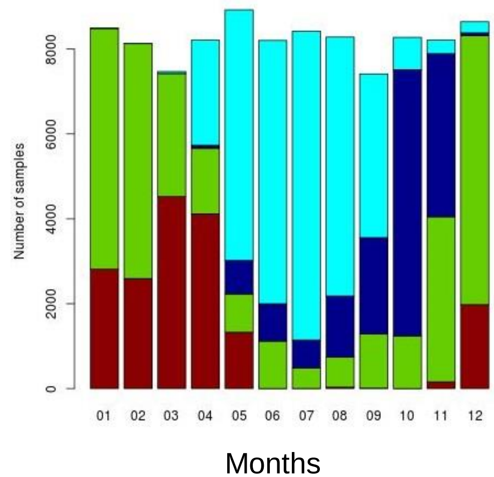


- New structuring variables : Oxygen, Nitrate, Silicate
- Accurate characterization of the spring bloom dynamics

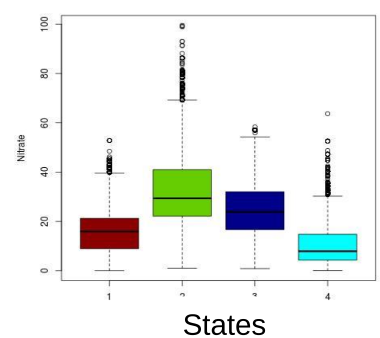
States dynamics



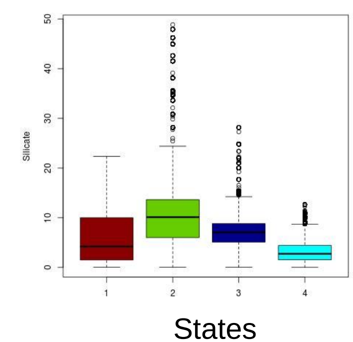
Frequency of states by months



Nitrate



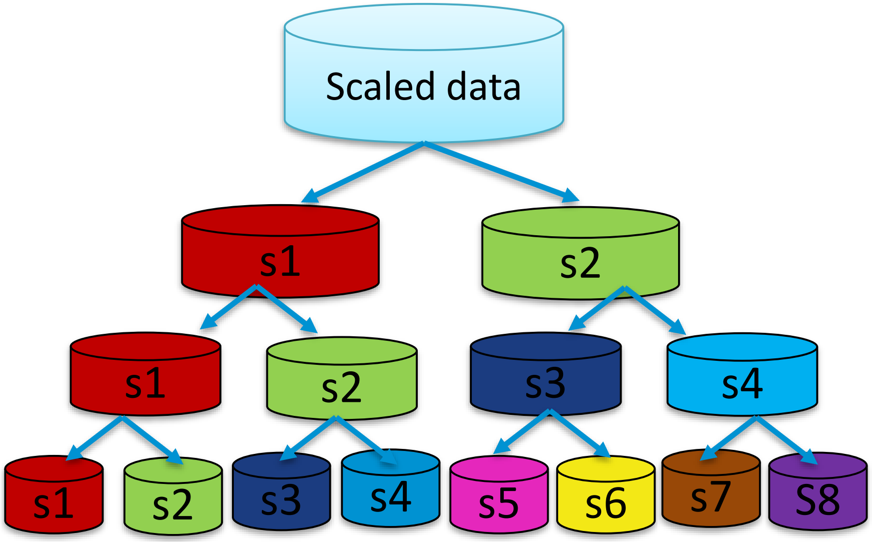
Silicate



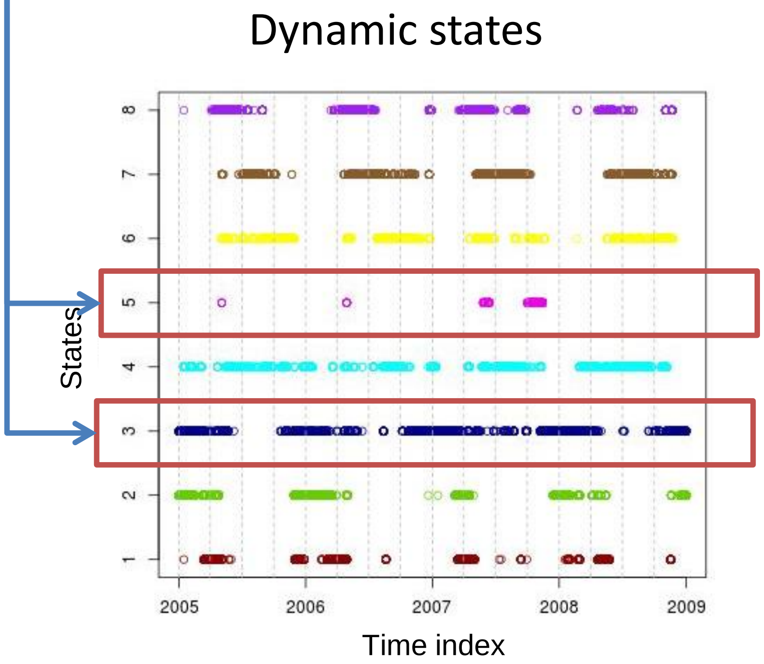
Correlation of each parameter for a given cluster :

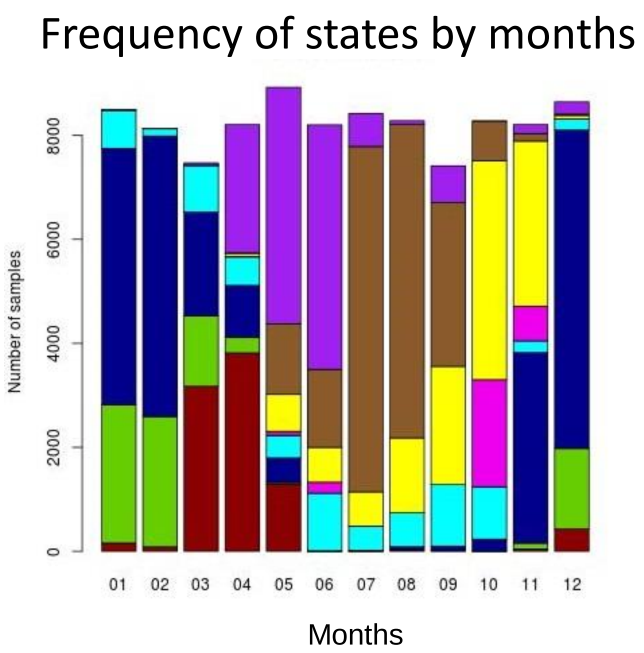
	Salinity	Turbidity	Temperature	Dissolved Oxygen	Nitrate	Phosphate	Silicate	PAR	Sea Level
S1	0.04	-0.08	-0.48	0.62	-0.16	-0.14	-0.06	-0.09	0.02
S2	-0.41	0.40	-0.39	0.05	0.53	0.34	0.47	-0.15	-0.002
S3	0.30	-0.11	0.30	-0.46	0.11	-0.02	0.02	-0.05	0.009
S4	0.13	-0.23	0.53	-0.19	-0.48	-0.19	-0.42	0.26	-0.02

3rd Spectral clustering

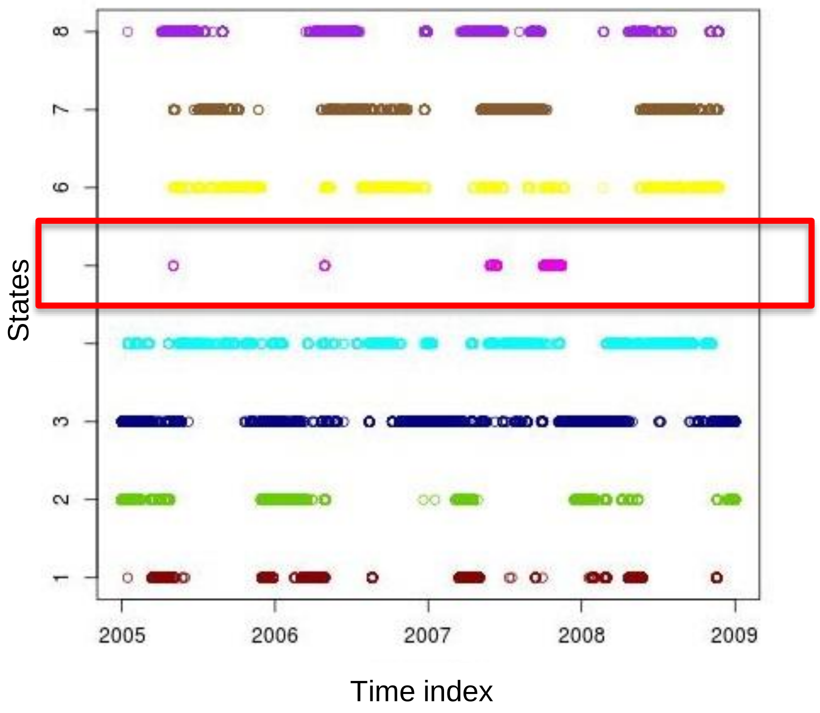


8 states with 2 different dynamics:
Regular (blue) vs. rare events (pink)

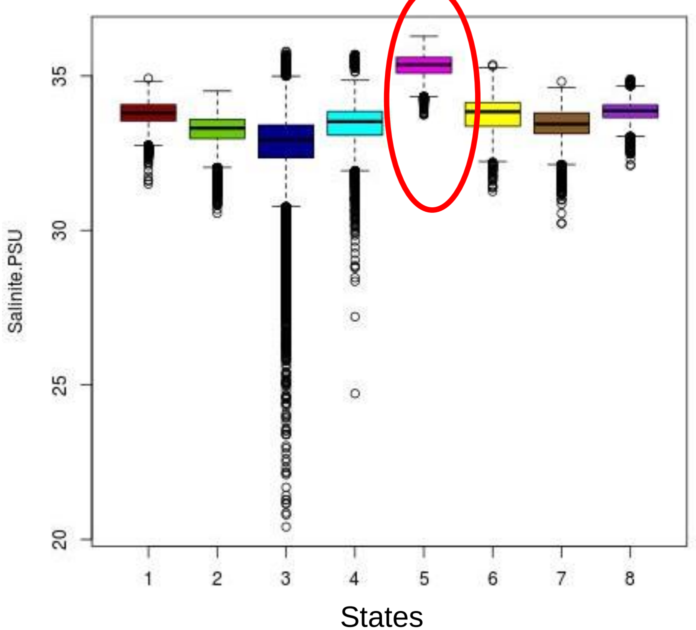


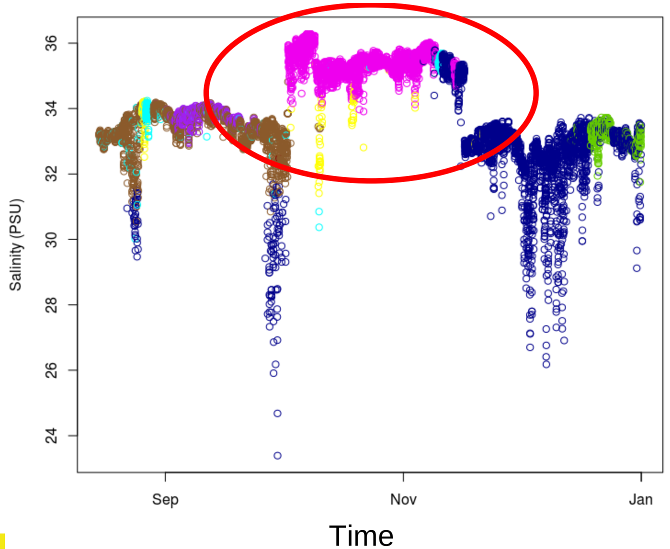


3[rd] Spectral clustering



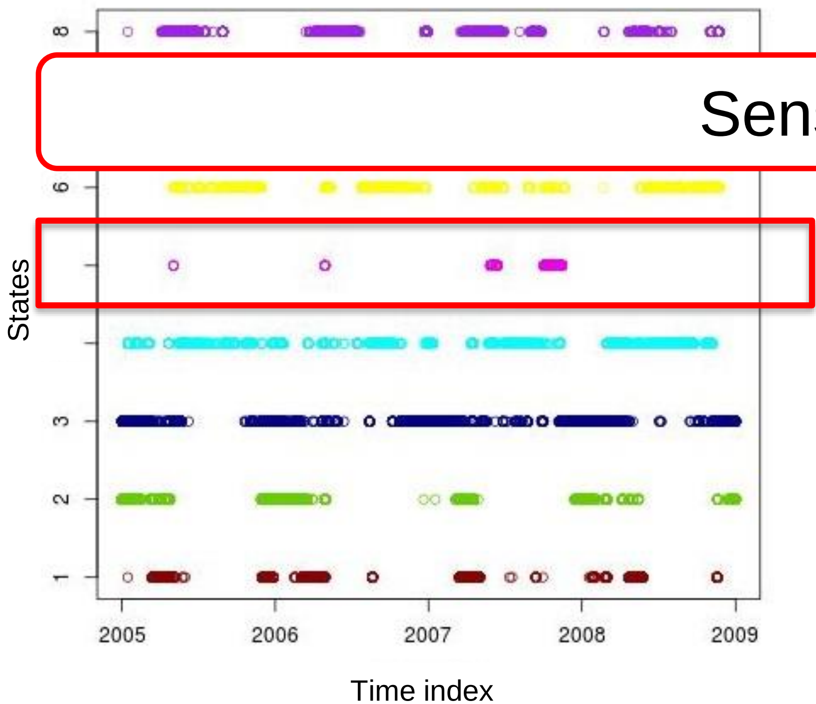
Salinity



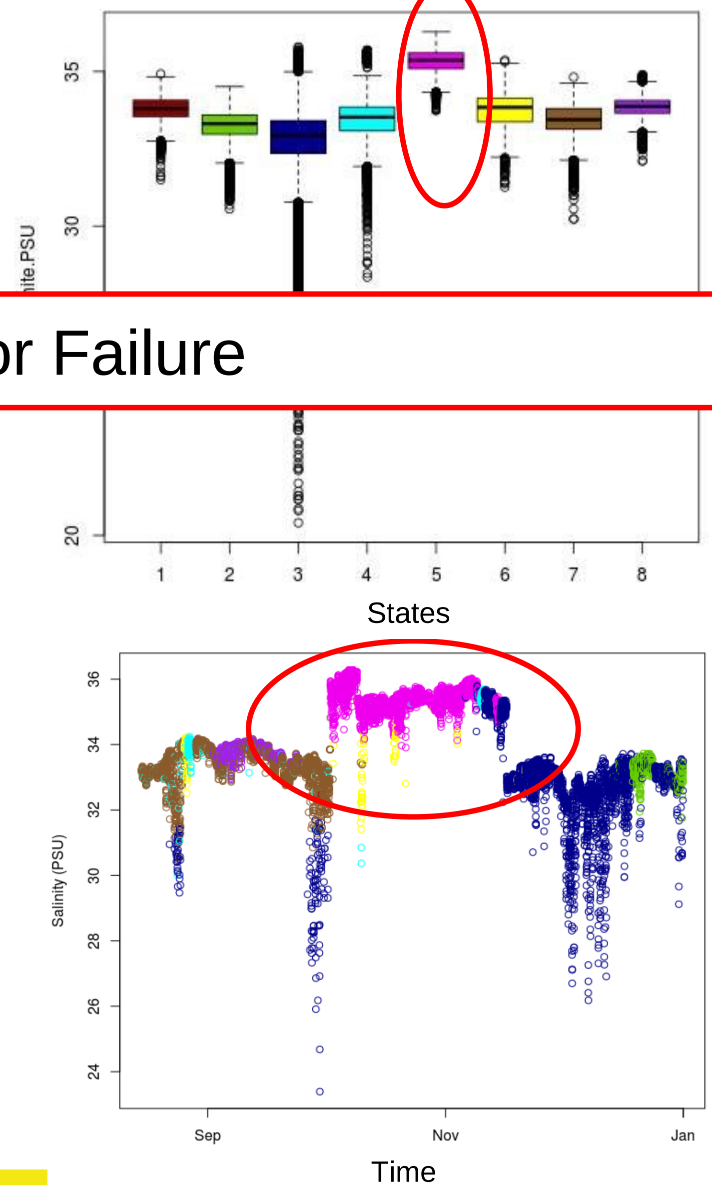


EXAMPLE OF STATES LABELISATION

3rd Spectral clustering



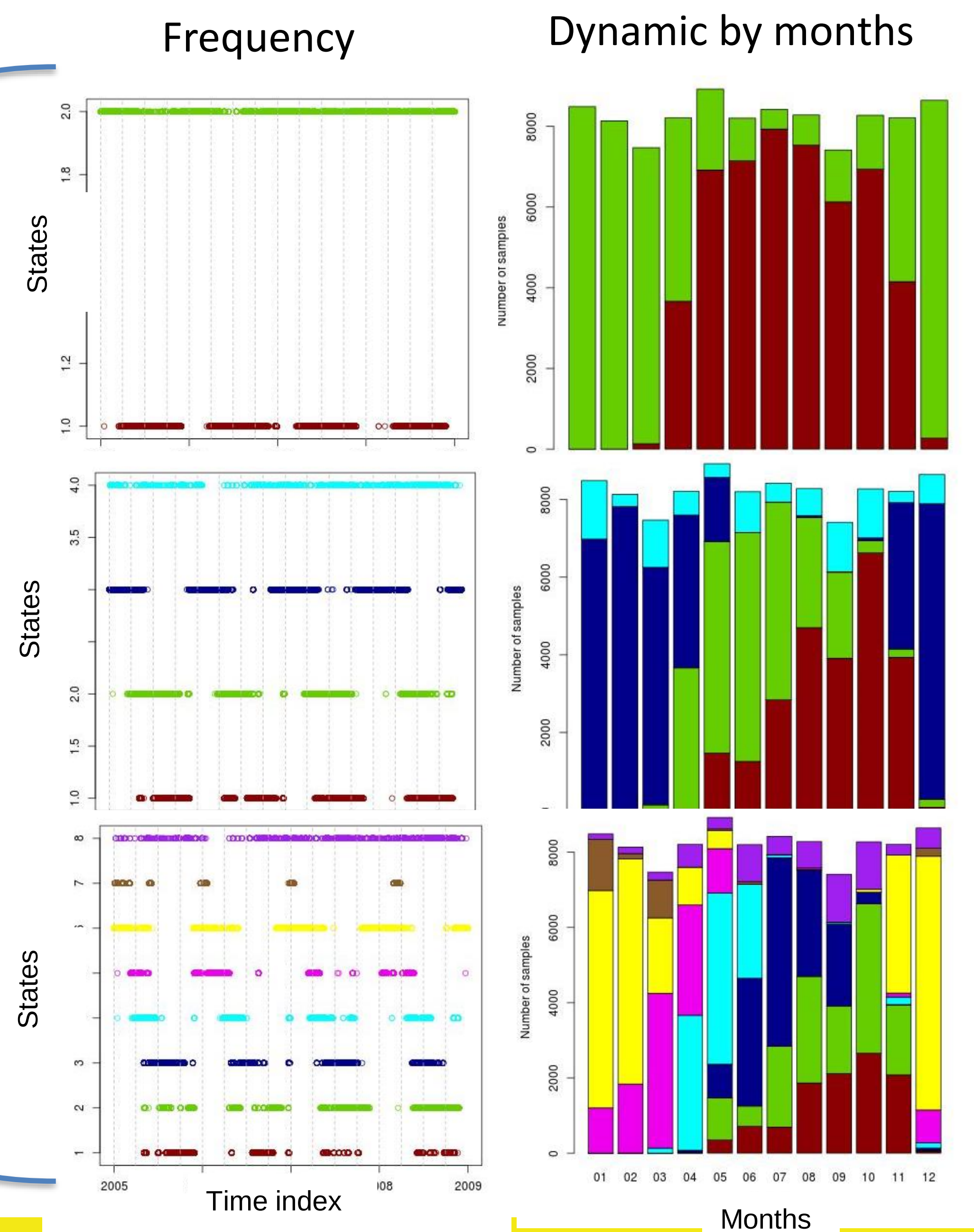
Sensor Failure



STATES DYNAMIC AND MAIN CONTRIBUTING PARAMETERS

Ifremer

After correction sensor failure



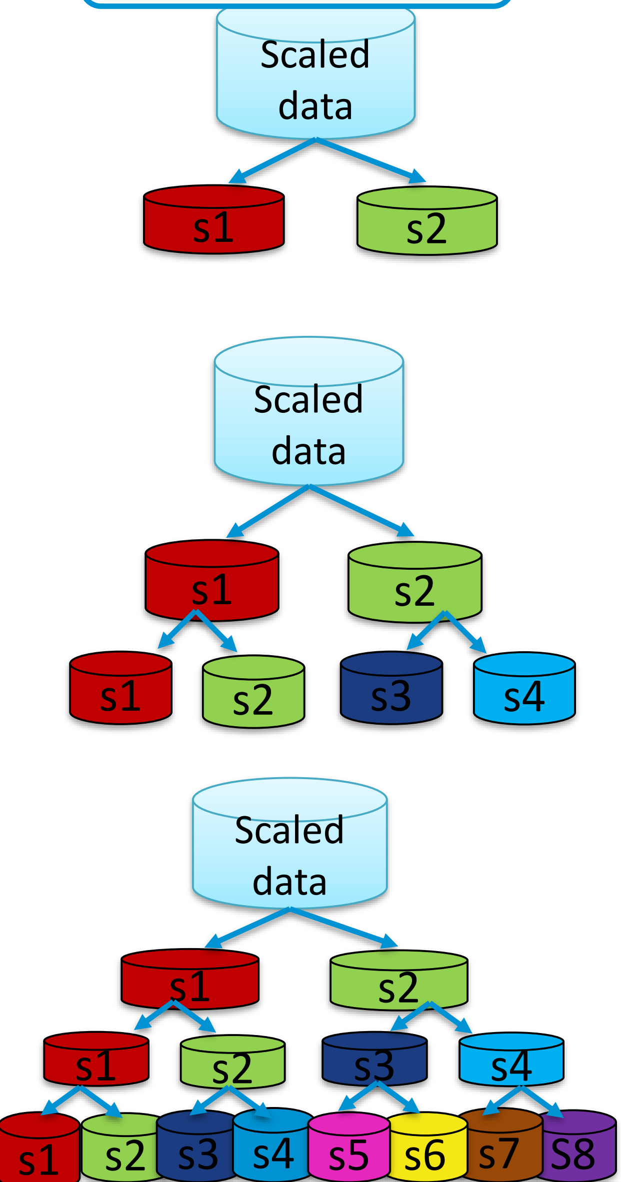
1st Spectral clustering

2nd Spectral clustering

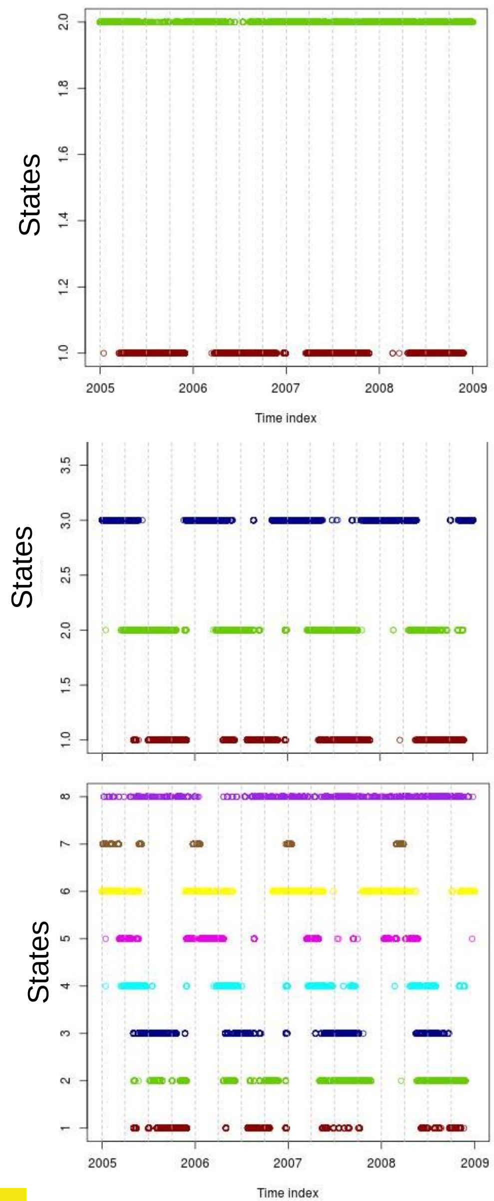
3rd Spectral clustering

STATES DYNAMIC AND MAIN CONTRIBUTING PARAMETERS

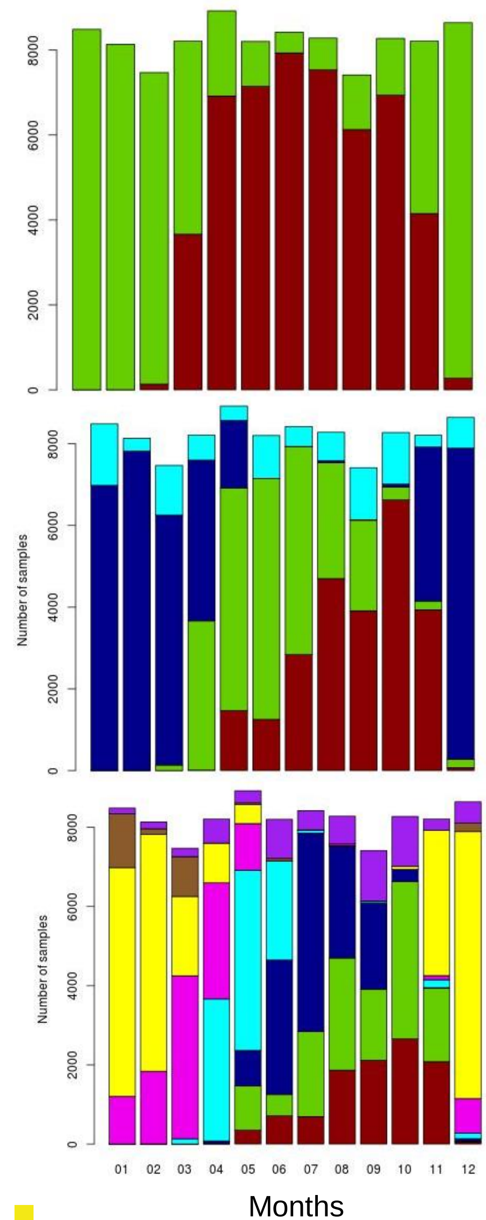
Results



Frequency



Dynamic by months



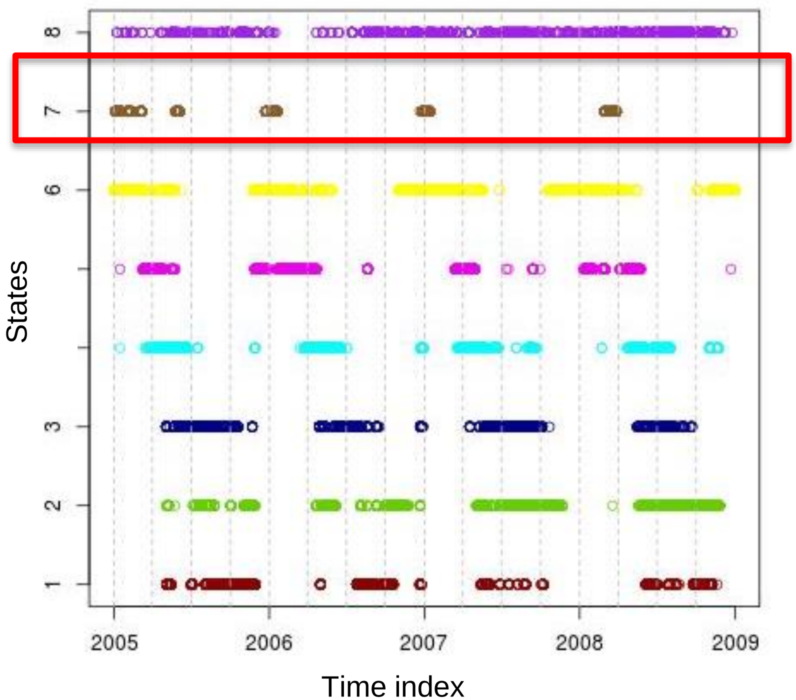
1st Spectral clustering

2nd Spectral clustering

3rd Spectral clustering

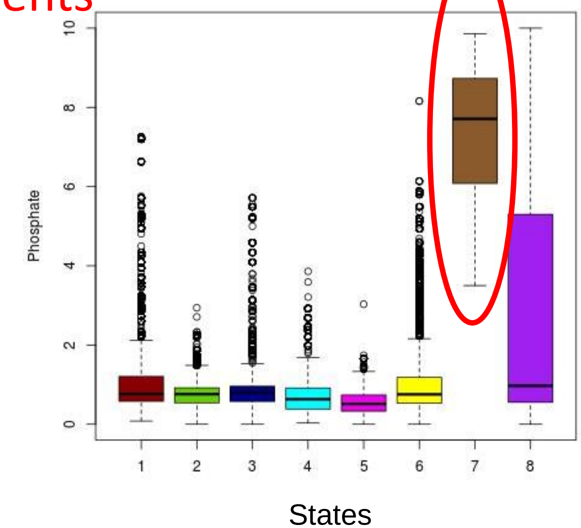
Intermittent Events : rare/extreme

3rd Spectral clustering



Rare/Extreme events

Phosphate

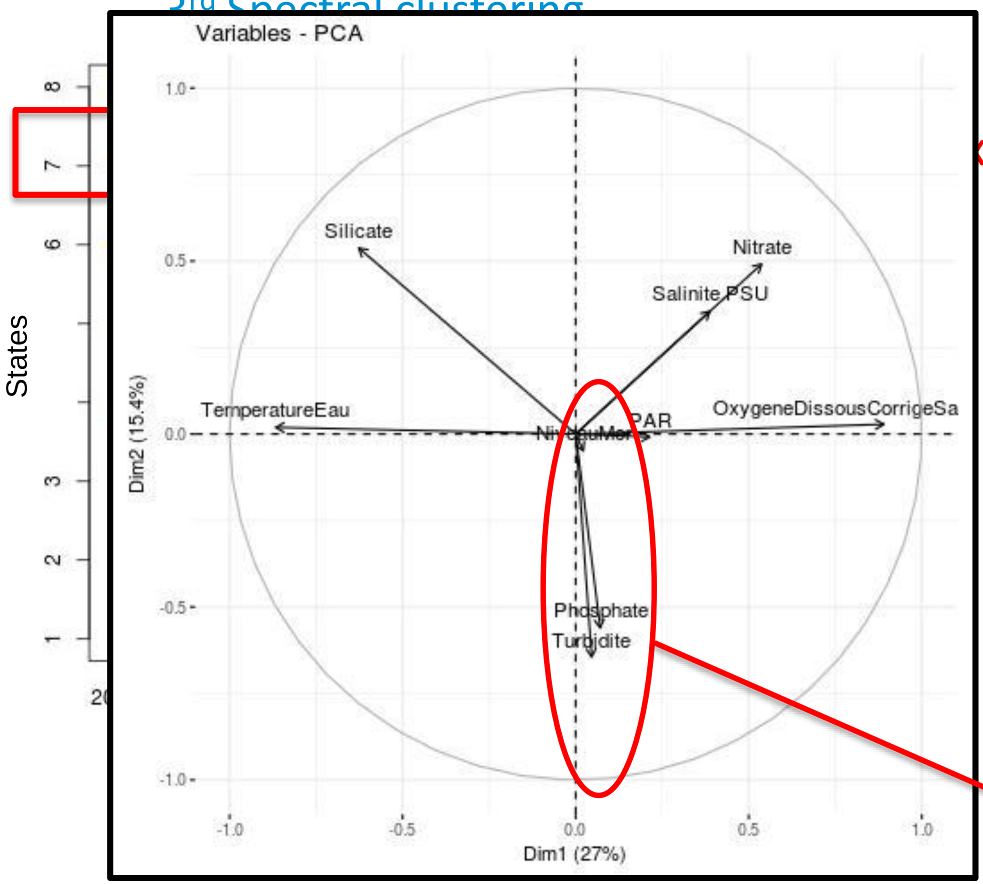


Phosphate Correlation State 7 = 0.62

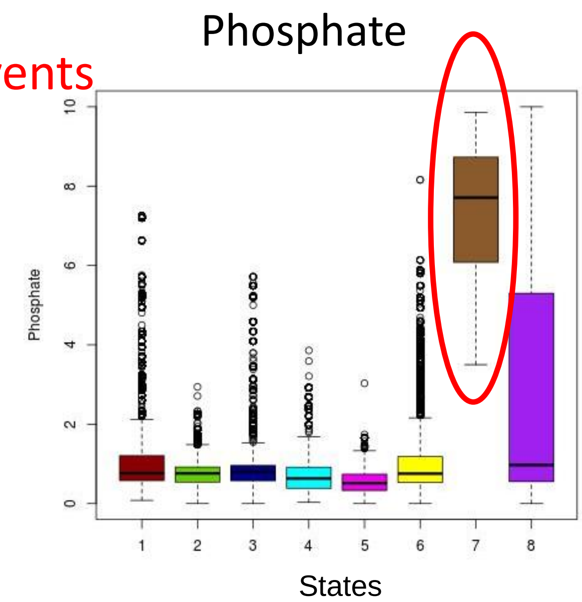
Detection of environmental states

Intermittent Events : rare/extreme

3rd Spectral clustering

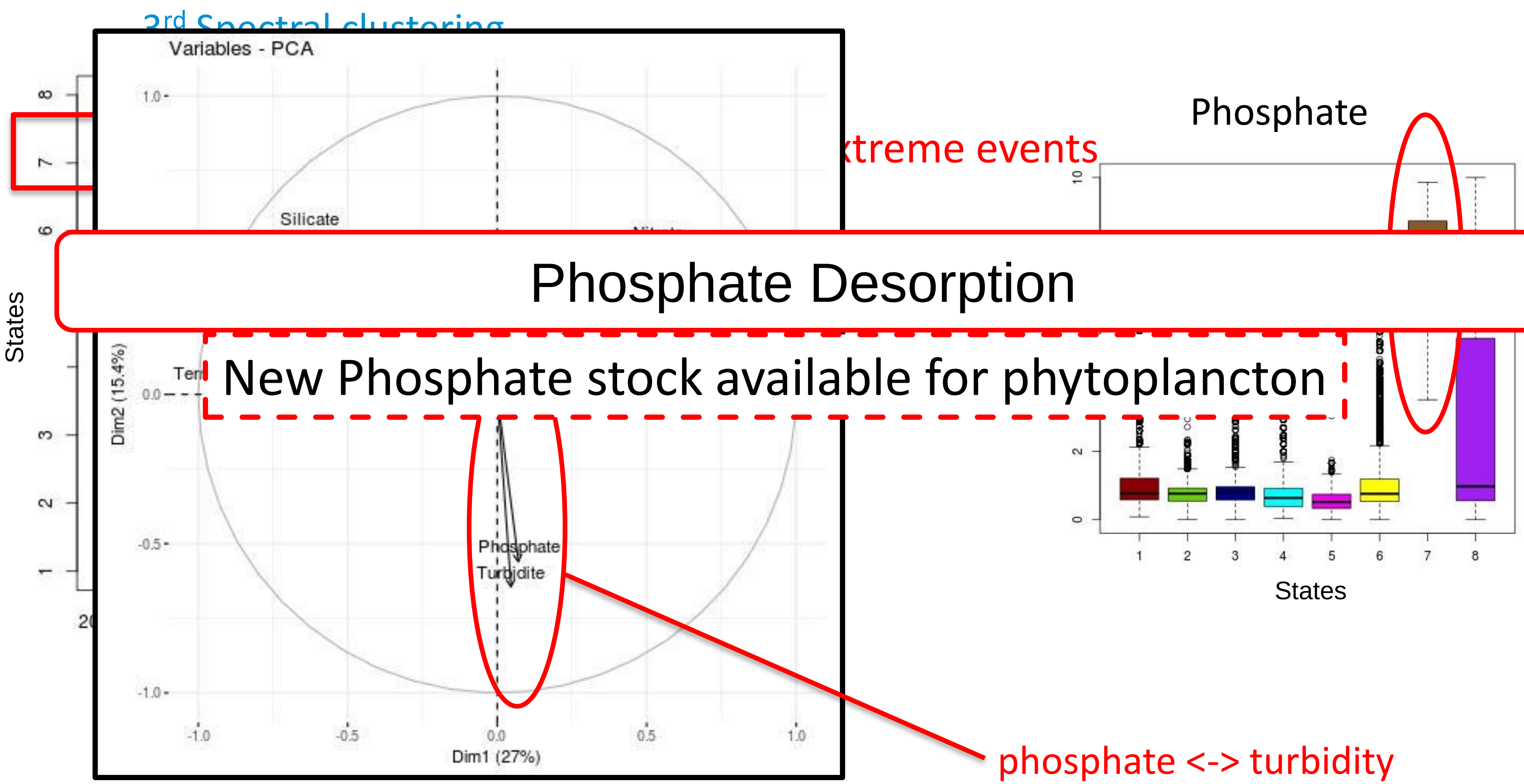


xtreme events



phosphate <-> turbidity

Intermittent Events : rare/extreme

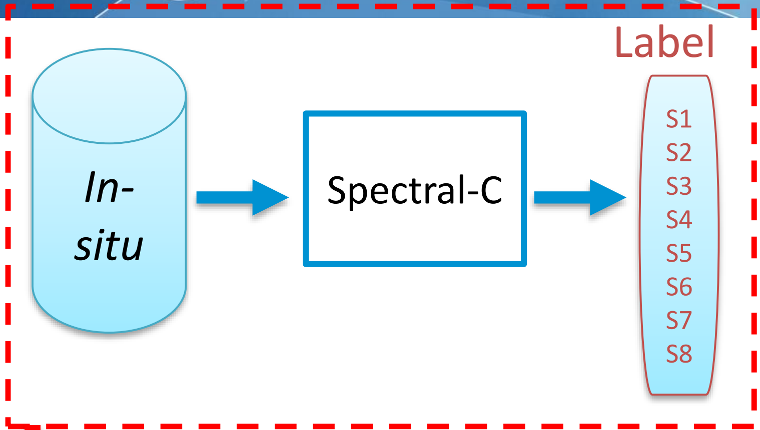


Phosphate Desorption

New Phosphate stock available for phytoplancton

phosphate <-> turbidity

CONCLUSION - PERSPECTIVES



Examples of labels:
Bloom, incl. HAB
Oxygen deficiency
Nutrient impulse
…

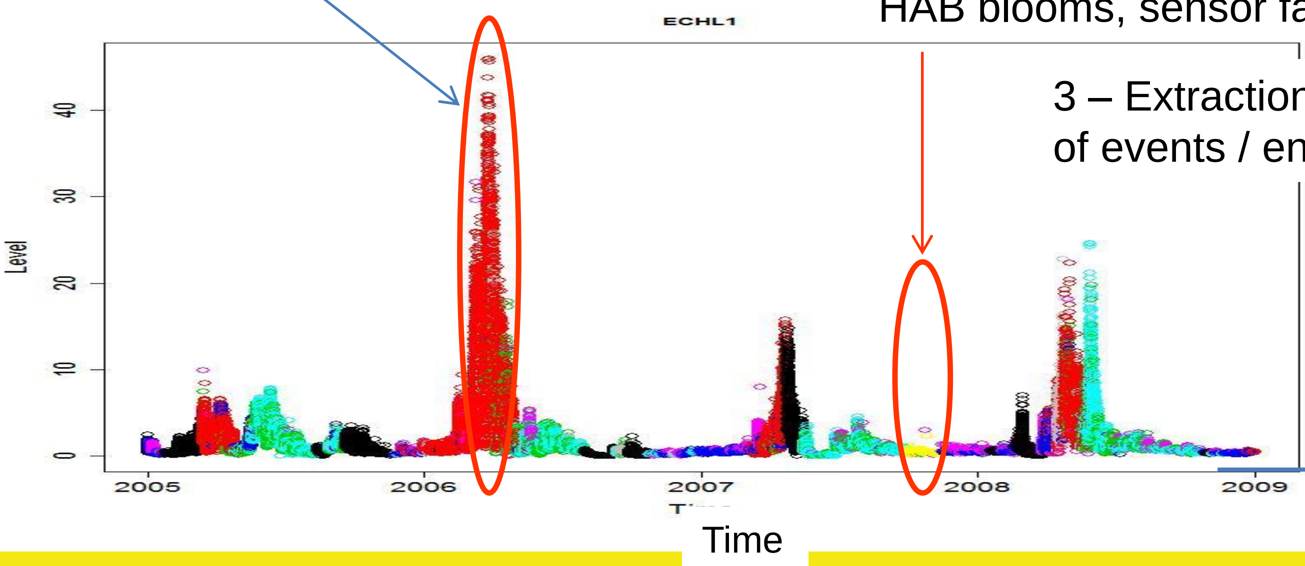
The spectral classification allowed to :

- Define **environmental states** in multi-parameter time series
- Detect, identify in time and space and **characterize states dynamics**
- **Extract <u>label</u> for frequent, rare or extreme events**

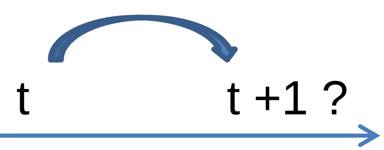
1 - Detection of usual events/states

2 - Detection of rare events/states : HAB blooms, sensor failures

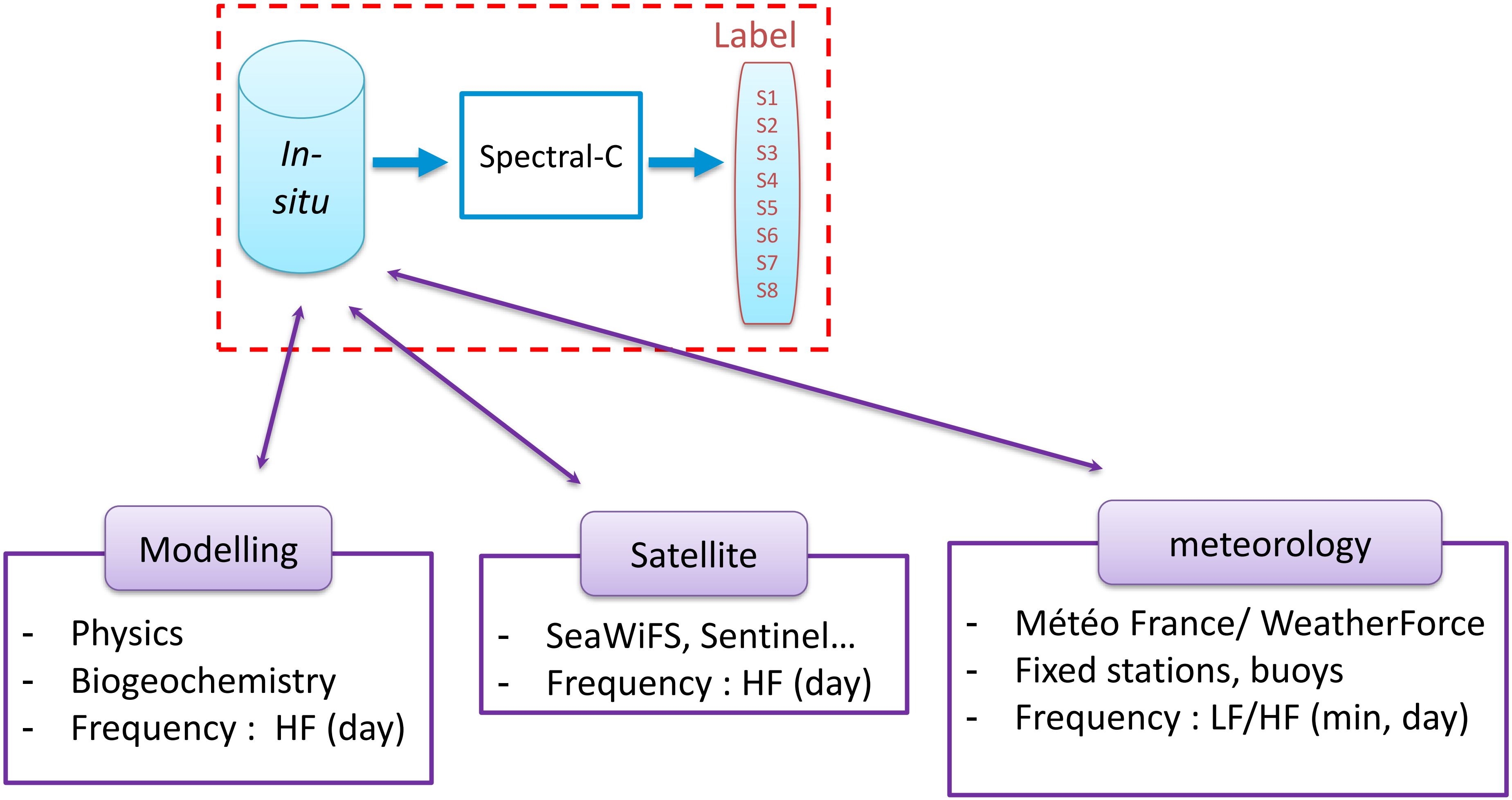
3 – Extraction of knowledge on dynamics of events / environmental states



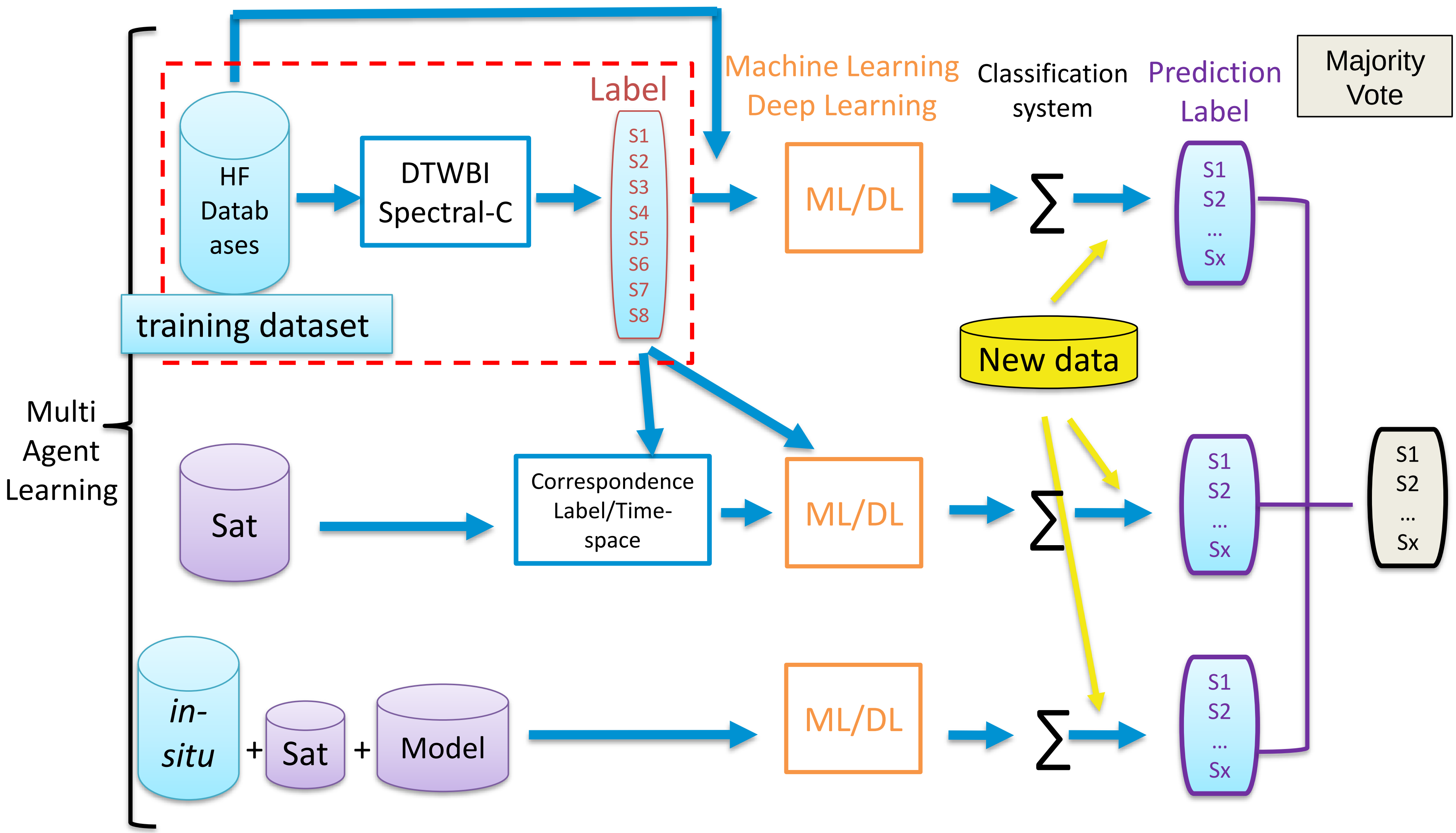
4 – Forecasting events



ADDING NEW DATA SOURCES



MACHINE LEARNING / DEEP LEARNING



Ifremer

Thank you for your attention

The authors want to acknowledge H2020 JERICO-Next for their financial contribution as well as the organizers.

This work has been also partly funded by the French government and the region Hauts-de-France in the framework of the project CPER 2014-2020 MARCO

Kelly Grassi's PhD is funded by WeatherForce as part of its R & D program "Building an Initial State of the Atmosphere by Unconventional Data Aggregation".