Incremental clustering of high-frequency coastal and marine observations for detecting changes in environmental conditions and phytoplankton communities, as well as early warning of HABs

Authors :
Carrat M., **Halawi Ghosn R.,** Hébert P.-A., Lefebvre A., Poisson-Caillault E.

Affiliation :
1 – Univ. Littoral, LISIC UR 4491, Calais, France
2 – IFREMER, Unité Littoral, Laboratoire Environnement et Ressources, Boulogne sur mer, France

This PhD project is funded by the Office Français de la Biodiversité (OFB) and by Ifremer (grant agreement OFB.21.0578)

- For centuries, the marine environment has been subjected to various sources of pollution.
- Causes of HABs and eutrophication: Excess Nutrients (N & P) from agricultural run off & sewage
- Effects of Harmful Algal Blooms:
  - ✓ Release Toxins in water
  - ✓ Reduce Biodiversity, and ecosystem services
  - ✓ Recreational activities

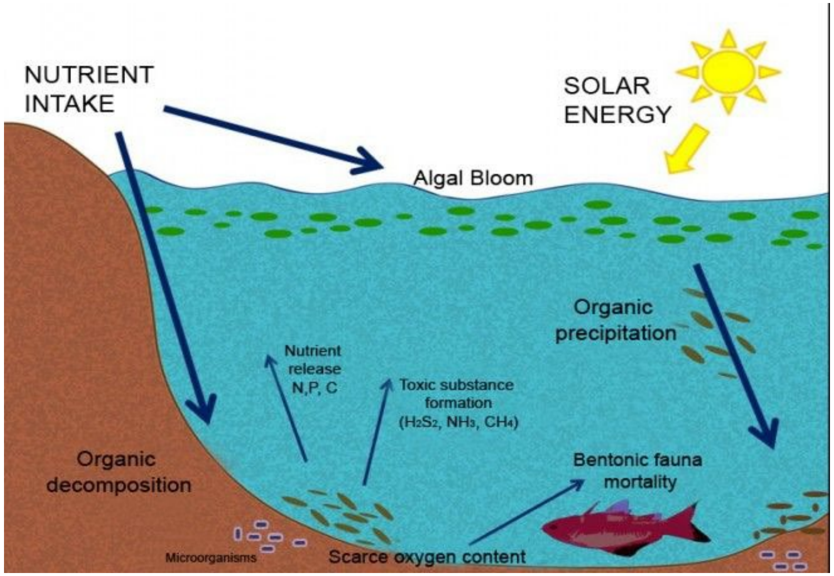


Figure-1 Eutrophication

Figure-2 *Paeocystis globosa* blooms in Boulogne (Ifremer)

Objective:

Our purpose is to:

1) Characterize environmental events (identify normal, recurrent and extreme events)

2) Understand phytoplankton dynamics

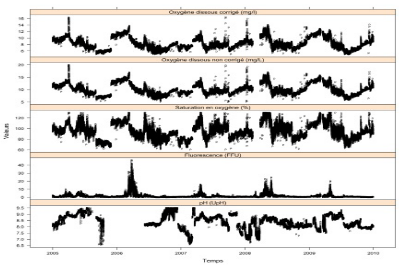
3) Predict HABs and develop an expert warning system (to help stakeholders, professionals and shell fish farmers)

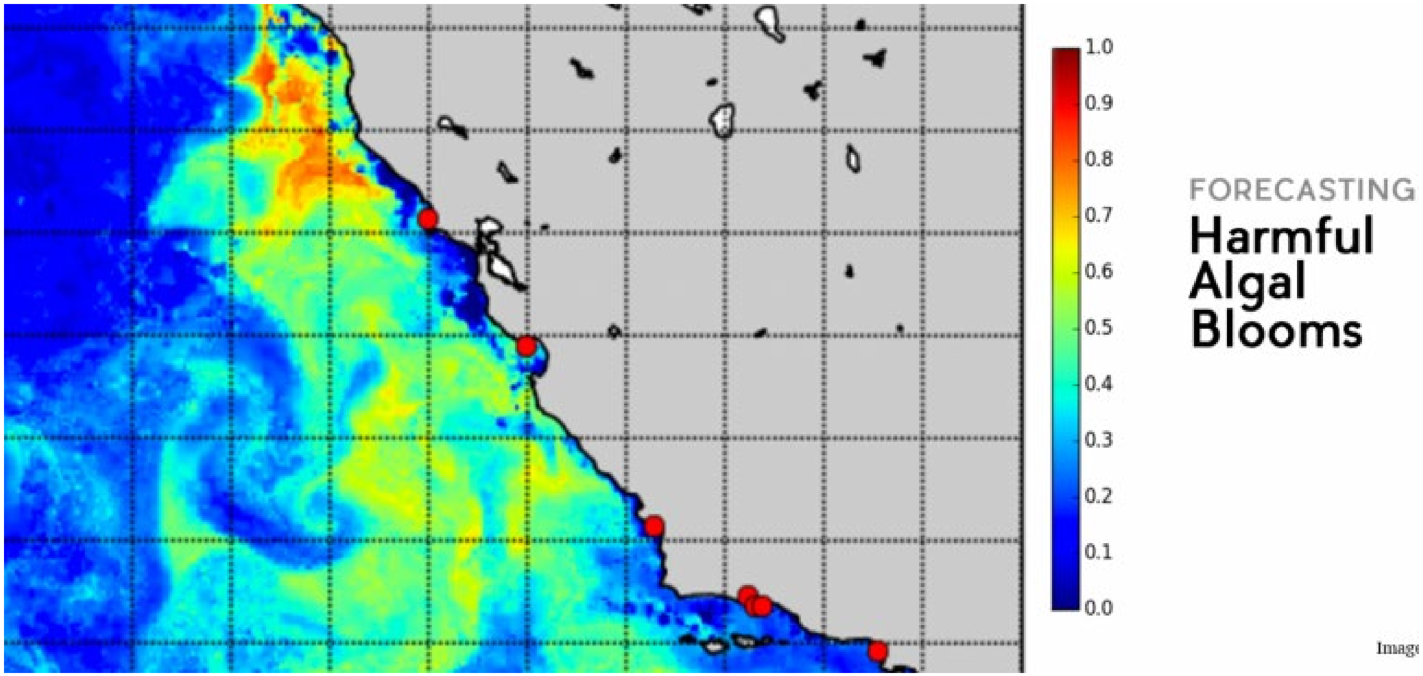
Approach

- Multi-source, multi-parameter, multi-criteria, multi-frequency approach
- Spatial and temporal scales
- We will then integrate the datasets

Data colllected from:

- Buoys
- Satellite
- Modelling
- Ferry Boxes





For Data from Buoys, Satellite and Modeling, we are using:

1) Multi-Level Clustering (MSC) to characterize the events
2) Random Forest for forecasting of HABs

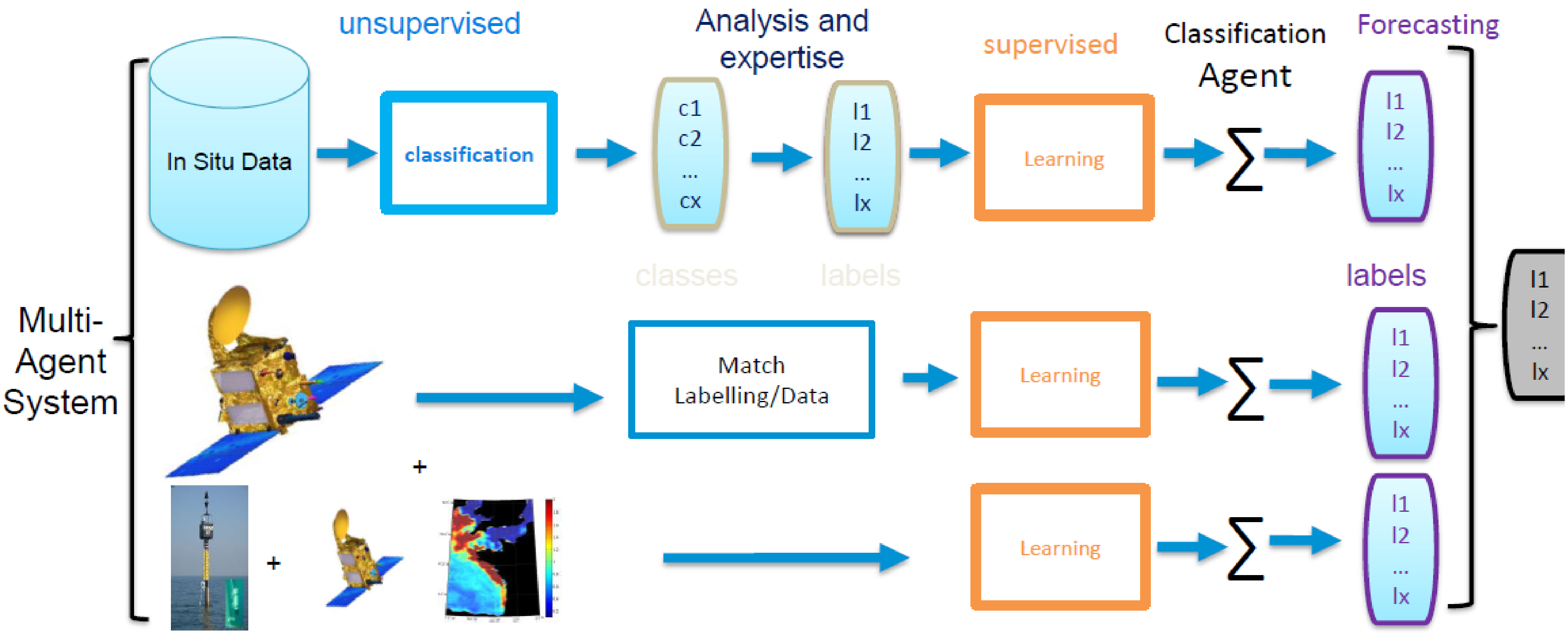


Figure-3 Summary of the methodology used for datasets from buoys, satellite and modelling

Methodology:

For Ferry Box dataset:

Our purpose is to propose a system capable of automatically:

1) visualizing and interpret data as it arrives

2) detecting changes in environmental conditions

3) Alerting on areas with high concentration of harmful algae

The system must be able to:

- adapt to newly acquired measurements and new or modified states
- retain a large amount of information.

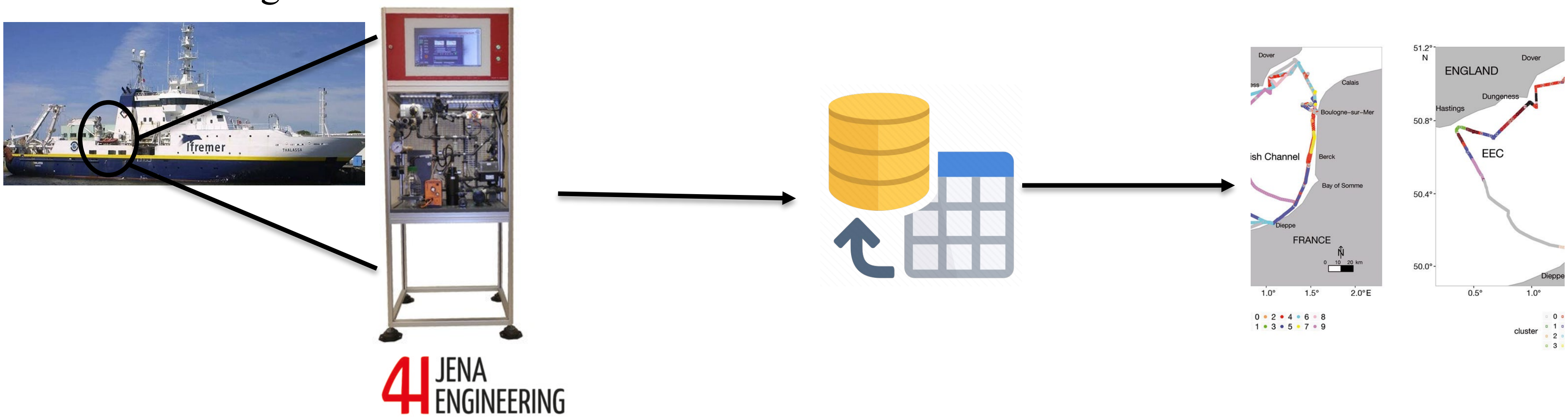


Figure-4 Tradional steps from collecting FB dataset till pre-processing

Newly Proposed Approach:

Incremental clustering with cold-start: a method to cluster a dataset and add data drop by drop or by batch without having to use the whole dataset.

Our newly proposed approach is based on Incremental Clustering (Kmeans approach in relevant space: PCA-kernel-spectral)

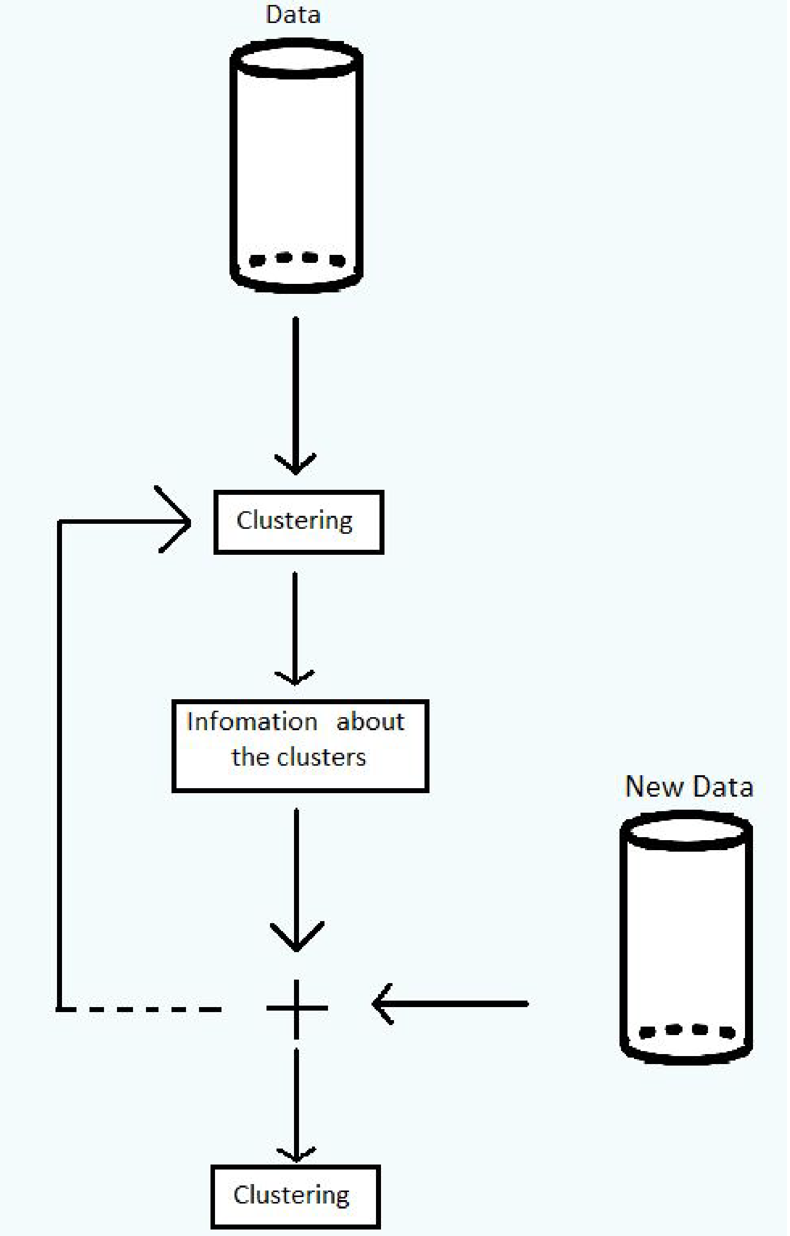


Figure-5 Incremental Clustering

K-Means Clustering Algorithm

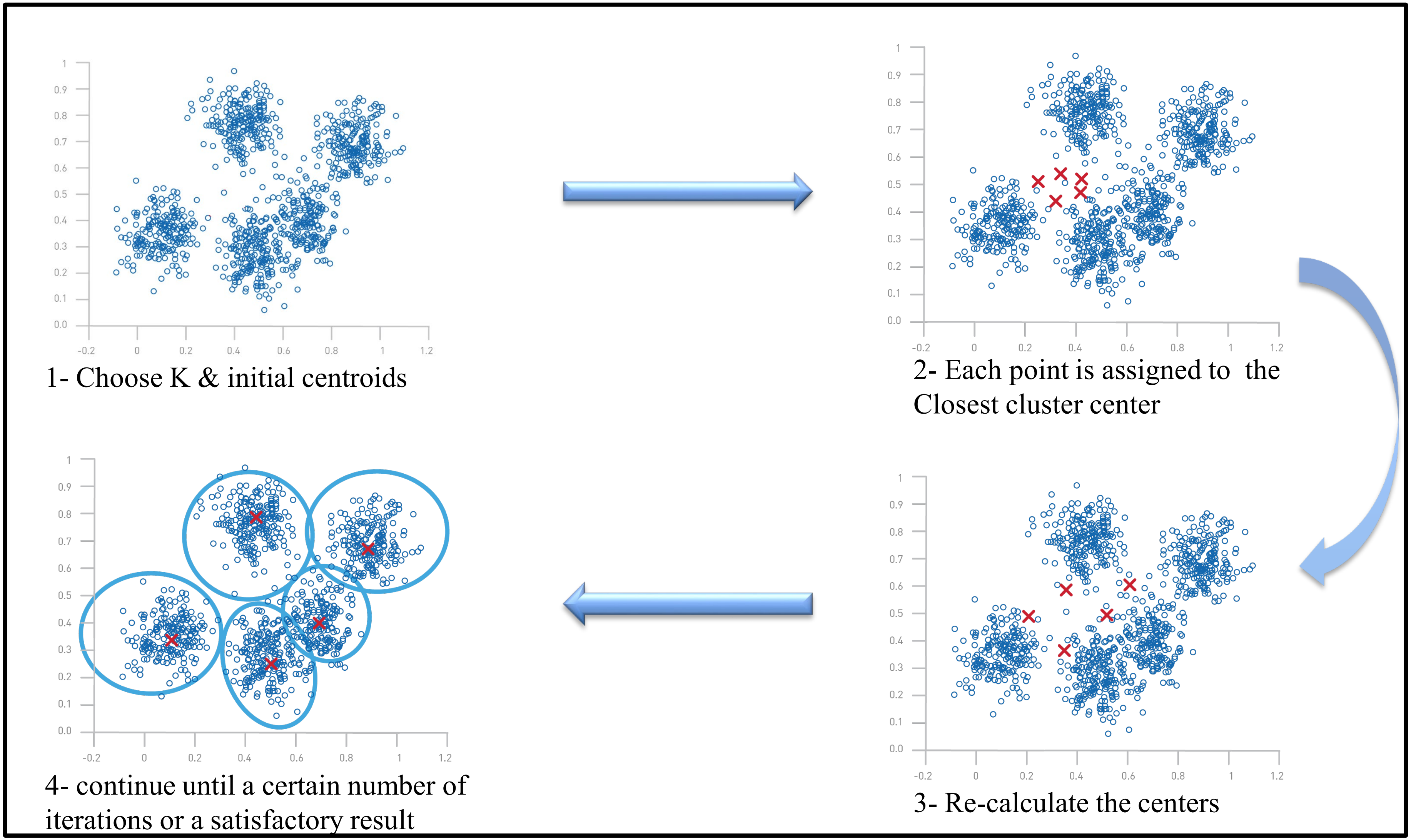


Figure-6 K-Means Incremental Clustering

Incremental K-Means

- Cold learning
- Step by step training of the model
- Identify the essential information to keep
- Update the model: Merge clusters or Add new clusters or Correct observed clusters

Return

1) Read a data batch → 2) Determine K (silhouette criteria/elbow) → 3) Kernel-K-means → 4) Reading a new batch → 5) K-means with previous centers (inclusion test for model update)

Incremental Clustering on Leg1 dataset

- 15 Clusters at the end
- Detects new eco-region and algal blooms in real time
- Add new information to anticipate events during cruises. For instance, we can identify blooms from our offices, and then take more samples

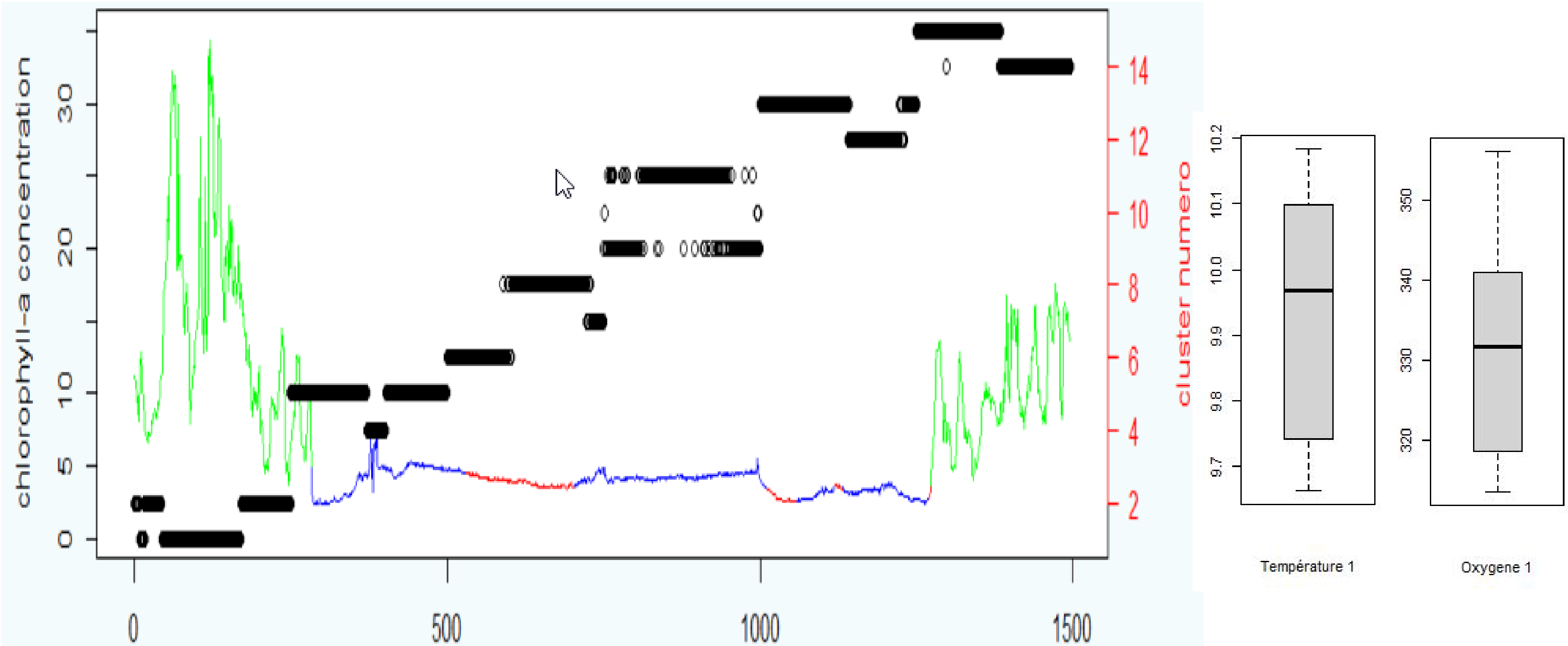


Figure-7 Results of K-Means Incremental Clustering on LEG1 DYPHYMA dataset

Incremental Clustering on Leg1 dataset

- Detects new eco-regions in near real time
- Characterize different events according to physiochemical variables (Temp, Salinity, DO)
- Separate English from French waters at high resolution in space
- Highlight heterogeneity for each kind of water (identify *Phaeocystis* blooms signals, riverine inputs..)

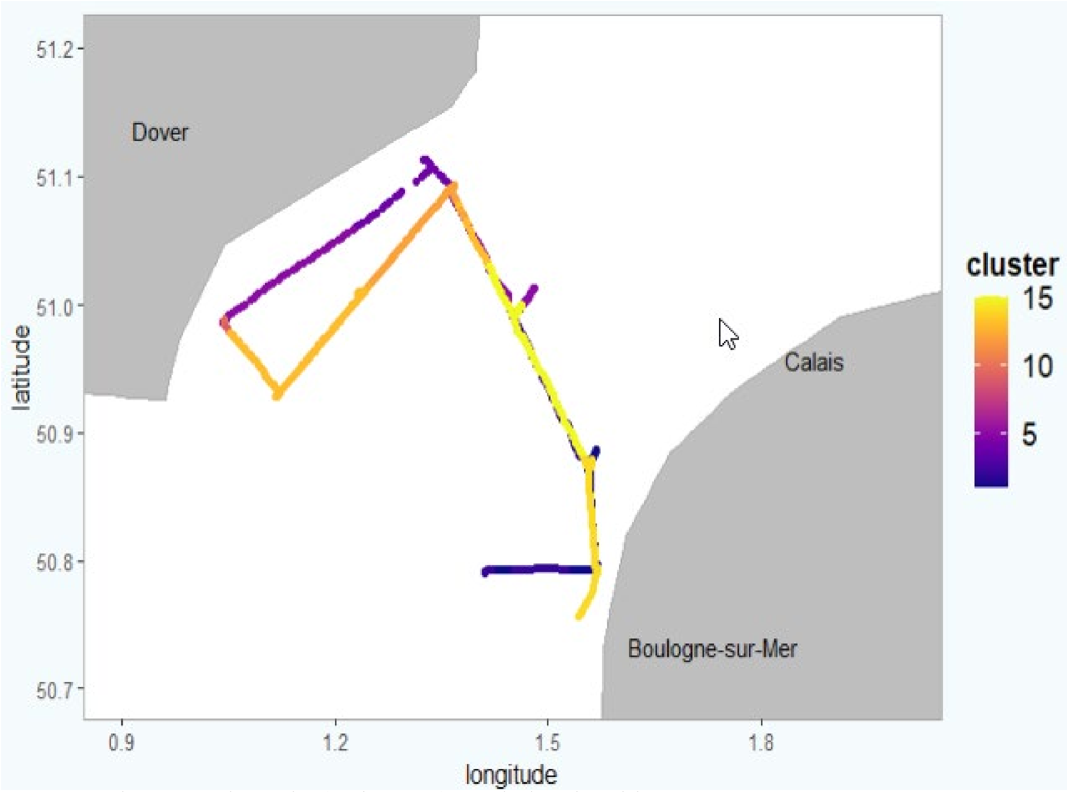


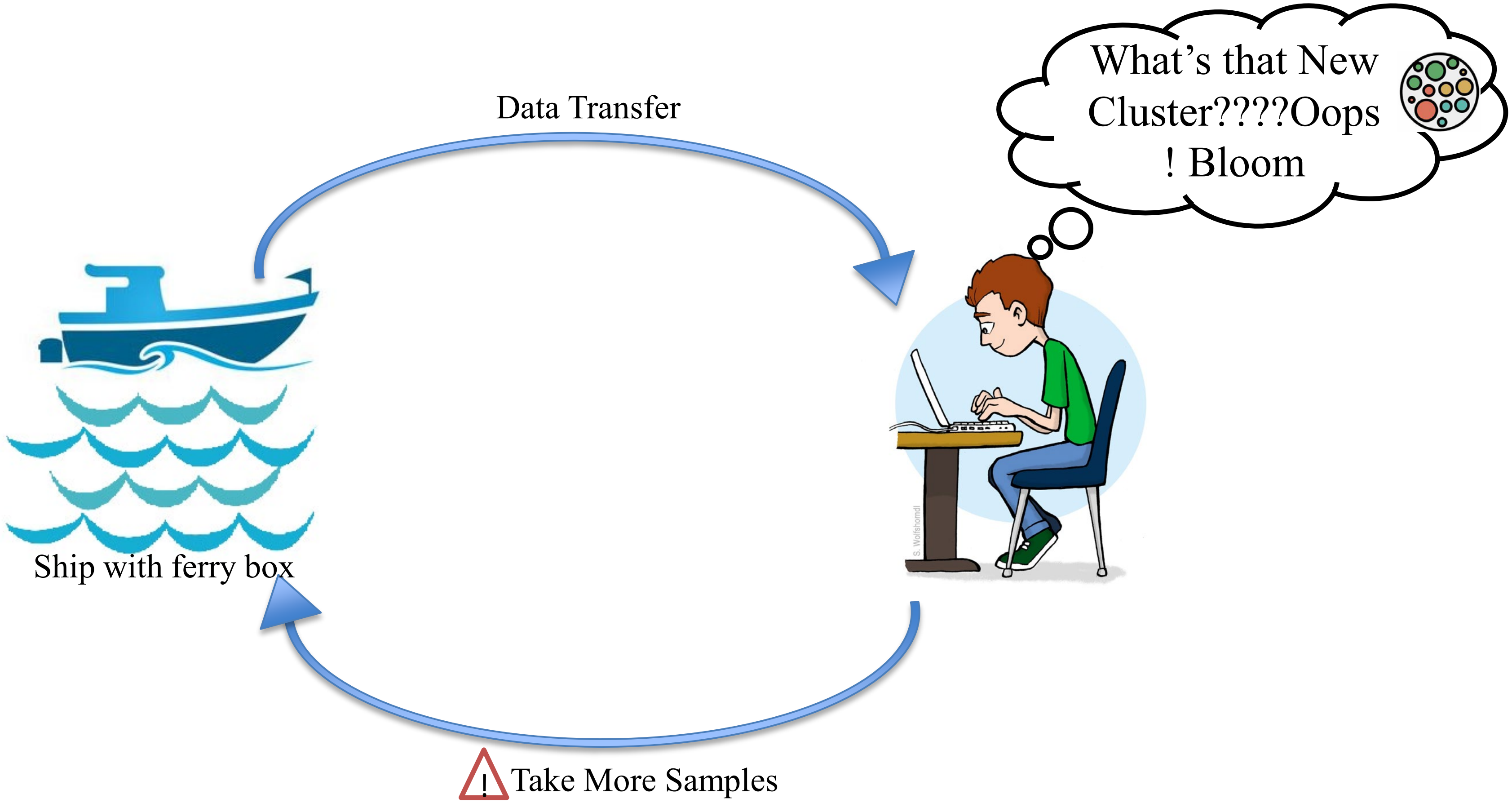
Figure-8 First trip (trajectory) made by the ship- DYPHYMA Leg1

Results:

Thus

1- Incremental Clustering detects new eco-regions in near real time

2- Can alert onboard technicians from the Hydro/phyto Lab or real-time setting of the FB sampling strategy, to take more samples to study blooms or new eco-regions



Thank you for your Attention!

This PhD project is funded by the Office Français de la Biodiversité (OFB) and by Ifremer (grant agreement OFB.21.0578)