Using Machine Learning to Characterize Phytoplankton Algal Blooms and Environmental States in the eastern English Channel

RAED HALAWI GHOSN[1], ÉMILIE POISSON-CAILLAULT[2], GUILLAUME WACQUET[1], ALAIN LEFEBVRE[1]

[1] *Ifremer, Laboratoire Environnement et Ressources, Department of Oceanography and Ecosystem Dynamics, Boulogne-sur-mer, France*

[2] *LISIC, Université Littoral du Cote d'opale, Calais, France*

Phytoplankton are photosynthetic organisms present at the base of the food web. The increase in anthropogenic activities leads to massive nitrate and phosphate riverine discharges into the coastal zone, and promotes an increase in phytoplankton biomass in coastal regions (and advecting offshore), thereby forming harmful algal blooms and fastening the eutrophication process. In the eastern English Channel (EEC), *Phaeocystis globosa* species bloom in the spring, and they are usually followed by diatom blooms of *Pseudonitzchia* complex. Here, we proposed a new approach to understand the dynamics of phytoplankton and reveal the environmental states (the change from one environmental state into another). As a contribution to the improvement of the Quality Assurance / Quality Control procedures, processed high-resolution raw data from the instrumented station MAREL Carnot (COAST-HF, IR-ILICO) was completed with an enhanced DTWBI data completion approach. Then, based on this optimized time series, the Multilevel Spectral Clustering (M-SC) approach was applied to reveal the different environmental states, including recurrent, rare, and extreme events. Following that, low-resolution data from the REPHY program / PHYTOBS-network, which contain information on phytoplankton species biomass and abundance, were linked to the MAREL Carnot clusters, and the main phytoplankton species associated with each cluster were determined. Inspired by the Margalef mandala concept and the Reynolds approach, this approach allows us to characterize the main controlling factors of HAB events in the EEC. Based on this labeling stage, a model based on machine learning will be developed to forecast such events. It will help to improve our capacity to promote early warning system to enhance, in near real time, the sampling strategy within monitoring networks, to warn shellfish farmers about such events, thereby limiting the socio-economical impacts. As a perspective, such an approach will be adapted to integrate data from modeling and ocean color products.