



**HAL**  
open science

# Caractérisation de la dynamique du phytoplancton, des événements récurrents aux événements extrêmes, grâce au Machine Learning

Alain Lefebvre, Kelly Grassi, André Bigand, P Chatelain, E Vincent, Émilie Poisson Caillault

► **To cite this version:**

Alain Lefebvre, Kelly Grassi, André Bigand, P Chatelain, E Vincent, et al.. Caractérisation de la dynamique du phytoplancton, des événements récurrents aux événements extrêmes, grâce au Machine Learning. Colloque EVOLECO - Evolution à Long terme des Ecosystèmes Côtiers, Nov 2021, La Rochelle, France. hal-03444190

**HAL Id: hal-03444190**

**<https://hal.science/hal-03444190>**

Submitted on 23 Nov 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Caractérisation de la dynamique du phytoplancton, des événements récurrents aux événements extrêmes, grâce au Machine Learning.

Lefebvre A.<sup>1</sup>, Grassi K.<sup>1,2,3</sup>, Bigand A.<sup>3</sup>, P. Chatelain<sup>3</sup>, E. Vincent<sup>3</sup>, Poisson-Caillault E.<sup>3</sup>

<sup>1</sup> IFREMER, Unité Littoral, Laboratoire Environnement et Ressources, Boulogne sur Mer, France ; email : alain.lefebvre@ifremer.fr

<sup>2</sup> WeatherForce, Toulouse, France

<sup>3</sup>LISIC, EA 4491, Université du Littoral Côte d'Opale, Calais, France

Mots clef : Machine Learning, Classification, Observation Augmentée, Eutrophisation, Harmful Algal Bloom.

## Résumé :

Le phytoplancton constitue la base des chaînes trophiques pélagiques et une part essentielle de la production primaire dans les zones côtières (De Jonge et Elliot 2001 ; Cloern et Jasby, 2008). Sous certaines conditions favorables de température, lumière, turbulence et sels nutritifs, la croissance de diverses espèces phytoplanctoniques dont les Harmful Algal Blooms (HABs) peut excéder celles des organismes brouteurs, entraînant leur prolifération. Les HABs peuvent provoquer des effets néfastes sur les organismes, via la production de toxines, des dommages mécaniques ou pour d'autres raisons, telle que l'anoxie du milieu due à une reminéralisation des algues non consommées accumulées (Lassus et al., 2016). Ces effets peuvent conduire à des mortalités massives de poissons et/ou d'invertébrés benthiques, à des épisodes de contamination des ressources les rendant impropres à la consommation humaine et engendrant des pertes économiques importantes. Ainsi, alors que certains taxons vont naturellement contribuer aux transferts énergétiques vers les niveaux supérieurs et contribuer ainsi à un fonctionnement optimal de l'écosystème, d'autres taxons, responsables de la formation de ces HABs, vont être qualifiés de toxiques et/ou nuisibles pour le fonctionnement de l'écosystème et sa qualité de manière générale.

Par sa capacité à répondre rapidement aux modifications de la qualité de son environnement, permettant ainsi d'évaluer une réponse à différentes sources de pressions, le phytoplancton est régulièrement utilisé comme un indicateur de qualité pour les directives et conventions de mer régionales (exemples de la Directive Cadre sur l'Eau (DCE - 2000/60/CE), de la Directive Cadre Stratégie du Milieu Marin (DCSMM - 2008/56/CE) et de la Convention d'Oslo et de Paris (OSPAR, 2010)). La biomasse du phytoplancton, son abondance et sa composition, la fréquence et l'intensité des efflorescences sont régulièrement utilisés dans les métriques développées pour leurs besoins.

La Manche orientale et la baie sud de la mer du Nord sont caractérisées par des proliférations anormales (augmentation de la durée et de l'intensité des efflorescences) de la Prymnesiophycée *Phaeocystis globosa* dont l'accumulation de biomasse a des conséquences néfastes sur le fonctionnement des écosystèmes pélagiques et benthiques. Par ailleurs, les blooms de *Phaeocystis* sont potentiellement associés à des développements de Diatomées du groupe des *Pseudo-nitzschia*, dont certaines espèces sont susceptibles de produire des toxines amnésiantes s'accumulant dans les produits de la mer (Delegrange et al., 2018 ; Lefebvre et Dezechache, 2020).

La compréhension du déterminisme de ces proliférations, et de manière plus globale, de la structure et la dynamique de la composante phytoplanctonique est donc un prérequis indispensable pour toute évaluation de la diversité marine (Garmendia et al., 2013). Le maintien des biens et des services écologiques et écosystémiques est également en partie lié à la dynamique du phytoplancton. Cependant, cette compréhension ne peut s'affranchir de l'étude de leur régulation environnementale (climat, lumière, nutriments...) et biologique (consommation, sédimentation, lyse...) impliquant des approches à différentes échelles d'observations (long-terme vs court terme, basse fréquence vs haute fréquence, régional vs global) et la prise en compte des interactions d'échelles. A ce jour, ces approches multi-échelles, multi-paramètres sont encore peu nombreuses. Les échelles spatiales et temporelles considérées par les programmes classiques de l'Observation Marine Côtière ne permettent que de couvrir une partie des différents processus impliqués dans les efflorescences du phytoplancton. Lorsque toutes les conditions sont réunies pour une approche multi-source, multi-paramètre et multi-échelle alors les méthodes habituelles de traitement des données s'avèrent inefficaces pour extraire le maximum d'information de ces bases de données complexes. Alors que les données basses fréquences (BF) continuent de livrer leurs secrets, la complexité des données HF (bases volumineuses, non linéarité, données manquantes,...) rend leur exploitation difficile.

Afin de contribuer au développement de système d'Observation Intégrée de la zone côtière, nos objectifs scientifiques sont de deux types : Méthodologique (M) et Écologique (E), et se déclinent ainsi :

- (M1) mettre en place un flux de données hydrodynamiques, biogéochimiques et biologiques adaptés à la problématique et aux zones à enjeux en proposant un recensement des bases et des portails contenant les informations les plus adéquates et les plus fiables possibles,
- (M2) adapter et optimiser les méthodes numériques d'analyses des séries temporelles BF et HF déjà développées dans des travaux antérieurs par les équipes encadrantes,
- (E1) identifier (semi-) automatiquement les événements habituels, récurrents (donc des états environnementaux), même s'ils sont d'amplitude, de durée différentes, dans des écosystèmes contrastés (Figure 1 A),
- (E2) mettre en évidence les événements extrêmes, intermittents et/ou rares qui font la particularité de ces écosystèmes (Figure 1, B),
- (E3) caractériser la dynamique de passage d'un état environnemental vers un autre en tenant compte des aspects multi-paramètre et multi-échelle (Figure 1, C),
- (E4) prédire l'évolution du système en considérant différents scénarios de modifications des paramètres de contrôle des efflorescences phytoplanctoniques (Figure 1, D).

Sur la base des connaissances acquises quant à la dynamique des états environnementaux et la hiérarchisation des facteurs de contrôle, il est envisagé de développer :

- (M3) un système d'alerte des efflorescences d'algues nuisibles ou potentiellement toxiques, devant permettre une diminution du temps nécessaire pour mettre en place des interdictions de pêche et de ramassage des coquillages dans les régions affectées par les HABs, tout en maintenant les autres régions actives, réduisant ainsi les impacts économiques sur le secteur,
- (M4) un système expert d'aide à la décision pour améliorer les évaluations environnementales et ainsi contribuer à une meilleure définition des programmes de surveillance et de mesures en support aux besoins des Directives et des Conventions de Mers Régionales,

-(M5) un jeu de données marines, de référence et labellisé de type « UCI Machine Learning Repository » afin de servir de ressources pour la recherche empirique et méthodologique en Machine Learning.

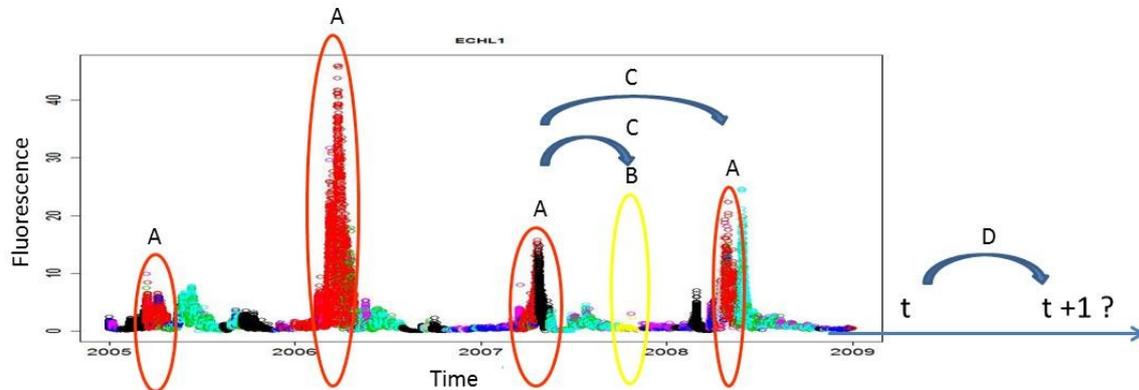


Figure 1. Évolution de la fluorescence mesurée par MAREL Carnot : identification des principaux états environnementaux et des éléments de dynamique à caractériser par classification et modélisation : (A) Événements courants, mais d'amplitude et de durée variables, (B) Événements rares, intermittents ou extrêmes, (C) dynamique entre les états et (D) prédiction des états.

Dans ce contexte, afin de préparer la phase de labellisation, étape préalable indispensable à l'apprentissage et donc à la prédiction, un système numérique Open Source basé sur plusieurs méthodes du Machine Learning a été proposé. Il inclut une nouvelle méthode de Classification Spectrale Multi-Niveaux (M-SC). Ce système doit permettre (i) de définir des schémas de fonctionnement des efflorescences à partir de données multi-sources et multi-échelles, (ii) de disposer d'un système de prédiction et d'alerte, (iii) de pouvoir adapter en temps (quasi) réel les stratégies d'échantillonnage. La mise en évidence d'états environnementaux favorables ou pas aux efflorescences algales permet de hiérarchiser les facteurs de contrôle et d'identifier des schémas de fonctionnements. Cette approche appliquée aux données de la station MAREL-Carnot (IR ILICO COAST-HF) et en considérant également des données BF, a permis la description (aux niveaux biogéochimique et taxonomique) d'événements récurrents mais aussi extrêmes, dont la période peut être infra-hebdomadaire ou même horaire (Figure 2). En plus de la caractérisation par état, cette approche multicritères identifie des schémas relationnels (approche de type DPSIR). Elle met en évidence des schémas de succession des facteurs d'influences et des communautés phytoplanctoniques avec un zoom sur les taxons nuisibles de la Manche orientale, *Phaeocystis globosa* et le groupe des *Pseudo-Nitzschia*.

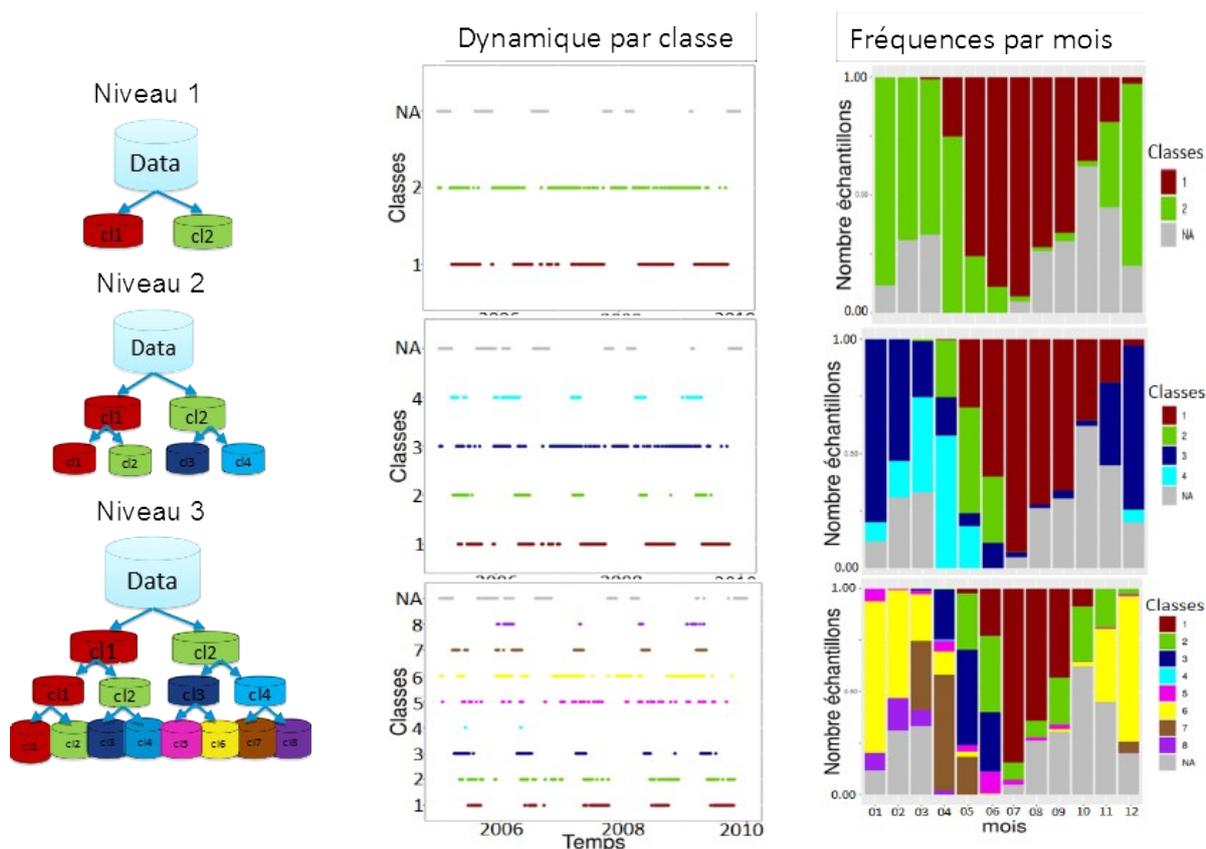


Figure 2. Application de la méthode de classification spectrale multi-niveaux (M-SC) à 3 niveaux aux données de la station instrumentée MAREL Carnot : Schémas de la classification par niveau (colonne de gauche). Fréquence d'occurrence de chaque état par mois et par niveau (colonne centrale) et dynamique de chaque état par niveau (colonne de droite). La couleur grise représente les données non classées (au moins une donnée manquante (#NA)).