



HAL
open science

Téledétection du phytoplancton par méthode neuronale : du global au régional, de la composition pigmentaire aux biorégions

Roy El Hourany

► **To cite this version:**

Roy El Hourany. Téledétection du phytoplancton par méthode neuronale : du global au régional, de la composition pigmentaire aux biorégions. Biodiversité et Ecologie. Sorbonne Université, 2019. Français. NNT : 2019SORUS095 . tel-02562426v2

HAL Id: tel-02562426

<https://theses.hal.science/tel-02562426v2>

Submitted on 20 Nov 2020

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.

Sorbonne Université

Ecole doctorale 129 Sciences de l'Environnement

Laboratoire d'Océanographie et du Climat : Expérimentations et Approches Numériques

Centre National de Télédétection, CNRSL

Télédétection du Phytoplancton par Méthode Neuronale *Du Global au Régional, de la Composition Pigmentaire aux Biorégions*

Par Roy ELHOURANY

Thèse de doctorat

Méthodes Statistiques appliquées à la Biologie Marine

Dirigée par Mme Sylvie Thiria, M. Ghaleb Faour et Mme Marie Abboud-Abi
Saab

Présentée et soutenue publiquement le 29 Mars 2019

Devant un jury composé de :

M. Gérard Biau	Professeur des Universités,	Président du Jury
Mme Bernadette Dorrizzi	Professeur des Universités,	Rapporteur
M. Huber Loisel,	Professeur des Universités,	Rapporteur
Mme Marina Levy	Directrice de Recherche,	Examineur
M. Hervé Claustre	Directeur de Recherche,	Examineur
Mme Sylvie Thiria	Professeur des Universités,	Directrice de Thèse
M. Ghaleb Faour	Directeur de Recherche,	Co-directeur de Thèse
Mme Marie Abboud-Abi Saab	Directrice de Recherche,	Co-directrice de Thèse

A ma famille

Qui n'a jamais cessé de me soutenir pour que je puisse atteindre mes objectifs

Résumé:

Cette thèse présente une approche novatrice d'analyse et d'observation de la structure de la communauté de phytoplancton à l'échelle mondiale et régionale à l'aide de données satellitaires (couleur de l'océan et température de surface) et d'observations in-situ. L'approche est basée sur des méthodes neuronales de classification, telles que les cartes auto-organisatrices (SOM) calibrées sur une grande base de données globale formée de mesures satellitaires collocalisées avec des mesures in-situ. Nous avons d'abord développé une méthode d'estimation des pigments phytoplanctoniques secondaires appliquée à l'océan global à partir de mesures satellitaires. Ensuite nous avons réalisé une étude fine de la Méditerranée où les groupes phytoplanctoniques (PFTs) ont été identifiés. En se servant des mesures de profondeur de la couche de mélange (MLD) fournies par les flotteurs ARGO, de la température de surface de la mer (SST) et de la concentration en chlorophylle-a (Chla) satellitaire, nous avons déterminé sept bio-régions basées sur le cycle annuel de ces variables en utilisant une SOM modifiée. Enfin ces bio-régions ont été caractérisées en termes de PFTs. Les méthodes utilisées nous ont permis d'évaluer les incertitudes sur les pigments et sur les PFTs. L'ensemble des méthodes proposées dans la thèse permettent d'effectuer des études similaires dans d'autres régions.

Mots-clés : Phytoplancton, pigments secondaires, Télédétection, Biorégions, Réseaux Neuronaux, Classification, Carte Auto-Organisatrice, Océan Global, Méditerranée.

Abstract:

This thesis presents a novel approach to analyze and observe the phytoplankton community structure at global and regional scale using satellite data (Ocean colour and Sea surface temperature) and in-situ observations. The approach is based on neural network classification methods, such as Self-Organizing Maps (SOM) trained on a large global database composed of satellite observations collocated with in-situ measurements. First, we developed a method to estimate secondary phytoplankton pigments from satellite measurements in the global ocean. Then we focused our studies on the Mediterranean Sea. Phytoplankton groups (PFTs) were identified from the secondary pigments estimated in the first phase. We then characterized seven bio-regions by clustering annual cycles MLD obtained from Argo floats, SST and Chla by using an advanced SOM. At last, these bio-regions were characterized in terms of PFTs. The methods developed in this thesis allowed us to estimate uncertainties on secondary pigments and PFTs. The applicability of these methods are broad and can be used to investigate other oceanic areas.

Keywords: Phytoplankton, secondary pigments, Ocean Colour, Neural network, Self-Organizing Maps, Global Ocean, Mediterranean Sea.

Remerciement

J'aimerais remercier tous ceux qui ont contribué au succès de cette thèse.

Je remercie en premier lieu ma directrice de thèse, Mme Sylvie THIRIA, pour son soutien, pour sa grande disponibilité tout au long de ma thèse, et pour son apport immense sur le plan scientifique et professionnel. Je tiens à lui exprimer ma grande gratitude pour son accompagnement journalier, son souci et sa confiance.

Mes remerciements sont également adressés à mon co-directeur de thèse, M. Ghaleb FAOUR, qui a dirigé mes travaux de recherche depuis mon Master 2 et m'a guidé tout au long de cette thèse avec entière implication, patience et confiance.

J'adresse toute ma reconnaissance à ma co-directrice de thèse, Mme Marie Abboud-Abi Saab, qui a mis à ma disposition ses vastes expériences, conseils et interprétations pour m'aider à accomplir ce travail.

Ma gratitude et mon respect à M. Michel Crépon qui m'a encadré avec enthousiasme durant ma recherche. Ses commentaires précieux émanent d'une grande connaissance dans le domaine de l'océanographie.

Mes remerciements sont également adressés aux membres du jury; M. Gérard Biau, Mme Bernadette Dorizzi, M. Hubert Loisel, Mme Marina Lévy et M. Hervé Claustre. Ils m'ont tous fait l'honneur d'être présents à ma soutenance de thèse et leurs remarques m'ont été très bénéfiques.

Je tiens à remercier le Conseil National de la Recherche Scientifique CNRS-Liban, pour m'avoir accordé une bourse doctorale pour effectuer la thèse.

Je remercie le Centre National de Télédétection et le LOCEAN qui m'ont accueilli pendant ces trois ans en m'offrant un cadre de travail très enrichissant.

Je tiens à remercier Mme Sarah Najem, M. Julien Brajard, M. Anastase Charantonis, M. Ali Fadel, et M. Mario Mhawij pour leur bienveillance et leur guide tout au long de ce parcours de thèse. Ma gratitude est également adressée à M. Carlos Mejia pour son attention et son expertise dans les notions de programmation informatique.

Je tiens à remercier M. Fouad Badran pour ses commentaires et corrections lors de la rédaction du manuscrit de thèse.

Ma profonde gratitude et mon obligeance va ensuite envers ma famille et mes amis. Leur foi en moi et leur soutien m'ont été primordiaux.

Table des Matières

Introduction Générale	5
Chapitre I: Télédétection Marine, Phytoplancton ; Généralités	8
1. Télédétection marine :	9
1.1. Propriétés optiques apparentes et inhérentes des océans :	9
1.2. Constituants de l'eau (optiquement significative):	12
1.3. Types des Eaux : Cas 1 et Cas 2 :	12
2. Le phytoplancton :	14
2.1. Notion de <i>Phytoplankton Size Class (PSC)</i> et <i>Phytoplankton Functional Type (PFT)</i> : ...	14
2.2. Méthodes d'identification et de caractérisation du Phytoplancton :	15
2.3. Etude du phytoplancton depuis l'espace :	18
Chapitre II: Télédétection de la Variabilité des Pigments Phytoplanctoniques Secondaires par Méthode Neuronale ; Approche Globale	21
1. Introduction:	22
2. Cartes de Kohonen	25
2.1. Idée de base	25
2.2. Architecture	25
2.3. Algorithme d'apprentissage	26
2.4. Propriétés	28
3. Estimation des pigments secondaires phytoplanctoniques à partir des images satellitaires en utilisant les SOM	29
Article 1: Estimation of secondary phytoplankton pigments from satellite observations using self-organizing maps (SOM)	30
1. Introduction	32
2. Materials	35
2.1. GlobColour Data:	35
2.2. AVHRR SST Data	37
2.3. HPLC pigment Dataset	37
2.4. The experimental Database	39
3. The Proposed Method	40
3.1. Self-organizing maps (SOM), the general concept	40

3.2.	Construction of the SOM.....	41
4.	Results.....	44
4.1.	Analysis of the SOM Organization and topology	44
4.2.	Cross validation results	46
4.3.	Estimation of pigment concentrations in the global ocean	48
5.	Discussion	54
5.1.	Comparison to other approaches deriving pigment concentrations.....	54
5.2.	Uncertainties and quality control	56
5.3.	Spatio-temporal variability of phytoplankton pigments	57
6.	Conclusion	59
Chapitre III :	La Méditerranée ; Caractéristiques Géographiques, Hydrodynamiques et Ecosystémiques	60
1.	Caractéristiques géographiques et hydrodynamiques	61
2.	Régime du vent.....	62
3.	Circulation globale.....	63
3.1.	Circulation de l'eau Atlantique	63
3.2.	Circulation intermédiaire et thermohaline	64
4.	Caractéristiques Ecosystémiques.....	65
Chapitre IV :	Téledétection du Phytoplancton de la Méditerranée ; Approche Régionale	67
1.	Introduction	68
Partie 1 : Téledétection du groupe de phytoplancton dominant au niveau de la Méditerranée en utilisant des cartes auto-organisatrices (SOM)		70
<u>Article 2: Phytoplankton diversity in the Mediterranean Sea from satellite data using Self-Organizing Maps.....</u>		<u>71</u>
1.	Introduction.....	72
2.	Materials.....	75
2.1.	The SOM-Pigments:.....	75
2.2.	Satellite Data	75
2.3.	Med HPLC pigment Dataset (Med-Pigments).....	76
3.	The Proposed Method.....	78
3.1.	Self-organizing maps (SOM), general concept	78
3.2.	SOM-PFT: calibration and validation procedure	79
4.	Results.....	82
4.1.	Analysis of the SOM-PFT	85

4.2.	Validation experiments.....	86
4.3.	Distribution of dominant PFTs in the Mediterranean Sea.....	90
5.	Discussion.....	92
5.1.	Phytoplankton dynamic in the Mediterranean Sea.....	93
6.	Conclusion.....	95
Partie 2 : Reconstruction de la variabilité spatio-temporelle des PSC au niveau la Méditerranée.		
..... 96		
1.	Introduction.....	96
2.	Matériels et méthodes.....	97
3.	Résultats.....	99
3.1.	Analyse de la variabilité des PSC.....	99
3.2.	Comparaison avec des séries temporelles de comptage de phytoplancton.....	102
4.	Discussion.....	107
4.1.	La dynamique du phytoplancton dans la Méditerranée:.....	107
4.2.	Evaluation de la méthodologie:.....	108
5.	Conclusion.....	110
Chapitre V : Régionalisation de la Méditerranée par la méthode de 2S-SOM..... 112		
1.	Introduction:.....	113
2.	Etat de l'art ; Biorégionalisation de la Méditerranée.....	115
3.	Base de Données.....	117
3.1.	Couleur de l'Océan.....	117
3.2.	Climatologie de profondeur de la couche de Mélange, Mixed Layer Depth (MLD)	119
3.3.	Base de données d'apprentissage :.....	120
4.	Méthodologie:.....	120
4.1.	Définition de l'Algorithme 2S-SOM.....	120
4.2.	Fonctionnement de l'algorithme 2S-SOM :.....	122
4.3.	Application de l'algorithme :.....	123
5.	Résultat.....	127
5.1.	Analyse de la 2S-SOM :.....	127
5.2.	Caractérisation des Cycles de Climatologie mensuelle de Chla, SST et MLD de Chaque classe	130
5.3.	Distribution géographiques des classes.....	133
5.4.	Etude de la variabilité du phytoplancton au niveau des biorégions Méditerranéennes	137

6. Conclusion:	140
Conclusion Générale :	142
Références:	147

Introduction Générale

Approximativement 70,8% de la surface de la Terre est recouverte par les océans et mers. L'océan global recèle plus de 60 % des phénomènes géophysiques et écosystémiques qui sont responsables de la régulation du climat du globe, et est aussi le siège d'une énorme production primaire due au phytoplancton. Ce dernier est le plancton végétal, c'est-à-dire l'ensemble des organismes végétaux chlorophylliens vivant en suspension dans l'eau. Il constitue « 50 % de la production primaire à l'échelle mondiale » et est à la base des réseaux trophiques océaniques. Le phytoplancton joue aussi le rôle de "pompe biologique de CO₂" en raison de sa capacité à fixer le dioxyde de carbone atmosphérique par photosynthèse pour produire de la matière organique. Son développement est fortement saisonnier et souvent constitué de successions de populations.

Le phytoplancton, les matières dissoutes et en suspension présentes en surface augmentent les propriétés d'absorption et de diffusion de l'eau de mer. Tous ces composants affectent, par conséquent, la signature spectrale de la lumière réfléchie et diffusée par la couche supérieure de l'océan. La "Couleur de l'Océan" varie ainsi du bleu profond des eaux pures au vert des eaux riches en phytoplancton ou en sédiments. La concentration en chlorophylle-a (Chl_a) est habituellement utilisée comme un indicateur de la quantité de biomasse présente dans l'océan. Ce pigment, qui est contenu dans la grande majorité des organismes photosynthétiques, absorbe fortement le rayonnement solaire dans les longueurs d'ondes bleues et faiblement dans la zone verte du spectre. Cette propriété spectrale permet de déterminer la Chl_a à partir des observations de la couleur de l'océan par satellite. Les images fournies par le premier capteur spatial de couleur de l'océan, CZCS (Coastal Zone Color Scanner) lancé en 1978, ont permis de bien comprendre la répartition globale de la biomasse marine (Gordon *et al.*, 1980, Hovis *et al.*, 1980).

L'observation de la couleur de l'océan depuis l'espace permet en effet d'étudier la variabilité spatiale et temporelle de la biomasse marine à l'échelle régionale et globale. De plus, le phytoplancton se comporte souvent comme un traceur passif, pour cela la télédétection de la couleur de l'océan peut être utilisée pour décrire la dynamique mesoéchelle des océans comme les tourbillons ou les méandres de grands courants tels que le Gulf Stream.

De nos jours, les capteurs multi-spectraux permettent d'obtenir des informations fines à une résolution spatio-temporelle satisfaisante, et laissent entrevoir la possibilité d'observer la variabilité non pas seulement de la Chl_a mais aussi des pigments secondaires qui sont associés au phytoplancton. Cette information est cruciale pour comprendre les écosystèmes marins et l'évaluation des flux d'énergie affectant les chaînes alimentaires. Pour quantifier cet effet, le signal satellitaire fourni par les capteurs de couleur de l'océan doit être analysé avec des algorithmes novateurs pour extraire l'information permettant d'obtenir les pigments secondaires.

Du point de vue régional, la mer Méditerranée présente un grand intérêt découlant de ses particularités de mer quasi fermée de grande dimension, de sa morphologie, de sa dynamique et son forçage externe qui en font un «modèle miniature» de l'océan mondial permettant une meilleure compréhension des processus complexes de l'océan global, qui va de la méso à l'échelle du bassin telle que la circulation thermohaline influencée par la formation d'eau dense (Lacombe *et al.*, 1981; Siokou-Frangou *et al.*, 2010). Les eaux de la mer Méditerranée sont parmi les plus pauvres en nutriment de l'océan mondial en surface tant qu'en profondeur, avec un gradient négatif Ouest-Est de nutriment et par conséquent, de phytoplancton et donc de Chl_a (Psarra *et al.*, 2005). Malgré ce caractère oligotrophe, la Méditerranée est un "hotspot" de biodiversité marine (Coll *et al.*, 2010) grâce à plusieurs caractéristiques tant du point de vue bathymétrique qu'hydrodynamique, mais vulnérable au changement climatique (The Mermex Group, 2011), d'où l'importance d'une étude approfondie de ce mésocosme.

Les études des communautés phytoplanctoniques à l'aide de la télédétection sont assez peu nombreuses en Méditerranée (Navarro *et al.*, 2014, 2017, Sammartino *et al.*, 2015, 2018; Di Cicco *et al.*, 2017). Ce domaine assez vaste nous a incités à orienter cette thèse autour de 3 axes principaux :

- 1- Des algorithmes ont été proposés pour estimer les pigments secondaires phytoplanctoniques à une échelle régionale. La plus grande étendue marine étudiée est l'Atlantique Sud, avec les travaux de (Bracher *et al.*, 2015). Le premier axe de ma thèse est constitué par le développement d'un algorithme valable à l'échelle de l'océan global pour l'estimation des pigments secondaires à partir des images

satellites de couleur de l'océan. Un premier aperçu d'une climatologie à l'échelle du globe des pigments secondaires est proposé.

- 2- Le deuxième axe de la thèse traite de l'application régionale de cet algorithme d'estimation de pigments secondaires en rapport avec l'observation de la dynamique spatio-temporelle des PFTs et PSC de la Méditerranée. Des exercices de validations sont effectués sur différents types de base de données (mesures de concentration pigmentaire et longues séries temporelles de comptage phytoplanctonique).
- 3- Finalement, étant donné la diversité des approches de régionalisation appliquée à la Méditerranée, le 3ème axe porte sur la proposition d'une biorégionalisation vis-à-vis des images satellitaires et observation in-situ. Le résultat sera interprété par la variabilité de la communauté phytoplanctonique obtenue par nos algorithmes développés. Cet axe conclut la thèse par une analyse synthétique vis-à-vis de la dynamique des masses d'eaux de la Méditerranée et des facteurs régissant la biomasse phytoplanctonique de ce milieu.

Ce manuscrit est articulé en 5 chapitres ; le premier présente des généralités sur les caractéristiques optiques de l'océan, sur le phytoplancton et sur sa télédétection. Ensuite le chapitre 2 se focalise sur le développement de l'algorithme global d'estimation des pigments secondaires, et est présenté sous forme d'article. Le chapitre 3 porte sur la description des caractéristiques hydrodynamiques et écologiques de la Méditerranée. Le chapitre 4 est subdivisé en 2 parties: en premier lieu nous élaborons une régionalisation de l'algorithme pour l'étude des PFT sous forme d'article, et en second lieu nous procéderons à l'observation des PSC avec un essai de représentation de la relation pigment/comptage phytoplanctonique. Finalement, Le chapitre 5 porte sur le résultat de la biorégionalisation de la Méditerranée et le manuscrit se conclut par une synthèse des différents aspects abordés.

Chapitre I: Télédétection Marine, Phytoplancton ; Généralités

1. Télédétection marine :

1.1. Propriétés optiques apparentes et inhérentes des océans :

L'océan n'est pas une masse d'eau pure. Ce gigantesque milieu aquatique en interaction (biologique, écologique, tectonique ...) constante avec le milieu aérien et aussi terrestre, contient une grande quantité de constituants diversifiés en type, en taille et en fonctions (matières organiques dissoutes, matières organiques en suspension, virus, bactéries, phytoplancton, zooplancton...). Ces constituants interagissent avec la lumière solaire par des processus de diffusion et d'absorption. Preisendorfer, (1961) a pu réaliser une distinction entre les propriétés optiques inhérentes (IOPs) dépendantes exclusivement des propriétés du milieu physique, et les propriétés optiques apparentes (AOPs) qui prennent en considérations les IOPs et les constituants de l'eau et d'éclairement en surface.

Outres les constituants du milieu, les propriétés optiques apparentes AOP sont dépendantes des conditions d'éclairement, c'est à dire de la distribution angulaire des photons incidents (cas de la lumière solaire), la nébulosité et les aérosols. La détermination de ces propriétés se fait à partir des mesures radiométriques des paramètres comme la luminance et l'éclairement.

La luminance L est le flux de puissance lumineuse reçu ou émis par une surface élémentaire dS . La luminance est un paramètre fondamental permettant de décrire le flux du rayonnement dans la direction (θ l'angle zénithal, ϕ l'angle azimutal) par unité d'angle solide et par unité d'aire en un point (x, y, z).

On note $L(\theta, \phi, \lambda, x, y, z)$ (en $W m^{-2} sr^{-1} nm^{-1}$)

Ou θ (angle zénithal) est compté à partir de l'horizontale. Si le soleil est directement sous l'observateur, le zénith a un angle de 90° . L'horizon a donc un zénith de 0° .

ϕ (angle azimutal) est compté à partir du nord à l'est, en sorte qu'un astre dans le nord à un azimut de 0° , un astre à l'est a un azimut de 90° .

λ , désigne la longueur d'onde, x, y, z les coordonnées du point.

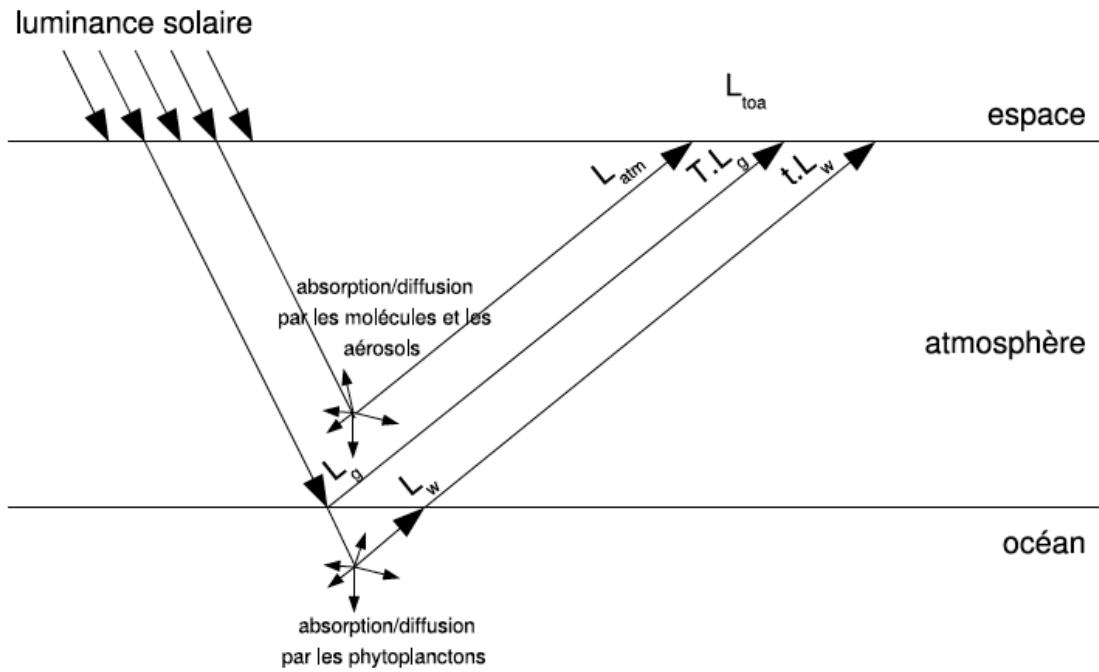


Figure I.1 *Trajet de la lumière solaire incidente, rétro-diffusée et réfléchi dans le système océan atmosphère (d'après Brajard, 2006)*

Le transfert radiatif étudie les modifications apportées par le milieu au flux radiatif L . Il est régi par des équations complexes (équations du transfert radiatif) qui n'ont pas de solutions analytiques et qui impliquent une description précise du milieu étudié. L'inversion de ces équations permettrait d'obtenir une description du milieu, au prix d'opérations complexes. En télédétection satellitaire de couleur de l'océan, on analyse le signal de luminance reçu au sommet de l'atmosphère L_{toa} qui provient du signal émis par le soleil qui traverse l'atmosphère et l'océan avant d'être rétrodiffusé vers l'espace (Figure I.1). Ce signal peut se décomposer de la façon suivante :

$$L_{toa} = L_{atm} + L_w \times t + L_g \times T \quad (I.1)$$

où L_{atm} est le signal de luminance rétrodiffusé par l'atmosphère, L_w la luminance sortant de la mer qui contient toute l'information sur le milieu marin pondéré par un terme de transmission diffuse t dû à la traversé de l'atmosphère, L_g représente la réflexion spéculaire du soleil (glitter) pondéré par un coefficient de transmission T . La difficulté de la télédétection marine provient de ce que le terme $L_{atm} \ll L_{toa}$ relation ou L_{atm} peut atteindre 90% de L_{toa} . Cette relation est directionnelle (voir eq. I.1).

On définit l'éclairement comme $E(\lambda)$ la somme des luminances intégrées provenant de toutes les directions reçues (ou émises) par une surface et est et dénoté par $E(\lambda)$.

Lorsque le rayonnement n'a lieu que dans une direction (c'est le cas du soleil) on peut écrire :

$$E(\lambda) = E_0(\lambda) \cos(\theta_s) \quad (1.2)$$

Où $E_0(\lambda)$ est la valeur de l'éclairement dans une direction perpendiculaire aux rayons du soleil et θ_s est l'angle entre la direction du rayonnement et la normale à la surface.

On définit alors une grandeur normalisée, la réflectance, qui permet de s'affranchir des fluctuations dues aux variations du rayonnement solaire et ainsi ne dépend et ainsi ne dépend que des caractéristiques du milieu et de la géométrie du rayonnement. Dans le cas d'un éclairement dû au soleil, la réflectance s'écrit :

$$R(\lambda, \theta, \varphi) = \frac{\pi L(\lambda, \theta, \varphi)}{E_0(\lambda) \cos(\theta_s)} \quad (1.3)$$

On en déduit une nouvelle formulation de l'équation (1.1)

$$R_{toa} = R_{atm} + R_w \times t + R_g \times T \quad (1.4)$$

Qui est l'équation utilisée par la suite pour trouver la composition de l'eau.

La réflectance, juste au-dessous de la surface de la mer, est due à la présence de différentes substances présentes sous forme particulaire ou dissoutes dans l'eau. L'analyse de la couleur de l'eau permet de retrouver par inversion ces substances. Des algorithmes ont été développés pour permettre de retrouver les constituants marins à partir de capteurs de couleurs de l'océan (Morel and Prieur, 1977; Gordon, 1997).

Les propriétés optiques apparentes de l'eau dépendent de sa composition (ex : l'indice de réfraction n , coefficients d'absorption "a" et de diffusion "b_b", fluorescence...). On a :

$$R_w(\lambda) = \frac{Gb(\lambda)}{a(\lambda) + b_b(\lambda)} \quad (1.5)$$

Ces propriétés sont additives, il est alors possible de les décomposer en plusieurs termes ayant différentes contributions liées aux constituants de l'océan. A titre d'exemple, les coefficients d'absorption a et de diffusion b_b sont décomposés de la sorte :

$$a(\lambda) = a_w(\lambda) + a_p(\lambda) + a_d(\lambda)$$

$$b_b(\lambda) = b_{bw}(\lambda) + b_{bp}(\lambda)$$

Avec w , p et d ont pour signification la contribution de l'eau de mer pure, des particules et de la matière dissoute.

1.2. Constituants de l'eau (optiquement significative):

La couleur de l'eau est essentiellement affectée par le phytoplancton, les sédiments, les substances chimiques dissoutes et aussi la structure moléculaire de l'eau. Le phytoplancton (qui fait l'objet de cette étude), contient de la chlorophylle, qui est le principal pigment qui absorbe la lumière des plantes photosynthétiques. Comme chez les végétaux terrestres, ce pigment absorbe le bleu et rouge du spectre et reflète le vert, de ce même concept, la biomasse phytoplanctonique est responsable du virage de la couleur de l'eau vers une couleur verdâtre.

Les sédiments absorbent et diffusent différentes longueurs d'ondes affectant par suite la couleur de l'eau. De même, les matières en suspension, qui absorbent essentiellement le bleu, reflètent ainsi une couleur jaunâtre.

D'où il est très important de prendre en considérations les constituants influençant sur l'optique de l'eau de mer en réalisant une étude sur la couleur de l'océan.

1.3. Types des Eaux : Cas 1 et Cas 2 :

La classification des eaux océaniques en "Cas 1" et "Cas 2" a été suggérée par (Morel and Prieur, 1977). Deux cas extrêmes peuvent être identifiés :

Le cas 1: qui constitue la majorité des eaux des océans (95%). La propriété optique de ce type est conditionnée exclusivement par la dynamique du phytoplancton renfermant la chlorophylle-a et de l'ensemble de ses produits associés (détritus organiques sous formes de particules ou de substances dissoutes et organismes hétérotrophes).

Le cas 1 est généralement typique des eaux de l'océan, éloignées des côtes et des embouchures riveraines.

Le cas 2: ce sont les eaux qui se trouvent au voisinage des zones côtières, qui sont influencées par le drainage des décharges et sédiments terrestres. Pour cette raison, les propriétés

optiques de ces eaux sont contrôlées par 3 facteurs : la chlorophylle, les sédiments mis en suspension et les matières dissoutes.

Dans la nature, on ne peut pas arriver à une séparation complète de ces 2 cas, et on peut avoir des eaux qui soient de type intermédiaire.

2. Le phytoplancton :

Le phytoplancton (plancton végétal) regroupe les organismes photosynthétiques microscopiques, vivant en suspension dans l'eau. Comme les végétaux terrestres, ces organismes sont munies de pigments chlorophylliens, ce qui leur a offert la capacité de synthétiser leurs propres matière organiques (glucides, lipides, protéines) à partir de substances minérales dissoutes dans l'eau de mer (le dioxyde de carbone et les sels nutritifs: nitrates, phosphates, silice, ammonium...) en utilisant l'énergie lumineuse : c'est le processus de la photosynthèse.

Il existe une grande diversité d'espèces, de tailles et de formes.

2.1. Notion de *Phytoplankton Size Class (PSC)* et *Phytoplankton Functional Type (PFT)* :

Sieburth, Smetacek and Lenz, (1978) ont établi une classification des organismes en fonction de leur taille (Phytoplankton Size Classe). Trois classes de taille sont essentielles dans l'étude de l'écologie phytoplanctonique :

- PICO phytoplancton ($T < 2 \mu\text{m}$)
- NANO phytoplancton ($2 \mu\text{m} < T < 20 \mu\text{m}$)
- MICRO phytoplancton ($T > 20 \mu\text{m}$)

En général, ces groupes présentent une répartition inégale dans l'océan, du point de vue géophysique et géographique. Le microphytoplancton est présent dans les hautes et moyennes latitudes, le nanoplancton est omniprésent dans l'ensemble des océans de la surface globale, à l'exception des gyres subtropicaux, alors que le picophytoplancton est dominant dans les gyres subtropicaux.

Une autre approche de classification du phytoplancton a été suggérée. Le nombre d'organismes, l'abondance relative des espèces, leurs traits biologiques et leurs rôles fonctionnels ont été utilisés pour décrire la structure des communautés de phytoplancton (Reynolds, 2006). Le concept de *Phytoplankton functional type (PFT)* a été introduit pour définir des groupes de phytoplancton par leurs fonctionnalités dans l'écosystème ;

La définition de «groupe fonctionnel» est sujette à différentes interprétations, regroupant le phytoplancton sur la base de divers rôles écologiques et d'exigences spécifiques. Ce terme

regroupe des espèces ayant des «traits morphologiques et physiologiques similaires» (Reynolds *et al.*, 2002): un groupe fonctionnel est composé de différentes espèces qui, à partir de la même ressource ou composante écologique, remplissent une fonction écologique commune (Blondel, 2003). Suivant leur métabolisme biogéochimique, les principaux groupes de phytoplancton taxonomiques peuvent être assemblés en quatre «groupes fonctionnels» spécifiques (Falkowski *et al.*, 2003; Litchman *et al.*, 2007; Sathyendranath *et al.*, 2014): fixateurs d'azote (propre aux procaryotes), calcifiants (y compris la classe taxonomique d'Haptophyceae, généralement connue sous le nom de coccolithophores), silicifieurs (représentés par la classe de Bacillariophyceae, généralement appelée diatomée, suivis par quelques producteurs de chrysophytes, silicoflagellés et xanthophytes, peu répandus en mer Méditerranée, et de diméthylsulfoniopropionate (DMSP) (faisant référence à des organismes de phytoplancton marin appartenant principalement au groupe des Dinoflagellates, et quelques Haptophytes).

2.2. Méthodes d'identification et de caractérisation du Phytoplancton :

Les méthodes d'identification et de caractérisation du phytoplancton sur le terrain ont beaucoup progressé depuis les premières approches fondées sur la microscopie optique jusqu'à l'utilisation actuelle de la télédétection par satellite. De nombreuses approches nouvelles ont été mises au point pour étudier le phytoplancton dans la mer d'une manière complémentaire à la microscopie. Les progrès incluent l'utilisation de la chromatographie liquide à haute performance (HPLC) pour caractériser les pigments en tant que marqueurs chimiques pour les groupes de phytoplancton, l'utilisation de la cytométrie en flux pour caractériser les cellules en fonction des propriétés d'auto-fluorescence et de diffusion de la lumière, la mise au point de techniques automatisées d'imagerie cellulaire et l'exploitation de méthodes moléculaires et d'approches de séquençage de gènes pour caractériser la biodiversité. Les propriétés optiques du phytoplancton, telles que les spectres d'absorption, de diffusion, d'excitation et d'émission de fluorescence, ont également été développées comme outils permettant de classer le phytoplancton en différents types. De plus, les classes de taille du phytoplancton peuvent être estimées sur le terrain avec filtration successive.

- Microscopie optique

La microscopie optique est la méthode traditionnelle d'identification du phytoplancton. L'épifluorescence et la microscopie électronique ont considérablement étendu les capacités

d'identification au microscope (Booth, 1993; MacIsaac et Stockner, 1993). Le microscope reste un outil essentiel pour identifier de nombreuses espèces de phytoplanctons en fonction de leurs caractéristiques morphologiques (Tomas and Hasle, 1997). Dans certaines méthodes, le phytoplancton est collecté sur des filtres ou des mailles avant l'identification et le comptage des cellules. Le microscope inversé ou méthode Utermöhl (Utermöhl, 1931) implique l'utilisation de chambres de décantation pour concentrer les échantillons avant analyse microscopique. Ceci est une approche efficace pour caractériser de nombreux types de microphytoplancton (c'est-à-dire une taille de cellule > 20 µm).

Malgré l'efficacité de la microscopie, la méthode présente certaines limites. Plus particulièrement, la microscopie demande du temps et dépend fortement de l'expertise de l'observateur pour discriminer les taxons. L'identification taxonomique étant basée sur des caractéristiques morphologiques, il peut être difficile, même pour les experts, d'identifier les cellules si celles-ci ont été endommagées lors de la manipulation. Le pico et le nanophytoplancton sont difficiles à identifier en microscopie optique en raison de leur petite taille et de l'absence de caractéristiques morphologiques distinctives.

- **Cytométrie en Flux**

Contrairement aux méthodes microscopiques manuelles, les cytomètres en flux sont réputés pour la rapidité avec laquelle ils sont capables de caractériser le phytoplancton dans un échantillon d'eau. Depuis environ trois décennies, cette méthode, adaptée à partir d'outils médicaux conçus pour le comptage des cellules sanguines, s'est révélée prometteuse pour l'étude du phytoplancton ainsi que des bactéries et virus de l'océan (Yentsch and Horan, 1989; Sosik, Olson and Armbrust, 2010).

En cytométrie de flux, les cellules de phytoplancton dans un échantillon d'eau de mer sont forcées individuellement à passer à travers une petite ouverture dans un champ lumineux. Lorsqu'une cellule traverse ce champ, ses propriétés optiques, telles que la fluorescence et la diffusion, sont mesurées. De cette manière, un grand nombre de cellules sont comptées rapidement (> 10⁵ cellules par seconde) à partir d'échantillons d'eau prélevés en mer.

La présence de la chlorophylle fluorescente permet de discriminer le phytoplancton des autres particules, et l'analyse cytométrique en flux permet le comptage, la quantification des propriétés cellulaires telles que la taille et la pigmentation, et un certain degré de discrimination taxonomique et optique basée sur la taille (par exemple, *Prochlorococcus*,

Synechococcus, picoeucaryotes, diatomées pennées, coccolithophorides) (Dubelaar, Gerritzen and Gerritzen, 2000; Collier, 2004).

La cytométrie en flux conventionnelle présente certaines limitations importantes. L'approche permet une analyse beaucoup plus rapide des cellules que la microscopie, mais les volumes d'échantillons sont encore faibles (de quelques dizaines de microlitres à quelques ml) et relativement peu d'échantillons discrets peuvent être traités par rapport aux échelles de temps et de la fréquence de changement dans l'environnement naturel.

- **Méthode HPLC**

L'analyse chromatographique des pigments est une approche non microscopique pour la caractérisation de la communauté de phytoplancton (Jeffrey and Mantoura, 1997; Roy *et al.*, 2011). La chromatographie en phase liquide à haute performance (HPLC), associée à la connaissance de la composition de pigment spécifique à un taxon, peut fournir des informations sur diverses classes de phytoplancton constituant des communautés naturelles (Mackey *et al.*, 1996).

Les pigments du phytoplancton peuvent être classés en trois groupes principaux: les chlorophylles (a, b et c), les caroténoïdes (les carotènes et leurs dérivés oxygénés, appelés xanthophylles) et les biliprotéines (phycoérythrine, phycocyanine et allophycocyanine). La chlorophylle-a, où plus spécifiquement la mono-vinyle-chlorophylle-a, est omniprésente dans tous les organismes phototrophes, à l'exception de *Prochlorococcus* qui possède la variante divinyl-chlorophylle-a.

La distribution de tous les autres pigments varie selon les taxons. Plusieurs pigments sont limités à des taxons spécifiques et peuvent donc être utilisés comme pigments de marquage - également appelés pigments de diagnostic où empreintes digitales de pigments - pour ces taxons (Jeffrey *et al.*, 1997).

L'HPLC permet la séparation, l'identification et la quantification de 25 à 50 pigments selon la méthode, en une seule analyse. Elle offre l'avantage unique de fournir simultanément une description complète de la composition de la communauté de phytoplancton sur l'ensemble du spectre de taille des cellules et une détermination précise de la concentration en chlorophylle-a, l'indice le plus largement utilisé pour estimer la biomasse de phytoplancton. Quelques pigments peuvent être affectés sans équivoque à une seule classe de phytoplancton. Par exemple, le divinyl-chlorophylle-a et le divinyl-chlorophylle-b sont

caractéristiques uniques des *Prochlorococcus* et à l'alloxanthine des cryptophytes. Cependant, l'interprétation des données de pigment n'est pas toujours simple et est sujette aux incertitudes liées à l'état physiologique et environnemental qui est la cause de variabilités de la composition des pigments. En particulier, certains pigments marqueurs sont partagés entre plusieurs taxons de phytoplancton, ce qui rend difficile l'identification des groupes.

2.3. Etude du phytoplancton depuis l'espace :

Les méthodes traditionnelles utilisant les données de couleur de l'océan sont axées sur l'estimation de la concentration de chlorophylle-a en tant qu'indice de la biomasse du phytoplancton. Les algorithmes standard donnant la chlorophylle (O'Reilly *et al.*, 1998, 2000) reposent principalement sur des relations empiriques qui utilisent des rapports de bande spectrale de la luminance (water-leaving radiance) ou de la réflectance, généralement situés dans la partie bleue et verte du spectre. Ces algorithmes sont relativement simples à appliquer et sont robustes face à une série de défis, notamment la mise en œuvre d'une correction atmosphérique importante (90% du signal). De plus on a à gérer la calibration des radiomètres embarqués qui peut dériver dans le temps. Les algorithmes d'inversion conçus pour la récupération des propriétés optiques intrinsèques (les coefficients de rétrodiffusion, les coefficients d'absorption), en plus de la concentration en chlorophylle, constituent une avancée notable par rapport aux algorithmes empiriques (Maritorena, Siegel and Peterson, 2002; Smyth *et al.*, 2002, 2006), et sont conçus pour donner de l'information sur la concentration totale de phytoplancton, la taille des particules, etc.

Récemment, les recherches se focalisent sur l'estimation de propriétés caractéristiques des différentes communautés phytoplanctonique, telles que le type dominant, la structure de taille ou la détection simultanée de plusieurs types fonctionnels, se sont multipliées. Aller au-delà de l'estimation de la chlorophylle pour détecter les types de phytoplancton à partir de données satellitaires nécessite une interprétation et une analyse plus approfondie des signaux de couleur de l'océan. Des études ont été réalisées avec des algorithmes dédiés permettant de détecter un seul taxon possédant des caractéristiques optiques distinctes connues pour affecter les signaux détectables depuis l'espace.

Bien que la validation de ces approches reste difficile, des tentatives ont été réalisées avec un certain succès, tel que la quantification spatiale des blooms de coccolithophoridés (Débora

Iglesias-Rodríguez *et al.*, 2006), les blooms de *Trichodesmium* près de la surface (Subramaniam and Carpenter, 1994; Westberry, Siegel and Subramaniam, 2005), des blooms de cyanobactéries (Wu *et al.*, 2010), des blooms de diatomées (Sathyendranath *et al.*, 2004) et des sargasses flottantes (Gower, Young and King, 2013). Certains algorithmes ont aussi permis d'identifier certains types de phytoplancton en utilisant des réflectances normalisées comme PHYSAT (Alvain *et al.*, 2005) ou encore l'abondance en chlorophylle (Uitz, Claustre, Morel and Stanford B Hooker, 2006; Aiken *et al.*, 2009).

En plus des classes biogéochimiques de phytoplancton, des classes de taille ont été également estimées à partir d'images satellitaires. Certaines de ces études sont basées sur des approches d'abondance utilisant la concentration en chlorophylle (Uitz, Claustre, Morel and Stanford B Hooker, 2006; Devred *et al.*, 2011; Hirata *et al.*, 2011) ou encore l'absorption du phytoplancton (Hirata *et al.*, 2008) et les relations avec la taille du phytoplancton. D'autres méthodes établissent une distinction entre les types de phytoplancton en utilisant les signatures optiques spectrales. Ces méthodes sont focalisées soit sur l'absorption (Ciotti and Bricaud, 2006; Devred *et al.*, 2006; Mouw and Yoder, 2010), soit sur les caractéristiques de diffusion (Loisel *et al.*, 2006; Kostadinov, Siegel and Maritorena, 2009). Les approches de diffusion qui sont basées sur la distribution totale de particules en suspension, ne différencient pas le phytoplancton de la charge totale de particules. Les approches basées sur la diffusion reposent sur l'hypothèse générale qui admet que la diffusion dans les eaux du large est essentiellement due au phytoplancton.

Les océanographes aujourd'hui ont accès en routine aux images de couleur de l'océan observées par les capteurs satellitaires tels que SeaWiFS, MODIS, VIIRS, MERIS et OLCI, qui peuvent donner des images de l'océan global à des échelles temporelles de 1 à 2 jours. D'autres stratégies d'observation (navires, bouées dérivantes ou bouées ancrées par exemple) ne peuvent pas avoir la couverture et la résolution des capteurs spatiaux, d'où la forte motivation pour extraire des informations sur les PFT à partir de capteurs satellitaires, même s'il est admis que l'identification des PFT à travers les méthodes de télédétection satellitaires ont une résolution grossière et des incertitudes relativement grandes. En effet, l'utilité des informations obtenues par satellite augmenterait fortement si les incertitudes associées aux produits étaient bien évaluées. Les informations de télédétection peuvent apporter des informations uniques et essentielles sur la réponse des types fonctionnels du

phytoplancton aux événements climatiques régionaux tels qu'El Niño. Par exemple, il est généralement admis que les diatomées ne contribuent pas significativement à la biomasse de phytoplancton dans la région équatoriale du Pacifique (Kobayashi and Takahashi, 2002; Dandonneau *et al.*, 2004), mais des blooms exceptionnels ont été signalés dans cette zone à partir d'observations in situ (Coale *et al.*, 1996) et leur large couverture spatiale a récemment été observée par télédétection satellitaire (Alvain *et al.*, 2008). Un bloom de diatomées peut se produire lorsque la remontée d'eau équatoriale est intense et ne se limite pas à la région côtière, ce qui est observé dans un événement tel que La Niña et qui donne lieu à des exportations massives de carbone organique vers le fond (Greene *et al.*, 1994; Cavender-Bares, Karl and Chisholm, 2001).

Chapitre II: Télédétection de la Variabilité des Pigments Phytoplanktoniques Secondaires par Méthode Neuronale ; Approche Globale

1. Introduction:

Au cours de la dernière décennie, la télédétection des propriétés optiques de la surface océanique a fourni des vues synoptiques de l'abondance et de la distribution des constituants de la surface de la mer; la lumière solaire interagit avec l'eau de mer par réflexion et absorption, de telle sorte que le signal ascendant transmis au satellite contient des informations relatives à la composition de l'eau de mer. Dans les eaux du large, loin de la côte, la lumière du soleil interagit principalement avec la forme des cellules du phytoplancton (réflexion), les pigments (absorption) et les débris (par exemple, le CDOM). Par conséquent, la lumière vue par le capteur satellite contient essentiellement des informations sur le phytoplancton. La couleur de l'océan a été une plate-forme efficace pour estimer la concentration de Chlorophylle a (Chla) dans les eaux de surface, fournissant des mesures synoptiques sur l'océan mondial (Longhurst et al., 1995; Antoine et al., 1996; Behrenfeld et Falkowski, 1997; Behrenfeld et al., 2005; Westberry et al., 2008).

En plus de la Chla, de nombreux pigments du phytoplancton interagissent avec la lumière du soleil, tels que la chlorophylle b (Chlb), la chlorophylle c (Chlc) et les caroténoïdes photosynthétiques, ou pour protéger la Chla et d'autres pigments sensibles au « photodamaging », tels que les caroténoïdes photoprotecteurs. Certains pigments ne sont présents que dans des groupes de phytoplancton spécifiques et sont donc des pigments indicateurs pour permettre leur identification, par exemple la fucoxanthine dans les diatomées et la péridinine dans les dinoflagellés (Letelier et al., 1993; Vidussi et al., 2001). L'identification de ces pigments par télédétection devrait permettre d'obtenir une distribution spatio-temporelle sans précédente, ce qui donnerait des indications importantes sur la composition du phytoplancton, l'absorption de la lumière, l'état physiologique (Behrenfeld et Boss, 2006) et la dynamique du réseau trophique marin et de la productivité primaire (Bracher et al., 2017). Il a été reconnu que la détection des principales caractéristiques de la communauté de phytoplancton à partir d'images satellitaires constituait un défi majeur dans le domaine de l'optique océanique.

L'absorption de la lumière par le phytoplancton porte les empreintes de différents types de pigments et peut être mesurée par mesures optiques. Plusieurs études récentes ont étudié le potentiel d'utilisation de données satellitaires pour calculer les concentrations de surface de pigments autres que la Chla.

Chase et al. (2013) ont projeté un grand ensemble de données mondiales de mesures d'absorption hyperspectrale de particules sur une base de fonctions gaussiennes et ont évalué ces fonctions gaussiennes spécifiques en relation avec l'absorption par des pigments ou des groupes de pigments spécifiques. La méthode a fourni des résultats robustes pour obtenir des concentrations de Chla, Chlb, Chlc, phycoérythrine (PE), des pigments photoprotecteurs et des caroténoïdes photosynthétiques.

Organeli et al. (2013) ont utilisé une approche multivariée appliquée au spectre de la quatrième dérivée du phytoplancton ou à des données d'absorption de particules pour extraire la Chla, sept pigments diagnostiques et les classes de taille de phytoplancton correspondantes sur le site de Boussole en mer Méditerranée.

Pan et al. (2010) ont mis au point des algorithmes empiriques basés sur des rapports de réflectance permettant d'estimer les concentrations des pigments principaux. Des algorithmes de rapport de bande ont été développés à partir de mesures radiométriques et de données de pigments mesurées dans les eaux côtières du nord-est des États-Unis. Cet algorithme a permis de calculer les concentrations de Chla, Chlb, Chlc et de neuf caroténoïdes différents. Cependant, de tels algorithmes de rapport de bande nécessitent une très grande base de données (> 400 collocations avec des données satellitaires) à partir d'une région spécifique pour dériver un algorithme régionalisé robuste.

Bracher et al. (2015) ont développé des modèles pour estimer les pigments de phytoplancton à partir de mesures hyperspectrales in situ et satellitaires de réflectance marine ($R_{rs}(\lambda)$) appliquées à l'océan Atlantique. Ces modèles sont basés sur une analyse empirique des fonctions orthogonales (EOF) des spectres $R_{rs}(\lambda)$ normalisés.

Dans ce travail, nous proposons d'utiliser des cartes auto-organisées (SOM) - (Kohonen, 2013), afin de mettre en évidence la relation entre les données satellitaires et in situ mesurées à la surface de l'océan. Les SOM sont des classifieurs neuronaux à apprentissage non supervisé qui ont été appliqués avec succès à la télédétection (Richardson, Risien et Shillington, 2003; Ehsani et Quiel, 2008; Hu et Weng, 2009; Iskandar et Iskhaq, 2010; Gorricha et Lobo, 2012), et plus spécifiquement aux mesures de la couleur des océans pour la caractérisation des aérosols (Niang et al., 2006; Diouf et al., 2013), la classification des spectres de radiance (Ainsworth et Jones, 1999) ainsi que pour l'analyse des spectres

d'absorption du phytoplancton (Chazottes et al., 2006; Bricaud et al., 2007). Cette étude vise à fournir une méthode globale, précise et robuste pouvant être utilisée pour extraire la composition de pigments HPLC du signal de télédétection et ainsi permettre de suivre la dynamique du phytoplancton.

L'utilisation de SOM permet d'attribuer automatiquement différents motifs de signal satellite et la composition de pigment associée en fonction des similitudes en termes de forme et d'amplitude. De plus, la méthode nous permet de prendre en compte un plus grand nombre de données in situ de manière très efficace en termes de temps de traitement, avec une flexibilité et une fiabilité accrues.

Dans ce chapitre, on décrit l'algorithme de base de la SOM, puis en suite le développement de la méthode d'estimation des pigments secondaires qui est présenté sous forme d'article.

2. Cartes de Kohonen

Une carte auto-organisatrice (Self Organizing Map) est un type réseau de neurones d'apprentissage non supervisé qui permet de quantifier des données multi dimensionnelles et de les visualiser dans un espace de faible dimension, généralement 1 ou 2 dimensions. Leur propriété principale est d'introduire au niveau de l'apprentissage un ordre topologique qui permet par la suite des visualisations très efficaces des données. L'algorithme d'apprentissage d'une SOM est une extension de l'algorithme bien connu des k-moyennes.

2.1. Idée de base

Ces modèles de représentation des données, utilisées en Intelligence Artificielle, sont inspirées comme beaucoup des modèles neuronaux par la biologie. Il s'agit pour les SOM de reproduire un des principes de fonctionnement du cerveau des vertébrés : des stimuli de même nature excitent une région du cerveau bien particulière. Les neurones sont organisés dans le cortex de façon à interpréter tous les types de stimuli imaginables. De la même manière, la carte auto organisatrice se déploie de façon à représenter de façon ordonnée un ensemble de données. Dans ce cadre chaque neurone se spécialise pour représenter un sous-ensemble bien particulier de ces données selon leurs similitudes. Techniquement, les neurones réalisent une quantification vectorielle de l'espace de données, c'est à dire qu'ils discrétisent l'espace des données, et affectent à chaque zone un représentant dans cet espace : le vecteur référent w .

2.2. Architecture

Si l'on décrit les cartes auto-organisatrices à l'aide d'un formalisme réseau de neurones, son architecture est constituée d'une grille (le plus souvent uni- ou bidimensionnelle). Dans chaque nœud de la grille se trouve un neurone. Chaque neurone est lié à un vecteur référent, responsable d'une zone dans l'espace des données (appelé encore espace d'entrée). Chaque neurone représente donc un sous ensemble des données initiales dont le représentant est le vecteur référent. Dans une carte auto-organisatrice, les vecteurs référents fournissent une représentation discrète de l'espace d'entrée. Ils sont positionnés de telle façon qu'ils mettent en évidence la forme topologique de l'espace d'entrée (multi dimensionnel) en explicitant les relations de voisinage sur la grille (en dimension 1 ou 2). Ces voisinages permettent une indexation facile (via les coordonnées dans la grille). Le voisinage est formalisé par l'utilisation

d'une fonction noyau K positive, symétrique et telle que $K(0) = 1$ et $\lim_{|x| \rightarrow \infty} K(x) = 0$. Elle permet de définir la zone d'influence autour de chaque nœud.

Pour résumer une carte SOM se caractérise par :

- Une quantification vectorielle : on approxime chaque point dans l'espace d'entrée par le vecteur référent le plus proche
- Un classifieur : on affecte à chaque neurone de la grille une étiquette correspondante à une classe; tous les points de l'espace d'entrée qui se projettent sur un même neurone appartiennent à la même classe et sont proches dans l'espace des données.

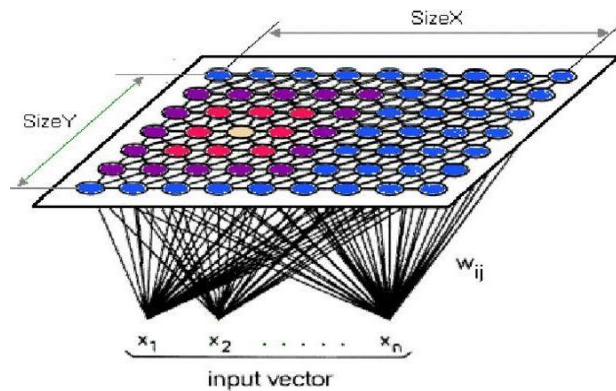


Figure II.1 : Représentation de l'architecture typique d'une SOM. x_i représentent les composantes d'un vecteur d'entrée, w_{ij} connote le vecteur référent d'un neurone j qui a capté l'observation i .

2.3. Algorithme d'apprentissage

Il existe plusieurs versions de l'algorithme des cartes topologiques. Nous présentons ici la forme « Batch » qui est celle utilisée dans cette thèse. L'ensemble D constitue l'ensemble d'apprentissage et l'apprentissage consiste à minimiser la fonction de coût qui est exprimée par :

$$J_{SOM}^T(\chi, W) = \sum_{z_i \in D} \sum_{c \in SOM} K^T(\delta(c, \chi(z_i))) \|z_i - w_c\|^2 \quad (2.1)$$

Où $c \in SOM$ indice du neurone de la grille SOM, $\chi(z_i)$ est la fonction d'allocation qui attribue chaque élément z_i de D à son vecteur référent $w_{\chi(z_i)}$. $\delta(c, \chi(z_i))$ est la distance discrète sur la grille de la SOM entre un neurone c et l'observation z_i attribuée.

Cette fonction de coût doit être minimisée par rapport au deux composantes χ et W . Il est facile de voir que pour un choix d'un ensemble de référents W , la fonction d'affectation qui minimise J_{SOM}^T (pour W fixé) est donnée par la relation suivante :

$$\chi(z_i) = \underset{c \in SOM}{\operatorname{argmin}} \left(\sum_{r \in SOM} K^T(\delta(r, c)) \|z_i - w_c\|^2 \right) \quad (2.2)$$

où la minimisation est prise sur l'ensemble des indices de la carte C. D'autre part, pour un choix de χ fixe, la fonction J^T_{som} est une fonction quadratique par rapport à W , elle atteint son minimum pour $\frac{\partial J_{som}^T}{\partial W} = 0$, ceci donne pour chaque i ($1 \leq i \leq n$) :

$$w_c^t = \frac{\sum_{i=1}^n K^T(\delta(X(z_i), c)) z_i}{\sum_{i=1}^n K^T(\delta(X(z_i), c))} \quad (2.3)$$

Nous présentons d'abord l'algorithme itératif qui permet de déterminer un minimum local de la fonction J^T_{som} pour un paramètre T fixé. On suppose connaître un ensemble de référents W^{t-1} . Cet algorithme se décompose en deux phases, phase d'affectation et phase de minimisation, comme le montre l'algorithme suivant :

2.3.1. Algorithme T fixé :

Phase d'Initialisation :

- Choix de la structure et de la taille de la carte ; Initialisation de l'ensemble W^0 des p référents initiaux, (en général, d'une manière aléatoire).
- Choix du nombre d'itérations N_{iter} ;

Etape itérative $t=1:N_{iter}$

- Phase d'affectation: l'ensemble des référents (W^{t-1}) déterminés à l'itération précédente étant fixé, définir la nouvelle forme d'affectation χ_t par l'équation (2.2) ;
- Phase de minimisation : χ_t tant fixé, choisir le nouvel ensemble des référents dont les éléments sont définis par la relation (2.3).

2.3.2. Algorithme standard

Phase d'Initialisation :

- Choix de la structure et de la taille de la carte, et des valeurs des paramètres d'apprentissage : T_{min} , T_{max} et le nombre d'itérations N_{iter} ;

- Initialisation de l'ensemble W^0 des p référents initiaux, (en générale, d'une manière aléatoire) ;

Etape itérative de $t = 1$ à N_{iter} faire :

- Mise à jour du paramètre T :

$$T(t) = T_{max} \times (T_{min}/T_{max})^{t/(N_{iter}-1)}$$

- Phase d'affectation: l'ensemble des référents (W^{t-1}) déterminés à l'itération précédente étant fixé, définir la nouvelle forme d'affectation χ_t par (2.2) ;
- Phase de minimisation : χ_t étant fixé, choisir le nouvel ensemble des référents dont les éléments sont définis par la relation (2.3).

2.4. Propriétés

- Similitude des densités dans l'espace d'entrée : la carte reflète la distribution des points dans l'espace d'entrée. Les zones dans lesquelles les vecteurs d'entraînement z_i sont tirés avec une grande probabilité d'occurrence sont cartographiées avec une meilleure résolution.
- Préservation des relations topologiques : des neurones voisins dans la grille occupent des positions voisines dans l'espace d'entrée (préservation des voisinages de la grille).

3. Estimation des pigments secondaires phytoplanctoniques à partir des images satellitaires en utilisant les SOM

Résumé :

Cette étude présente une méthode d'estimation des pigments secondaires de phytoplancton à partir d'observations satellitaires de la couleur de l'océan. Nous avons d'abord compilé un grand ensemble de données d'entraînement composé de 12 000 échantillons; chaque échantillon est composé de dix concentrations de pigment mesurées par HPLC de phytoplancton in situ, de produits GlobColour de concentration de Chlorophylle-a (Chla) et de réflectance de télédétection ($R_{rs}(\lambda)$) à différentes longueurs d'onde, ainsi que de mesures de la température de surface de la mer AVHRR (SST). L'ensemble de données résultant regroupe une grande variété de situations rencontrées entre 1997 et 2014.

La relation non linéaire entre les composantes in situ et satellitaires a été identifiée à l'aide d'une carte auto-organisatrice (SOM). La SOM a permis d'obtenir des estimations fiables de la concentration de Chla et de neuf pigments différents à partir d'observations par satellite. Une procédure de cross-validation a montré que les estimations étaient robustes pour tous les pigments ($R^2 > 0.75$ et RMSE moyen = $0,016 \text{ mg.m}^{-3}$).

Une association cohérente de plusieurs pigments de phytoplancton indiquant la dynamique spécifique à un groupe de phytoplancton a été montrée à l'échelle globale. Nous avons également calculé les incertitudes pour l'estimation de chaque pigment.

Estimation of secondary phytoplankton pigments from satellite observations using self-organizing maps (SOM)

Roy El Hourany ^{1,2*}, Marie Abboud-Abi Saab ³, Ghaleb Faour ², Olivier Aumont ¹, Michel Crépon ¹ & Sylvie Thiria ^{1,4}

¹ IPSL/LOCEAN, Sorbonne Université (Université Paris VI, CNRS, IRD, MNHN), Paris, France,

² National Center for Remote Sensing, National Council for the Scientific Research, Beirut, Lebanon,

³ National Center for Marine Sciences, National Council for the Scientific Research, Batroun, Lebanon,

⁴ Versailles Saint-Quentin-en-Yvelines University, Versailles, France

Corresponding author: Roy El Hourany (roy.elhourany@locean-ipsl.upmc.fr)

Journal of Geophysical Research: Oceans,

Submitted on 7 August 2018, Accepted on 6 February 2019

DOI: <https://doi.org/10.1029/2018JC014450>

Key points

- The dynamic of phytoplankton communities can be observed while reconstructing accessory pigments variability from ocean color observations
- Self-Organizing Maps allow an accurate retrieval of different phytoplankton pigment concentrations from satellite observations
- A consistent global approach is established to estimate accessory pigment concentrations along with the uncertainties

Abstract

This study presents a method for estimating secondary phytoplankton pigments from satellite ocean colour observations. We first compiled a large training data set composed of 12000 samples; each sample is composed of ten in-situ phytoplankton HPLC measured pigment concentrations, GlobColour products of Chlorophyll-a concentration (Chla) and remote sensing reflectance ($R_{rs}(\lambda)$) data at different wavelengths, in addition to AVHRR sea surface temperature measurements (SST). The resulting dataset regroups a large variety of encountered situations between 1997 and 2014. The non-linear relationship between the in-situ and satellite components was identified using a self-organizing map (SOM), which is a neural network classifier. As a major result, the SOM enabled reliable estimations of the concentration of Chla and of nine different pigments from satellite observations. A cross-validation procedure showed that the estimations were robust for all pigments ($R^2 > 0.75$ and

an average RMSE=0.016 mg.m⁻³). A consistent association of several phytoplankton pigments indicating phytoplankton group specific dynamic was shown at a Global scale. We also showed the uncertainties for the estimation of each pigment.

Plain Language Summary:

The knowledge of phytoplankton variability is essential to the understanding of the marine ecosystem dynamics and its response to environmental changes. This paper presents a new approach to estimate phytoplankton pigment concentrations from satellite observations by using an artificial neural network, the so-called Self Organizing Map. This neural network was calibrated using a large data set of in-situ pigment observations from oceanic cruises along with ocean colour satellite data provided by the Globcolour project and AVHRR sea surface temperature. This approach allows an accurate estimation of phytoplankton pigment concentrations and their related uncertainties. Moreover, the method allows to reproduce the spatio-temporal variability of pigments concentration and the dynamics of phytoplankton groups. A particular attention is given to the Southern Ocean whose phytoplankton communities are specific.

1. Introduction

Marine ecosystems are a major sink for atmospheric CO₂ (Häder *et al.*, 2014). The net transfer of CO₂ from the atmosphere to the oceans and then sediments, is mainly a direct consequence of the combined effect of the water solubility (physical pump) and the biological pump (Hülse *et al.*, 2017). The biological carbon pump is a key natural process and a major component of the global carbon cycle that regulates atmospheric CO₂ levels, transferring both organic and inorganic carbon fixed by phytoplankton in the euphotic zone to the ocean interior (Chisholm, 1995; Hülse *et al.*, 2017). Understanding the response of the biological carbon pump to global change is required to accurately predict the future impact of the increase in atmospheric CO₂ due to the human activities (Passow and Carlson, 2012). The dissolution of anthropogenic CO₂ in the ocean and the subsequent formation of carbonic acid has already resulted in a decrease of 0.1 pH unit and will continue to lower pH by an additional 0.2–0.3 pH units by the end of the century (Bhadury, 2015). This decline in ocean pH is referred to as ocean acidification (Orr *et al.*, 2005). At the same time, warming should increase the mean surface temperatures by an average of 3 °C, leading to longer periods of stratification with fewer deep mixing events (Sarmiento *et al.*, 2004) and a less efficient physical pump. Increased stratification is expected to lead to nutrient limitation and an increase in average irradiance in the euphotic layer, where phytoplankton grow (Schulz *et al.*, 2007). The knowledge of the space and time heterogeneity of phytoplankton abundance is essential to understand the marine ecosystem dynamics and responses to environmental changes (Mann and Lazier, 2006).

Phytoplankton are distributed among a large number of groups of microscopic photosynthesizing protists and cyanobacteria which contribute nearly 50% to the total primary production of Earth by fixing about 50 Gt carbon per year (Baumert, 2008). In order to understand and quantify the phytoplankton abundance and characteristics at the surface, in-situ methods evolved over the past years passing from microscope count method to flow cytometer and HPLC pigment diagnosis. In fact, HPLC enables the identification of 25 to 50 phytoplankton pigments within a single analysis for which each phytoplankton group is associated with specific diagnostic pigments. Therefore, HPLC measurements are now widely used to determine phytoplankton species in-situ. However, these methods are time-

demanding with a low cost-effectiveness and their applications are limited to accessible zones where sea water sampling can be made. Besides, it has been shown that many pigments overlap between unrelated groups leading to misinterpretation. Nevertheless, these methods are considered as references for phytoplankton identification.

In the last decade, remote sensing of surface optical properties has provided synoptic views of the abundance and distribution of sea surface constituents; the downward sunlight interacts with the seawater through backscattering and absorption in such a manner that the upward signal transmitted to the satellite contains information related to the composition of seawater. In the open ocean far from the coast (case 1 waters), the sunlight mainly interacts with the phytoplankton cell shape (backscattering), pigments (absorption) and debris (e.g. CDOM). Therefore, the light seen by the satellite sensor contains information on phytoplankton. Ocean color has been an effective platform to estimate the Chlorophyll-a concentration (Chla) in surface waters, providing synoptic measurements over the world ocean (Longhurst *et al.*, 1995; Antoine *et al.*, 1996; Behrenfeld and Falkowski, 1997; Behrenfeld *et al.*, 2005; Westberry *et al.*, 2008).

Besides Chla, many other pigments in phytoplankton interact with the sun light, such as chlorophyll b (Chlb), chlorophyll c (Chlc) and photosynthetic carotenoids (PSC), or in protecting Chla and other sensitive pigments from photodamage, such as photoprotective carotenoids (PPC). Some pigments only occur in specific phytoplankton groups and are thus indicator pigments for their identification, e.g., fucoxanthin in diatoms and peridinin in dinoflagellates (Letelier *et al.*, 1993; Vidussi *et al.*, 2001). The identification of these pigments by remote-sensing would provide an unprecedented spatio-temporal distribution (Kostadinov *et al.*, 2017), that would give powerful insights on the phytoplankton composition, light absorption, physiological state (Behrenfeld and Boss, 2006) and the dynamics of the marine food web and marine productivity (Bracher *et al.*, 2017a). This was early recognized that detection of major characteristics of the phytoplankton community from remote sensing images was a major challenge in ocean optics.

Phytoplankton absorption bears the imprints of different types of pigments and can be measured by optical measurements. Several recent studies have investigated the potential of using continuous optical data to derive surface concentrations of pigments other than Chla.

Chase et al. (2013) decomposed a large global data set of hyperspectral particulate absorption measurements into Gaussian function components and assessed the magnitude of specific Gaussian functions in relation to the absorption by specific pigments or pigment groups. The method provided robust results for obtaining concentrations of Chla, Chlb, Chlc, phycoerythrin (PE), PPC and PSC. Organelli et al. (2013) used a multivariate approach applied to fourth-derivative spectra of phytoplankton or particulate absorption data to retrieve Chla, seven diagnostic pigments and the corresponding phytoplankton size classes at the Boussole site in the Mediterranean Sea. Pan et al. (2010) developed empirical algorithms based on reflectance ratios to approximate key phytoplankton pigment concentrations. The band-ratio algorithms were developed from radiometric measurements collocated with pigment data measured in northeastern US coastal waters. This algorithm has successfully derived the concentrations of Chla, Chlb, Chlc and nine different carotenoids. However, such band-ratio algorithms require a very large database (>400 collocations with satellite data) from a specific region to derive robust regionalized algorithm. Bracher et al. (2015) developed models to estimate phytoplankton pigments from hyperspectral in-situ and satellite measurements of remote-sensing reflectance ($R_{rs}(\lambda)$) which were applied to the Atlantic Ocean. These models were based on empirical orthogonal functions (EOF) analysis of normalized $R_{rs}(\lambda)$ spectra.

In the present work, we propose to use self-organizing maps (SOM)- (Kohonen, 2013), in order to evidence the relationship between satellite and in-situ data measured at the ocean surface. The SOM are unsupervised neural classifiers which were successfully applied in remote sensing for pattern extraction (Richardson, Risien and Shillington, 2003; Ehsani and Quiel, 2008; Hu and Weng, 2009; Iskandar and Iskhag, 2010; Gorricha and Lobo, 2012), and more specifically to ocean color measurements for aerosol characterization (Niang *et al.*, 2006; Diouf *et al.*, 2013), radiance spectra classification (Ainsworth and Jones, 1999) as well as for phytoplankton absorption spectra analysis (Chazottes *et al.*, 2006; Bricaud *et al.*, 2007). This study aims at providing a global, accurate and robust method that can be used to retrieve the HPLC pigments composition from the remote sensing signal and consequently permits to track the phytoplankton dynamics. The use of SOM gives the ability to automatically assign different patterns of satellite signal and the associated pigment composition based on the similarities in terms of shape and amplitude. In addition, the method allows us to take into

account a larger in-situ dataset in a very efficient manner in terms of processing time, with a higher flexibility and reliability.

The paper is articulated as follows. In section 2, we briefly describe the global ocean database co-located with satellite data (D_{pigment}) used to calibrate the SOM we developed to estimate the pigment concentration. The SOM and the calibration procedure are described in section 3. In section 4, we present the results and the validation of the method. In section 5, we estimate the pigment concentration in the global ocean. The results are discussed in section 6. A conclusion is presented in section 7.

2. Materials

The proposed method is based on a statistical clustering of complementary available remotely sensed parameters and in-situ HPLC pigment data by using a SOM. For that, we used a rigorous dataset of in-situ HPLC pigment data collocated with satellite data, regrouping a large diversity of encountered situations and multiple parameter combinations between 1997 and 2014. Our approach is based on the spectral diversity of the water leaving signal, in conjunction with the variability of the phytoplankton pigments. The application of such approach will allow the spatio-temporal reconstruction of pigments. In this section, both the data and the methodology are described and explained.

2.1. GlobColour Data:

To extend existing time-series beyond that provided by a single satellite sensor, the ESA initiated the GlobColour project (<http://www.globcolour.info/>) to develop a satellite based ocean colour data set to support global carbon-cycle research. It aims at satisfying the scientific requirement for a long (10+ year) time-series of consistently calibrated global ocean colour information with the best possible spatial coverage. This has been achieved by merging data from SeaWiFS, MODIS, VIIRS, MERIS and OLCI.

The GlobColour project provides a continuous data set of merged level 3 (Mapped, 4km) daily remote sensing reflectance ($R_{rs}(\lambda)$). This product is generated for each instrument, using the corresponding level 2 data. The merged $R_{rs}(\lambda)$ are then computed as the weighted average of all single-sensor products. The 547-560 nm bands are submitted to a specific processing just

before averaging to prepare a more consistent merging between the sensors. These bands were fitted to the $R_{rs}(555)$ of SeaWiFS. The main reason behind this choice is that SeaWiFS is widely considered as the highest quality sensor with the best match to in-situ observations, and is commonly used in peer literature (Belo Couto *et al.*, 2016). The primary cause for using this dataset was to increase the number of match-ups between HPLC in-situ data and satellite parameters.

Basically, ocean-colour sensors measure the diffuse sunlight backscattered by the ocean. The principle of detecting phytoplankton groups from space relies on their spectral contributions to the $R_{rs}(\lambda)$ which in turn is determined by the spectral absorption (a , m^{-1}) and backscattering (b_b , m^{-1}) coefficients of the ocean (pure water and various particulate and dissolved matters) using this simplified formula (Morel and Gentili, 1996):

$$R_{rs}(\lambda) = G \times b_b(\lambda) / (a(\lambda) + b_b(\lambda)) \quad (1)$$

where G is a parameter mainly related to the geometry of the situation (sensor and solar angles) but also to environmental parameters (wind, Inherent Optical Properties, aerosols).

The contribution of the phytoplankton to the $R_{rs}(\lambda)$ can be explained by its pigment content, which absorbs the light at specific wavelengths, and its physical structure which scatters the light as a function of the wavelength.

In this study, we used $R_{rs}(\lambda)$ at four different wavelengths (412, 443, 490 and 555 nm). These $R_{rs}(\lambda)$ were downloaded between 1997 and 2014. Each of these $R_{rs}(\lambda)$ depends on several biogeochemical and physical factors such as the influence of phytoplankton pigments on the variability of $R_{rs}(443)$, $R_{rs}(490)$ and $R_{rs}(555)$, especially the 490 nm wavelength due to the maxima of absorption of several secondary pigments near this wavelength.

We also added the Chla concentration which gives an important information on the total phytoplankton abundance. The Chla was estimated with the OC5 algorithm (Gohin, 2011) and is provided on a daily basis via the GlobColour portal. The use of the 4 $R_{rs}(\lambda)$ along with the Chla OC5 in the classification procedure will help to identify the non-linear relationship between $R_{rs}(\lambda)$ -Chla-pigments and then to accurately estimate the pigment composition. The $R_{rs}(\lambda)$ data were the primary variables to be used in this approach, evaluating the optical relationship with the corresponding pigment composition. This is mainly due to the diversity

of the $Rrs(\lambda)$ spectra that will be analyzed by the SOM in terms of shape and amplitude. Adding the Chla OC5 as a second parameter, allows to take into account the inter-relationship between auxiliary pigments and Chla, and the relationship between Chla OC5 and $Rrs(\lambda)$ (on which the OC5 algorithm is based).

2.2. AVHRR SST Data

Since the pigment concentrations are characterized by a well-defined seasonal cycle, the use of SST in the algorithm of pigment retrieval permits to better fit the relationship between the in-situ HPLC and satellite data (Pan *et al.*, 2013). Therefore, adding SST should promote the identification of pigments co-varying with the seasonal variability of this physical factor. The SST data were downloaded at 4 km resolution and at a daily frequency between 1997 and 2014, using a product of the AVHRR instruments aboard NOAA polar-orbiting satellites: version 5.3 level 3 global 4km sea surface temperature (Casey *et al.*, 2010; Saha *et al.*, 2018) (https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.nodc:AVHRR_Pathfinder-NCEI-L3C-v5.3). Current retrieval algorithms for SST from AVHRR are based largely upon the multi-channel sea surface temperature (MCSST) algorithm (McClain *et al.*, 1985) which may be written as:

$$SST=A_1 + A_2*T_4 + Y (T_4 - T_5) \quad (2)$$

where A_1 and A_2 are constants determined through a least-squares fit to in-situ data, T_4 , T_5 , are brightness temperatures as derived from channels 4, 10.3-11.3 μ m and 5, 11.5-12.5 μ m, and Y is a weighting factor based on the knowledge of known absorption coefficients (Emery *et al.*, 1994).

2.3. HPLC pigment Dataset

In parallel to the remotely sensed data, the in-situ HPLC database used in this work is a compilation of different published datasets such as MAREDAT, NOMAD and SeaBASS, and several oceanic campaigns such as GEP&CO, TARA Ocean Expedition, Polarstern and Labrador Sea expeditions (Table A1.1). This database gathers 10 different pigments: Total Chlorophyll-a (Chla), Divynil-Chlorophyll-a (DVChla), Chlorophyll-b (Chlb), Divynil-Chlorophyll-b (DVChlb), 19'Hexfucoxanthin (19HF), 19'Butfucoxanthin (19BF), Fucoxanthin (Fuco), Peridinin (Perid),

Alloxanthin (Allo), Zeaxanthin (Zea). Allo and Zea are both included in PPC pigments while Fuco, Perid, 19HF and 19BF are PSC pigments.

All pigments except Chla are considered as Diagnostic Pigments (DP). A Diagnostic Pigment Analysis (DPA) can be applied to classify phytoplankton types from HPLC pigment data (Table A1.2) (Vidussi *et al.*, 2001). DPA defines an ensemble of DP for specific phytoplankton groups that can be quantified with respect to the sum of all the relative DP concentrations (i.e. $DP/\Sigma DP$) to estimate the relative abundance of a specific phytoplankton group. This approach has been used in different studies at both global and regional scales (Vidussi *et al.*, 2001; Marty *et al.*, 2002; Uitz, Claustre, Morel and Stanford B Hooker, 2006; Hirata *et al.*, 2011; Sammartino *et al.*, 2015; Di Cicco *et al.*, 2017; Mayot *et al.*, 2017).

Table A1.1: Compilation of HPLC databases, and their contributions in term of the sampling period, zone and number of observations.

Campaign/Dataset	Zone	Period	N Obs
MAREDAT ^a	Global ocean	1997-2008	10340
NOMAD ^b	Global ocean	1997-2003	1080
GeP&Co ^c	Pacific-Atlantic-Indian Oceans	7/2000-9/2002	1205
Polarstern ^d	Atlantic ocean	10/2007-5/2010	396
Labrador ^e	Labrador Sea	2005-2014	253
Tara Oceans Expedition ^f	Global ocean; Med	9/2009-10/2013	410
SeaBASS ^g	Global ocean	1997-2014	1836

^a (Luo *et al.*, 2012), ^b (Werdell and Bailey, 2005), ^c (Dandonneau *et al.*, 2004), ^d (Bracher *et al.*, 2015), ^e (Fragoso *et al.*, 2016), ^f (Pesant *et al.*, 2015), ^g <https://seabass.gsfc.nasa.gov/>

Table A1.2: Phytoplankton groups and their associated major pigments

Phytoplankton Size Class (PSC)	Phytoplankton functional group (PFT)	Pigment
Micro-phytoplankton (20 to 200µm)	Diatoms ^a	Fucoxanthin (Fuco)
	Dinoflagellates ^b	Peridinin (Perid)
Nano-phytoplankton (2 to 20µm)	Prymnesiophytes ^c	19'HexFucoxanthin (19HF)
	Chromophytes ^c	19HF and 19'ButFucoxanthin (19BF)
	Chlorophytes ^c	Chlorophyll-b (Chlb)
	Cryptomonads ^d	Alloxanthin (Allo)
Pico-phytoplankton (<2µm)	Cyanobacteria-Synechococcus ^e	Zeaxanthin (Zea)
	Prochlorococcus ^e	Divynil-Chlorophyll a and b (DVChla, DVChlb)

^a (Jeffrey, 1980), ^b (Jeffrey and Hallegraeff, 1987), ^c (Wright and Jeffrey, 1987), ^d (Gieskes and Kraay, 1983), ^e (Guillard *et al.*, 1985).

2.4. The experimental Database

A total of 15520 HPLC duplicate free samples were compiled in this database, limited to the first optical depth which is about 15-35m (D'Ortenzio *et al.*, 2009). The duplicates were identified by comparing the coordinates (latitude, longitude), the date (Year, month and the day) and the Depth. Vector data with similar or close values within these parameters were averaged. Due to their different optical properties (Mitchell *et al.*, 1991; Fenton *et al.*, 1994; Arrigo *et al.*, 1998), 3761 HPLC samples originating from the Antarctic Ocean were excluded (latitude > 50 °S). Afterwards, satellite match-ups were retrieved by extracting the nearest available pixel to the 11759 remaining HPLC samples coordinates in a 3x3 box, within a time lapse of +1 day. To specialize the applicability of the method to oceanic waters which are characterized by low to moderate Chla levels, in particular in comparison with coastal waters, the 95% percentile of the data was used which corresponds to a threshold at 3 mg.m⁻³ for the Chla values. High values above this percentile were flagged and replaced by a missing value in the remaining data. For consistency, the 95% percentile was then used to flag all high values of each variable. At this stage, 853 measurements with more than 14 missing variables were eliminated from the database. The resulting database contains 10906 sample, from which 7594 samples are collocated with satellite images between 1997 and 2014. The geographic location of this database samples is presented in Figure A1.1.

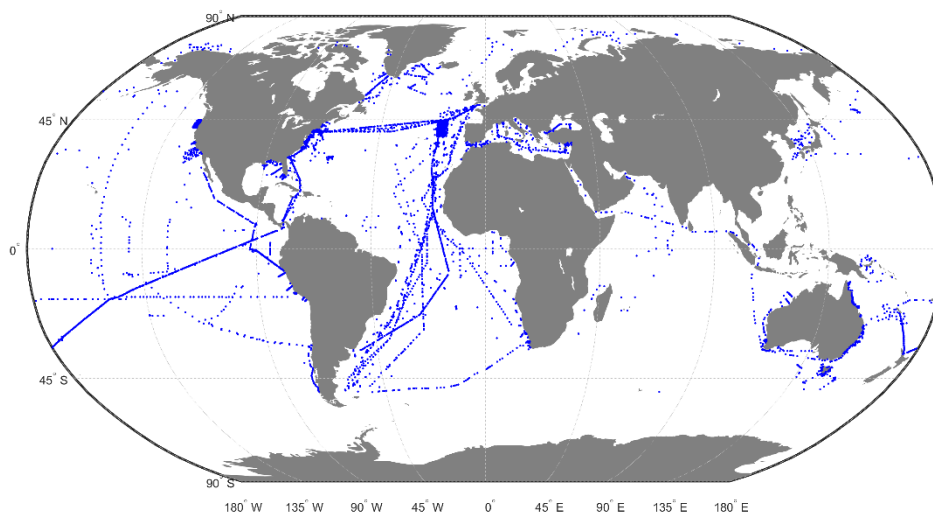


Figure A1.1: Geographic location of the HPLC in-situ measurements

Despite the scattered missing values, the remaining data contain information which may efficiently contribute to the calibration of the SOM and this is due to the learning algorithm dedicated to the SOM method that will take into account the inter-variable relationship (see section 3.1).

Finally, the experimental database ($\mathbf{D}_{\text{pigment}}$), that will serve to train the SOM, has a dimension of 10906x16, where 10906 are the number of samples (individuals), 16 are the number of components (10 HPLC pigments: in situ Chla, DVChla, Chlb, DVChlb, 19HF, 19BF, Fuco, Perid, Allo and Zea, and 6 Satellite derived variables: OC5 Chla, 4 Rrs(λ) and SST).

3. The Proposed Method

3.1. Self-organizing maps (SOM), the general concept

The SOM algorithms (Kohonen, 2013) constitute powerful nonlinear unsupervised classification methods. They are unsupervised neural classifiers, which have been commonly used to solve environmental problems (Cavazos and Cavazos, 1999; Richardson, Risien and Shillington, 2003; Liu and Weisberg, 2005; Liu, Weisberg and Mooers, 2006; Niang *et al.*, 2006; Reusch, Alley and Hewitson, 2007). SOM aims at clustering vectors of a multidimensional database (\mathbf{D}) into classes represented by a fixed network of neurons (the SOM map). The self-organizing maps are defined as an undirected graph, usually a rectangular grid of dimension $p \times q$. This graph structure is used to define a discrete distance (denoted by δ) between the neurons of the map which present the shortest path between two neurons. Moreover, SOM enables the partition of \mathbf{D} in which each cluster is associated with a neuron of the map and is represented by a prototype that is a synthetic multidimensional vector (the referent vector \mathbf{w}). Each vector \mathbf{z}_i of \mathbf{D} will be assigned to the neuron whose referent \mathbf{w} is the closest, in the sense of the Euclidean Norm (EN), and will be called the projection of the vector \mathbf{z}_i on the map. A fundamental property of a SOM is the topological ordering provided at the end of the clustering phase: two close neurons on the map represent data that are close in the data space. Indeed, the neurons are gathered in such a way that two close vectors of \mathbf{D} are projected on two “relatively” close neurons (with respect to δ) on the map. The estimation of the referent vectors \mathbf{w} of a SOM and the topological order is achieved through a minimization process in which the referent vectors \mathbf{w} are estimated from a learning data set (The DPIG data base in the present case). The cost function is of the form:

$$J_{SOM}^T(\chi, W) = \sum_{z_i \in D} \sum_{c \in SOM} K^T(\delta(c, \chi(z_i))) \|z_i - w_c\|^2 \quad (3)$$

where $c \in SOM$ indices the neurons of the SOM map, χ is the allocation function that assigns each element z_i of D to its referent vector $w_{\chi(z_i)}$. $\delta(c, \chi(z_i))$ is the discrete distance on the SOM between a neuron c and the neuron allocated to observation z_i and K^T a kernel function parameterized by T (where T stands for “temperature” in the scientific literature dedicated to SOM) that weights the discrete distance on the map and decreases during the minimization process.

This cost function takes into account the proper inertia of the partition of the data set D and ensures that its topology is preserved.

SOMs have frequently been used in the context of completing missing data (Jouini *et al.*, 2013), so the projected vectors z_i may have missing components. Under these conditions, the distance between a vector $z_i \in D$ and the referent vectors w of the map is the Euclidean distance that considers only the existing components (the Truncated Distance or TD hereinafter). The use of the TD allows to take into account the information embedded in the incomplete data described in the section 2.

3.2. Construction of the SOM

Phase 1: Training phase:

In the present study, the SOM map is constituted by a two-dimensional rectangular grid (200x100 =20 000 referents) trained using D_{pigment} , minimizing the $J_{SOM}^T(\chi, W)$ cost function. In order to equitably distribute the weights through the training procedure, the 16 parameters were normalized with their variances. So, each parameter contributes to build the SOM. Using a number of neurons larger than that of the training dataset allowed to refine the discretization of w and therefore to obtain a more accurate pigment estimation. Several experiments were made to find the ideal SOM size and have shown a significant increase of the general performance of the method at estimating pigment concentrations when the number of neurons increases to a certain extent (5.000, 10.000 and 20.000 neurons).

In the case of 20 000 neurons map, at least half of the neurons of the SOM has captured a sample of the database, which permitted to define a referent vector w for these neurons. The

second half of the neurons defined their \mathbf{w} through the topological ordering using the equation 3. In other terms, the discrete distance $\delta(c, \chi(z_i))$ between the neighboring neurons and the kernel K^T was used to determine the referent vector \mathbf{w} of each neuron that has not captured any data (Sarzeaud and Stephan, 2000). This proves the interest of the topological order provided by the SOMs maps which is used to accurately interpolate referent vectors for neurons which have not captured any data.

At the end of the training phase, each neuron of the SOM (denoted SOM-Pigments) was therefore associated with a referent vector \mathbf{w}_k constituted of 16 components, with $k \in \{1...20000\}$.

Phase 2: Pigment retrieval:

In the second phase which is an operating phase, we estimate the pigment concentrations using the different satellite images; The 6 ocean satellite observations (4 Rrs(λ), satellite Chla and SST) of a pixel P_j are projected onto the SOM. Doing so, the projected parameters are normalized with the corresponding variances of $\mathbf{D}_{\text{pigment}}$ to maintain an equal weight among the parameters and are assigned with the closest best matching neuron using the TD defined in 3.1. At the end of the assignment phase, each pixel is associated with a referent vector \mathbf{w}_k corresponding to the best matching neuron, which includes the 10 pigment concentrations.

For this phase, level 3 mapped 4km daily images of SST, Chla and Rrs(λ) at 4 wavelengths (412, 443, 490 and 555 nm) were used to estimate the pigment concentrations.

Phase 3: Cross-validation:

In order to evaluate the performance of the SOM on a global scale, the statistical parameters R^2 and RMSE have been calculated via the following cross-validation procedure implemented for each estimated pigment:

- Step 1: The initial data set was randomly segmented into 20 different partitions: 95% of the data served as a training set and 5% as a test set.
- Step 2: At each iteration, the SOM was trained with the training set.
- Step 3: We applied the pigment retrieval procedure described in phase 2.

- Step 4: The pigments concentrations estimated from the retrieval procedure were compared to the observed test set.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Obs_i - Est_i)^2}{n}} \quad (4)$$

Phase 4: Uncertainty calculation and quality control:

In order to compute the uncertainty on the estimated pigments, an ascending hierarchical clustering was applied to the SOM in order to cluster similar neighboring neurons into 500 Big-clusters (approximately 40 neurons per cluster). Then, for each Big-cluster j ($j=1...500$), a standard deviation vector (**Std_j**) was calculated, whose components x_i ($i=1...16$) are the 16 parameters of **D_{pigment}**. The standard deviation for component x_i belonging to cluster j is denoted **Std_j(x_i)** and is estimated from the values of the neurons clustered within the Big-cluster j . As a result, each neuron k belonging to the Big-cluster j will be associated with its **Std_j** and will be denoted **Std_k**

$$Std_k(x_i) = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x}_i)^2}{n}} \quad (5)$$

Where n is the number of neurons k in the Big-cluster j , x_i is a component ($i=1...16$) of a neuron k , \bar{x}_i is the mean of the component x_i computed from the neurons in Big-Cluster j :

$$\bar{x}_i = \frac{1}{n} \sum_{i=1}^n x_i \quad (6)$$

Moreover, in the operational phase, a quality control was applied based on the difference between the $Rrs(\lambda)$ spectrum projected on the SOM map and its corresponding selected neuron. A mask (M) is generated for each pixel p using the four $Rrs(\lambda)$ as follows:

$$\Delta Rrs(\lambda)_p = \frac{Rrs(\lambda)_p - Rrs(\lambda)_k}{Std_k(Rrs(\lambda))} \quad (7)$$

Where λ represents the wavelengths 412, 443, 490 and 555 nm, $Rrs(\lambda)$ is the value of $Rrs(\lambda)$ at wavelength λ for pixel p , $Rrs(\lambda)_k$ is the value of $Rrs(\lambda)$ at wavelength λ of the selected neuron k , and $Std_k(Rrs(\lambda))$ is the standard deviation for $Rrs(\lambda)$ as described above. $\Delta Rrs(\lambda)_p$ is a

function of $\text{Std}_k(\text{Rrs}(\lambda))$, it is the ratio of the difference between the $\text{Rrs}(\lambda)$ of p and its associated neuron k and the $\text{Std}_k(\text{Rrs}(\lambda))$.

The mean of the 4 $\Delta\text{Rrs}(\lambda)_p$ is attributed to each pixel of the resulting image;

$$\overline{\Delta\text{Rrs}(\lambda)}_p = \frac{\sum_{\lambda=412,443,490,555}^4 \Delta\text{Rrs}(\lambda)_p}{4} \quad (8)$$

The quality control Mask (denoted M_p) was defined using a threshold at $\overline{\Delta\text{Rrs}(\lambda)}_p = \pm 2 \text{ STD}$. therefore, values of $\overline{\Delta\text{Rrs}(\lambda)}_p$ above this threshold are rejected. This allows to filter the pixels with abnormal $\text{Rrs}(\lambda)$ spectrum that were not observed within the D_{pigment} used to train the SOM.

4. Results

4.1. Analysis of the SOM Organization and topology

Once the training phase of the SOM is done, the relationship between the in-situ pigments and those estimated from the satellite observations was assessed. A correlation analysis was performed between the parameters within the initial database and within the SOM. As a result, the correlations are efficiently preserved (Table A1.3). Therefore, the SOM is representative of the initial database D_{pigment} , reproducing accurately the relationship between the parameters.

Table A1.3: Comparison between the correlation matrix for the 16 variables of the initial database (white), and the correlation matrix of the 16 variables within the SOM after the training phase (gray). This comparison clearly indicates how SOM preserved the correlations between variables.

		SOM															
In-Situ Data	Var	Chla	DVCh la	Chlb	DVCh lb	19HF	19BF	Fuco	Perid	Allo	Zea	Chla OC5	Rrs(412)	Rrs(443)	Rrs(490)	Rrs(555)	SST
	Chla		-0.14	0.52	-0.09	0.51	0.36	0.79	0.53	0.52	0.00	0.70	-0.47	-0.48	-0.30	0.33	-0.32
	DVChla	-0.18		-0.08	0.18	-0.09	0.01	-0.15	-0.09	-0.16	0.38	-0.17	0.12	0.09	0.07	-0.07	0.20
	Chlb	0.39	-0.06		0.02	0.39	0.38	0.37	0.34	0.39	0.02	0.41	-0.44	-0.44	-0.25	0.23	-0.31
	DVChlb	-0.08	0.23	0.09		-0.07	0.01	-0.09	-0.03	-0.07	0.13	-0.10	-0.01	-0.02	-0.03	-0.08	0.12
	19HF	0.55	-0.15	0.35	-0.04		0.78	0.37	0.40	0.33	0.01	0.31	-0.42	-0.45	-0.37	0.07	-0.36
	19BF	0.38	-0.05	0.37	0.02	0.77		0.22	0.33	0.27	-0.04	0.14	-0.37	-0.40	-0.35	-0.02	-0.39
	Fuco	0.81	-0.17	0.36	-0.07	0.41	0.29		0.52	0.47	-0.08	0.69	-0.44	-0.46	-0.31	0.28	-0.29
	Perid	0.54	-0.10	0.31	-0.02	0.45	0.38	0.48		0.35	-0.01	0.44	-0.38	-0.40	-0.31	0.14	-0.18
	Allo	0.52	-0.19	0.34	-0.04	0.40	0.24	0.44	0.31		-0.14	0.40	-0.36	-0.36	-0.24	0.16	-0.25
	Zea	-0.02	0.39	-0.03	0.15	0.06	0.02	-0.04	0.01	-0.13		-0.01	0.07	0.06	0.14	0.14	0.32
	Chla OC5	0.74	-0.19	0.34	-0.09	0.37	0.20	0.67	0.42	0.39	-0.01		-0.56	-0.55	-0.27	0.50	-0.25
	Rrs(412)	-0.48	0.10	-0.41	-0.01	-0.48	-0.43	-0.45	-0.38	-0.36	-0.01	-0.58		0.97	0.68	-0.14	0.44
	Rrs(443)	-0.49	0.09	-0.40	-0.02	-0.50	-0.44	-0.46	-0.40	-0.35	-0.02	-0.57	0.97		0.81	0.01	0.45
	Rrs(490)	-0.30	0.08	-0.19	0.00	-0.38	-0.35	-0.29	-0.28	-0.23	0.08	-0.30	0.68	0.81		0.51	0.39
	Rrs(555)	0.30	-0.07	0.20	-0.04	0.09	0.02	0.25	0.14	0.15	0.13	0.43	-0.12	0.03	0.51		0.02
SST	-0.34	0.21	-0.26	0.10	-0.37	-0.42	-0.31	-0.21	-0.23	0.20	-0.28	0.42	0.43	0.34	-0.05		

Besides, the 16 components affected to the neurons of the SOM seemed to be well organized; Diagonal gradients are found for most of the parameters (Figure A1.2). Such results denote a coherent relationship between the variables. The organization of the pigments on the SOM with respect to the satellite data helps understanding the link between the phytoplankton pigments and the satellite signal. The in-situ and satellite Chla are closely related which may explain the satisfying correlation between satellite and in-situ Chla data ($r=0.70$, $pVal < 0.001$). The Fuco follows a diagonal gradient on the SOM (top left, bottom right) and thus exhibits an organization close to that of both in-situ and satellite Chla. As a consequence, there is a strong relationship between these two pigments (Chla in-situ vs. Fuco: $r=0.80$, Chla OC5 vs. Fuco: $r=0.67$, $pVal < 0.001$). The Chlb and Perid mainly exhibit a vertical gradient (top to bottom), while the Allo is organized with high values simultaneously on the top and left sides of the SOM map. 19HF and 19BF are both covariant ($r=0.77$, $pVal < 0.001$), and present a diagonal gradient (top right, bottom left) whilst both Chla and Fuco present a gradient following the other diagonal. DVChla and Zea present a diagonal gradient (bottom right, top left) and DVChlb is maximum in the middle of the right side of the map.

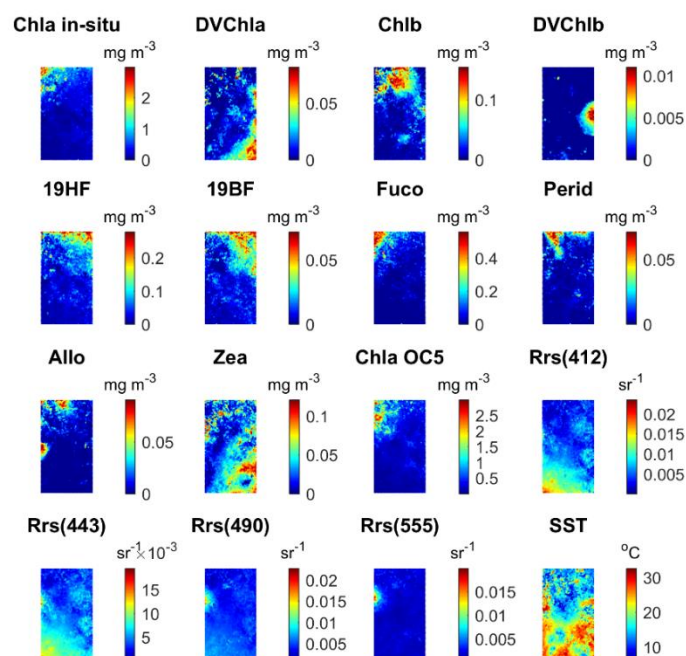


Figure A1.2: Organization of the 16 variables on the SOM map after the training phase. Each map represents the values recorded by the neurons of the SOM for the 16 variables. The topological organization of these neurons reflects the inter-variable relationship.

Most of the diagnostic pigments follow a negative gradient of SST ($r < -0.21$, $pVal < 0.001$), while Zea, DVChlb and DVChla co-vary with SST (high values of Zea, DVChlb and DVChla coincide with high SST values) ($r > 0.1$, $pVal < 0.001$). Noting that Zea is an indicator of Cyanobacteria communities and that both DVChlb and DVChla are diagnostic pigments of Prochlorococcus (Cyanobacteria), this link suggests a correlation between SST and Cyanobacteria spatio-temporal patterns. In a general view, the SOM topology seems to follow a certain regionalization; Chla, Chlb, 19HF, 19BF, Fuco, Perid and Allo pigments components show high concentrations at the top of the SOM, whereas low concentrations are found at the bottom. In contrast, DVChla, DVChlb and Zea display an opposite structuring. This organization promotes a subdivision into oligotrophic waters (bottom) where cyanobacteria dominate and more productive waters (top) where the other species are abundant.

4.2. Cross validation results

Table A1.4 shows the cross-validation results of the 10 pigments for each experiment with an increasing number of neurons, and Figure A1.3 presents the scatterplots that compare the observed pigment concentrations versus the estimated values for the test sets using 20 000 neurons. The RMSE and the R^2 coefficients between the estimated and the observed pigments both improve when a higher number of neurons is specified; For example, the accuracy of the Fuco prediction shifts from an R^2 of 0.5 to 0.78 when the SOM is increased from 5000 to 20000 neurons. Overall, when 20000 neurons are prescribed, the R^2 coefficients for the different pigments range between 0.75 and 0.89 with an average RMSE of $0.012 \text{ mg} \cdot \text{m}^{-3}$. Furthermore, the inspection of the scatterplots (Figure A1.3) shows that there is no bias in the estimation procedure.

Table A1.4: Statistical results of the cross-validation result for the 10 pigments. Non-log transformed data were evaluated to calculate the RMSE.

	N Neur=20000		N Neur=10000		N Neur=5000		pVal	N Obs
	R^2	RMSE [mg m^{-3}]	R^2	RMSE [mg m^{-3}]	R^2	RMSE [mg m^{-3}]		
Chla_{SOM}	0.84	0.22	0.79	0.23	0.55	0.24	0.001	4179
DVChla	0.77	0.01	0.56	0.02	0.48	0.02	0.001	2502
Chlb	0.85	0.01	0.69	0.02	0.64	0.02	0.001	2900
DVChlb	0.89	0.001	0.49	0.002	0.43	0.01	0.001	383
19HF	0.75	0.02	0.73	0.03	0.58	0.03	0.001	4284

19BF	0.79	0.01	0.68	0.01	0.61	0.01	0.001	4192
Fuco	0.87	0.02	0.76	0.02	0.50	0.02	0.001	4382
Perid	0.80	0.01	0.35	0.02	0.54	0.01	0.001	3113
Allo	0.76	0.01	0.50	0.02	0.41	0.01	0.001	2215
Zea	0.79	0.01	0.74	0.01	0.54	0.02	0.001	4262

The R^2 coefficient between the Chla estimated by the SOM ($Chl_{a_{SOM}}$) and the in-situ Chla is equal to 0.84 with an RMSE of $0.24 \text{ mg} \cdot \text{m}^{-3}$. When predicting $Chl_{a_{SOM}}$, the SOM provides a tool to validate the spatial reconstruction by comparing the output with the satellite $Chl_{a_{OC5}}$ images.

This comparison performed between daily images shows a good agreement between the two Chla, scoring an R^2 of 0.85, and an RMSE of $0.13 \text{ mg} \cdot \text{m}^{-3}$ (Figure A1.4). Besides, a slight over-estimation of $0.01 \text{ mg} \cdot \text{m}^{-3}$ by the $Chl_{a_{SOM}}$ compared to $Chl_{a_{OC5}}$ was highlighted for Chla values less than $1 \text{ mg} \cdot \text{m}^{-3}$. A saturation is observed at values near $3 \text{ mg} \cdot \text{m}^{-3}$ and it is mainly due to the limitation of the Chla values of the learning database, which were used to train the SOM. On the spatial scale, the Chla SOM reproduce the patterns initially observed in the original product, which proves that the reconstruction is efficient.

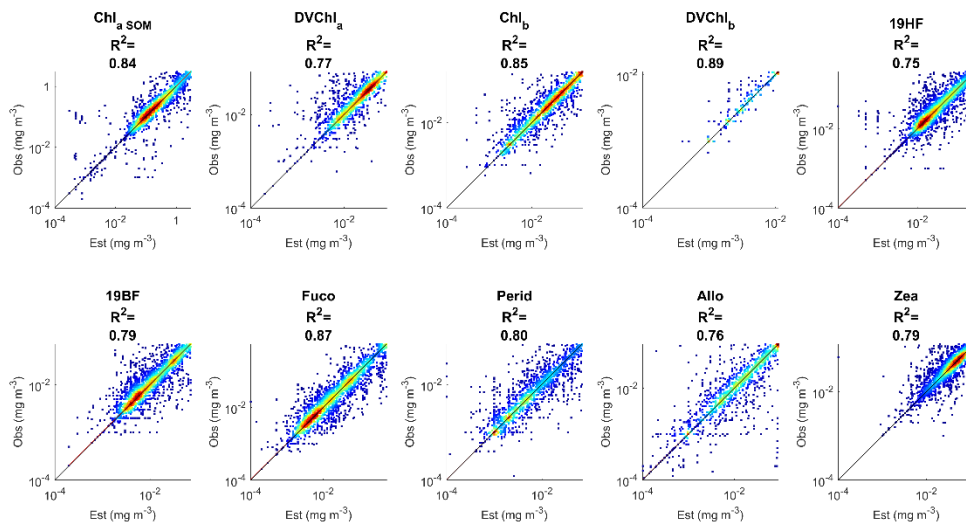


Figure A1.3: Scatter plots illustrating the cross-validation cumulative results for the 10 pigments, estimated vs. observed. The R^2 above each scatter plot represents the regression coefficient.

4.3. Estimation of pigment concentrations in the global ocean

The method described in section 3 has been used to generate daily images of pigments concentrations at global scale using satellite reflectance, Chla and SST products presented in section 2 (Figures A1.5, A1.6 and A1.7).

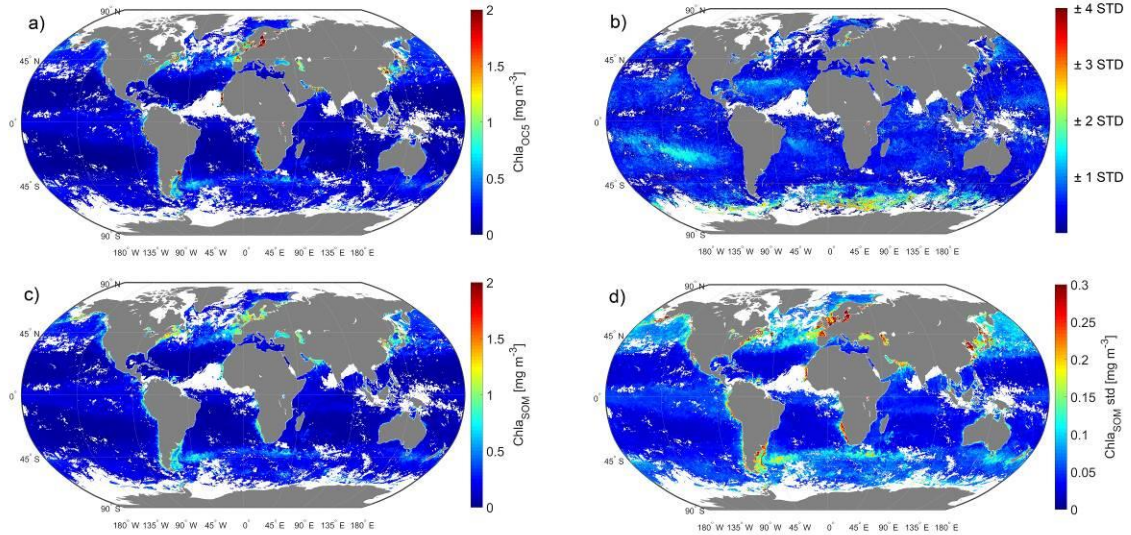


Figure A1.5: Eight day composite (2-9 April 2017) of a) $Chla_{OC5}$, b) $\Delta Rrs(\lambda)_p$, c) $Chla_{SOM}$, d) $Chla_{SOM}$ uncertainties.

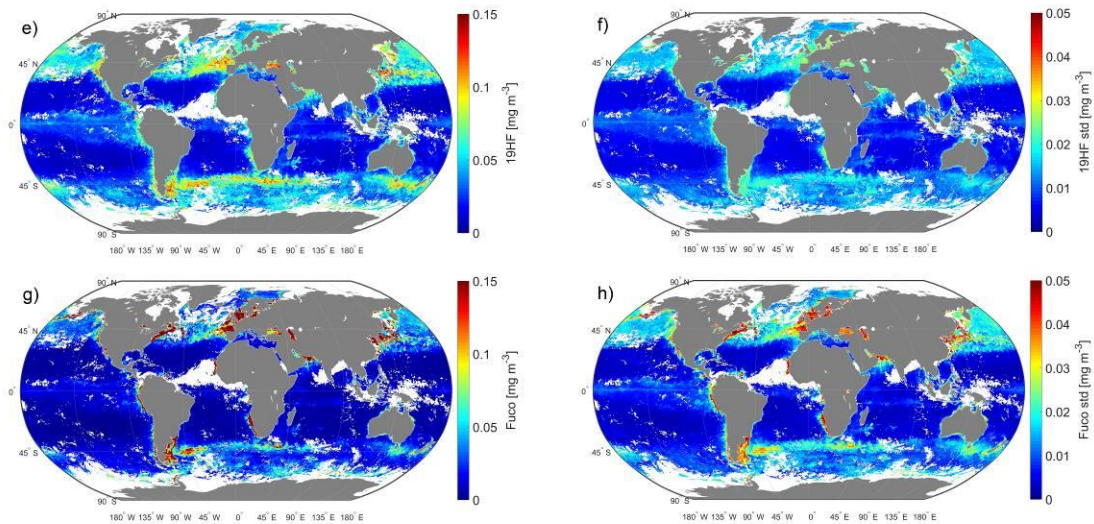


Figure A1.6. Eight day composite (2-9 April 2017) of e) 19HF, f) 19HF uncertainties, g) Fuco, h) Fuco uncertainties.

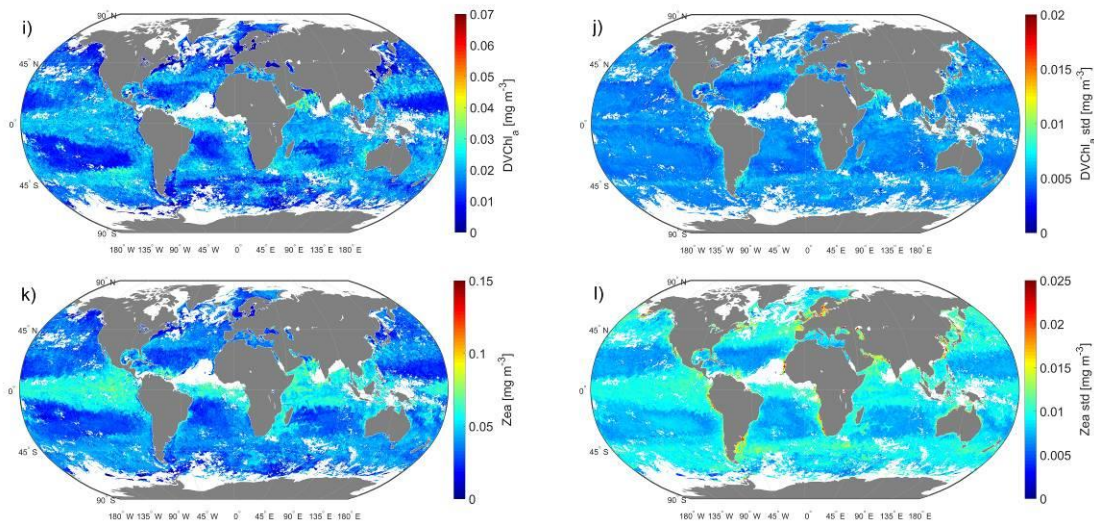


Figure A1.7: Eight day composite (2-9 April 2017) of i) DVChla, j) DVChla uncertainties, k) Zea, l) Zea uncertainties.

Furthermore, a quality control mask was determined from the mean difference $\overline{\Delta Rrs(\lambda)}_p$ between the satellite $Rrs(\lambda)$ spectrum projected on the SOM map and the $Rrs(\lambda)$ spectrum of the best matching neuron (see section 3, phase 4 for a detailed description) (Figure A1.5b). To quantify the quality of the pigments concentrations estimated by our method, a 10-years climatology (2006-2016) of $\overline{\Delta Rrs(\lambda)}_p$ has been generated and is displayed in Figure A1.8.

Values of $\overline{\Delta Rrs(\lambda)}_p$ remain relatively low over most of the ocean both in winter and in summer. However, the Southern Ocean exhibits high values (above ± 2 STD) in winter, which may even exceed ± 4 STD in some areas. In addition to that, the southern subtropical gyres of the Pacific Ocean are also characterized by relatively high uncertainties that do not yet exceed drastically 2 STDs. A 2 years (2012-2013) daily analysis was performed in order to evaluate the frequency of pixels that are characterized by high values of $\overline{\Delta Rrs(\lambda)}_p$ during that period. The near shore pixels have a 60% frequency to be flagged mainly due to the fact that most case 2 water data were excluded from the initial database. In the open ocean, the frequency drops to less than 20% except in the Southern Ocean where this frequency is higher, reaching 45% in the circumpolar current. This quite large uncertainty in that region is explained by its specific optical characteristics as well as by the exclusion of the data collected in the Southern Ocean from the training database and the very limited availability of satellite observations

(less than 50 observation per pixel through a 2 years period). Consequently, one may call into question the reliability of the SOM-Pigments to predict accurately the pigment concentrations in the Southern Ocean during the winter season. Yet, the 2 years frequency analysis shows that some observations in the Southern Ocean fall under the threshold. They mainly correspond to situations that do exist in other regions of the ocean and thus, that have been accounted for in the training phase. Therefore, in this case, the estimated pigments should be reliable. In the following, pixels with a ± 2 STD deviation are flagged to insure a relevant representation while tracking stable patterns of 6 major pigments concentration (Chla, DVChla, Chlb, 19HF, Fuco and Zea), in winter (Dec-Feb) and summer (July-Sept). Seasonal climatology composites were calculated and are represented in Figures A1.9 and A1.10 for the period ranging between 2006 and 2016.

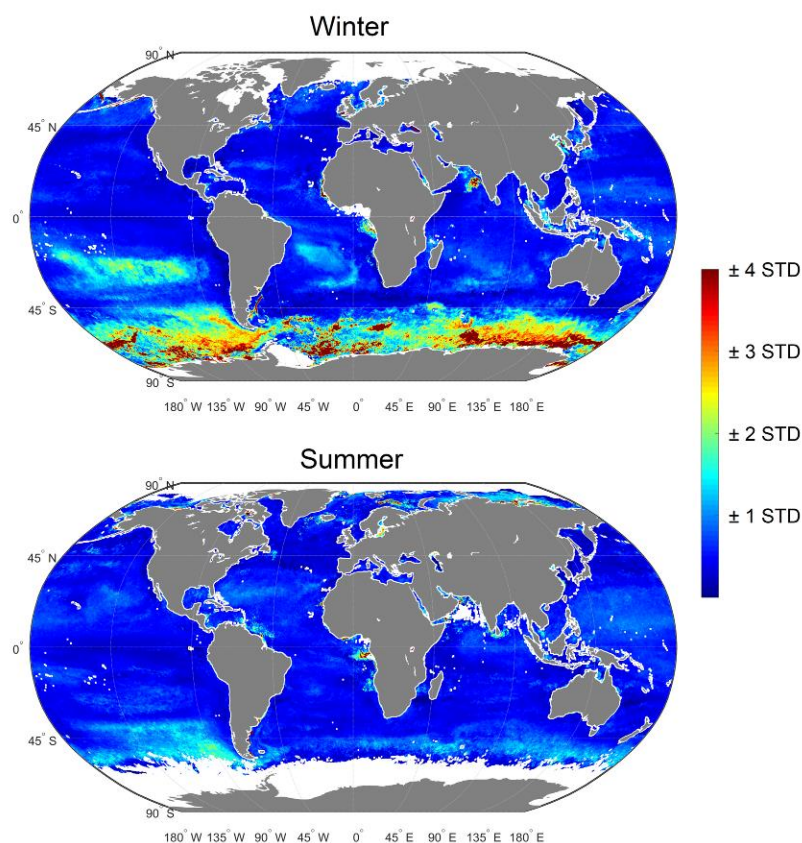


Figure A1.8: Seasonal climatology of the quality control mask as a function of STD.

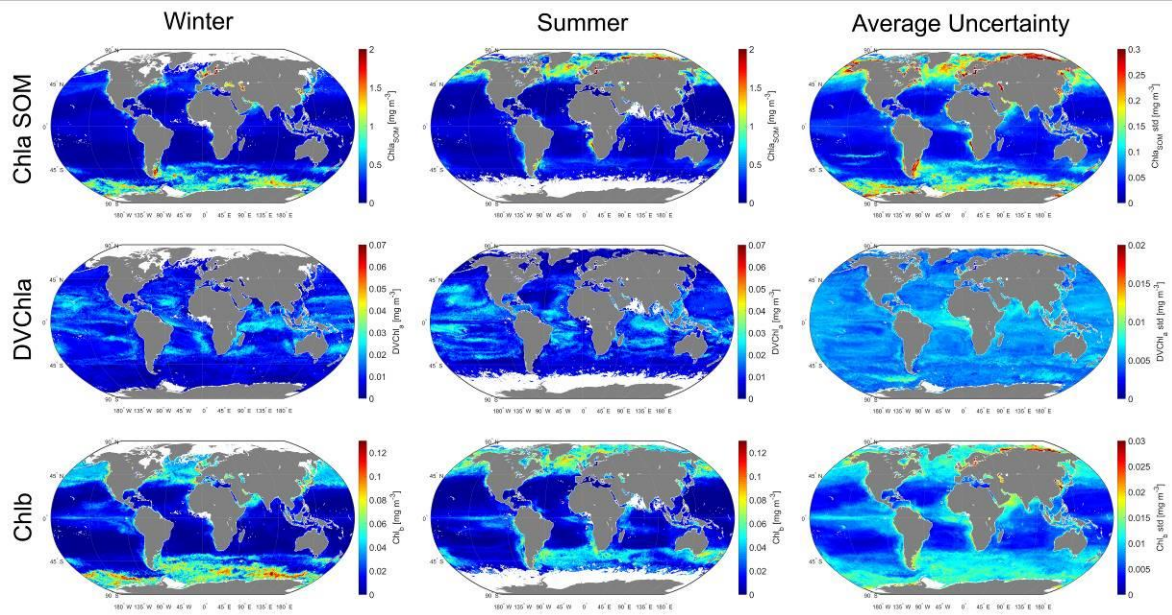


Figure A1.9: Seasonal climatology (Winter-Summer) generated between 2006-2016 of the estimated pigment concentrations via SOM and their corresponding uncertainties for Chla SOM (top line), DVChla (middle line), Chlb (bottom line).

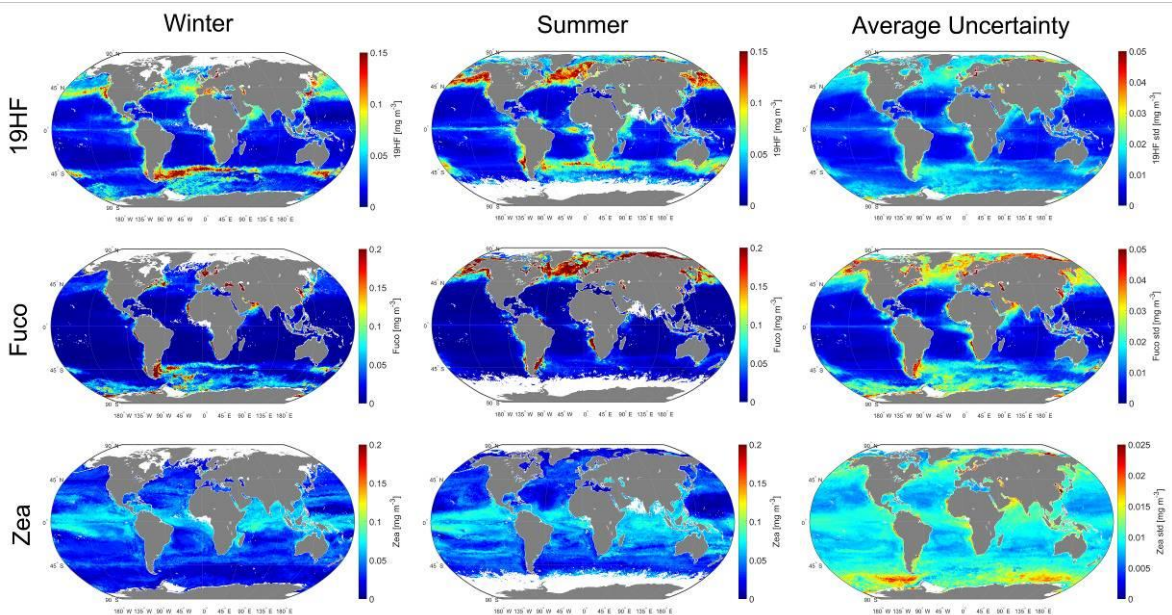


Figure A1.10: Seasonal climatology (Winter-Summer) generated between 2006-2016 of the estimated pigment concentrations via SOM and their corresponding uncertainties for 19HF (top line), Fuco (middle line), Zea (bottom line).

The results show specific patterns for every pigment. We note that the dynamics of Fuco is similar to that of Chla: High concentrations are observed in the Antarctic Sea in the winter season, and a pronounced maximum of $\text{Fuco} > 0.2 \pm 0.05 \text{ mg}\cdot\text{m}^{-3}$ is observed in the Northern

part of the Atlantic and of the Pacific in summer. High concentrations of Fuco $>0.2\pm0.05$ mg.m⁻³ are also found in Eastern Boundary Upwelling Systems such as off Peru, California and in the Benguela upwelling. This close relationship between Fuco and Chla was expected from the analysis of the SOM topology (see Section 4.1).

19HF (Figure A1.10) presents a spatio-temporal variability with stable patterns above 45°N (winter) and 45°S (summer) at the level of the subpolar/temperate interface. High concentrations of $>0.15\pm0.05$ mg.m⁻³ of 19HF and $>0.05\pm0.001$ mg.m⁻³ of 19BF are recorded in winter above 45°S associated with the circumpolar current and in summer around 45°N.

Chlb shows a pronounced variability in the subpolar regions, north of 45°N and south of 45°S latitudes. In winter, the Antarctic Sea shows higher concentrations of Chlb; while in summer, higher concentrations are observed in the Northern part of the Atlantic and the Pacific, reaching values above 0.12 ± 0.03 mg.m⁻³.

Zea and DVChla pigments are restricted to the subtropical and tropical regions and both present minimal concentrations in the high latitudes throughout the year. Zea shows relatively constant concentrations around 0.1 ± 0.01 mg.m⁻³ between 30°N and 30°S. The largest concentrations are reached near the equatorial divergences. In contrast, DVChla is characterized by a more patchy distribution and tends to be maximum at about 0.04 ± 0.01 mg.m⁻³ along the boundaries of the subtropical gyres.

To further assess the seasonal patterns of these pigments, the pigment with the highest concentration per pixel is represented in Figure A1.11. This simple analysis is designed to estimate the spatio-temporal distribution of the dominant pigments at the global scale. However, this simple diagnostic should be cautiously interpreted since the pigment to chlorophyll or pigment to carbon ratios are highly variable. As a consequence, a dominant pigment does not necessarily mean that the associated species dominate the phytoplankton community.

First, the 19HF pigment seems to dominate in the northern area of the Pacific and the Atlantic, and around the Circumpolar Current during both the winter and summer seasons. Meanwhile, Fuco is the most abundant pigment in coastal areas and in semi enclosed and enclosed basins. During the winter season, this pigment dominates along the Antarctic coast

and downstream of the islands in the Southern Ocean. In summer, Fuco dominates in the northern Pacific and Atlantic coasts and around the arctic. Chl_b maxima are observed mainly in the Southern Ocean in winter. Both DVChl_a and Zea are restricted to the tropical and subtropical regions and are almost absent at high latitudes. DVChl_a dominates in the subtropical gyres while Zea is the most abundant pigment over large areas of the tropical domain, with the exception of the upwelling and coastal regions. Both pigments are characteristic of cyanobacteria and their distributions are consistent with observations (Flombaum *et al.*, 2013).

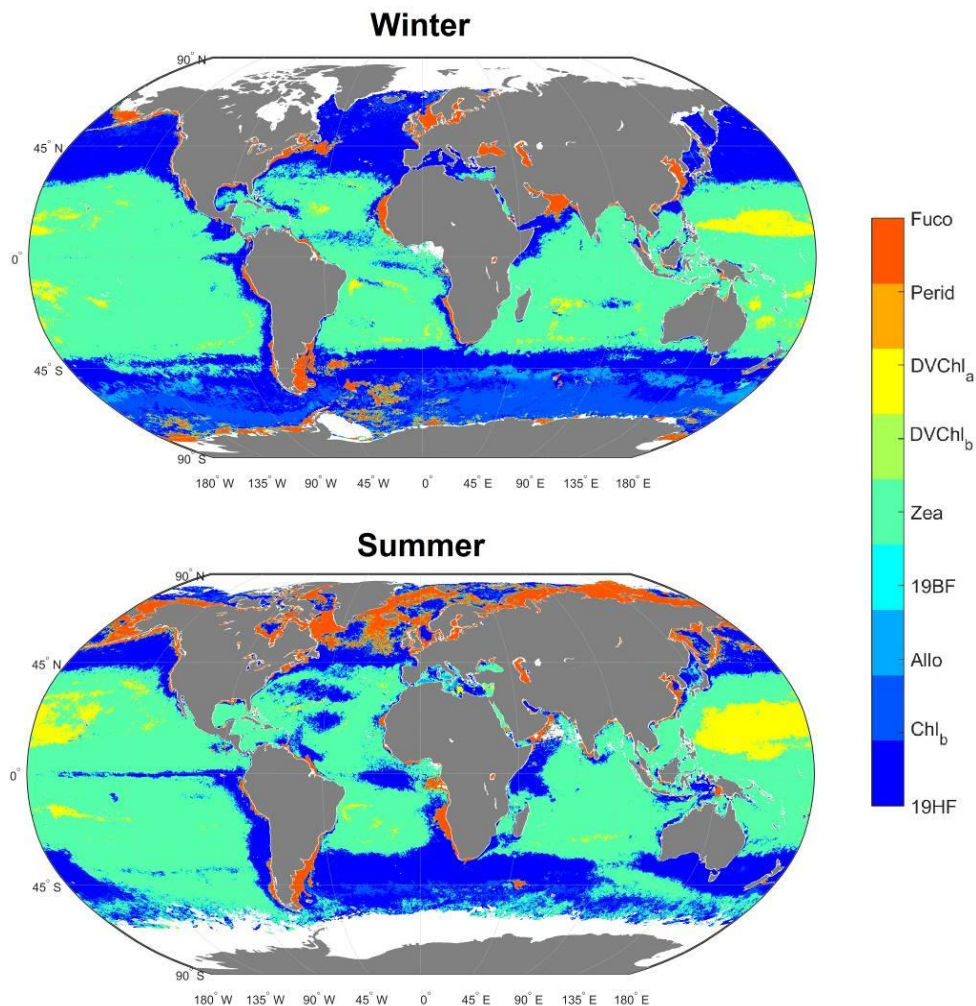


Figure A1.11: Global projection of the pigment with the maximum concentration at each pixel. The seasonal climatology were used to generate these MaxDP images.

5. Discussion

The originality of our method is to model the relationship between satellite observations and phytoplankton pigments by partitioning a large database in a very large number of small clusters using the SOM. This efficient neural network clustering method amounts to model the multi-dimensional relationship between pigments and satellite observations by a piecewise continuous function. The clustering allows us to easily take into account the multifactorial aspect of the relationship and the different orders of magnitude of the parameters.

5.1. Comparison to other approaches deriving pigment concentrations

Several attempts have been done to derive pigment concentrations from ocean color observations; Pan et al.(2010) developed $Rrs(\lambda)$ band-ratio algorithms for retrieving pigment concentrations. These algorithms are represented by third-order polynomial functions using $Rrs(\lambda)$ band ratio of either 490/550 nm or 490/670 nm (for SeaWiFS; for MODIS changed accordingly to MODIS bands 488 and 547 nm). They calibrated these functions by using satellite ocean color observations collocated with in-situ pigments measurements. Validation of their results with collocated satellite (SeaWiFS and MODIS) reflectance data and pigment concentrations showed very accurate predictions for several pigments (Chla, Chlb, Perid, Fuco, Allo and Zea). The RMSE ranged from 0.23 to 0.29, and the R^2 from 0.65 to 0.90. This method was modified for the northern South China Sea using globally derived relationships and locally identified links between pigment concentrations and SST (Pan *et al.*, 2013) . They achieved an accuracy similar to Pan et al. (2010). Compared to our SOM results, the quality of these estimations is similar. However, based on a larger data set, our SOM offers a robust tool to estimate pigment concentrations at the global scale.

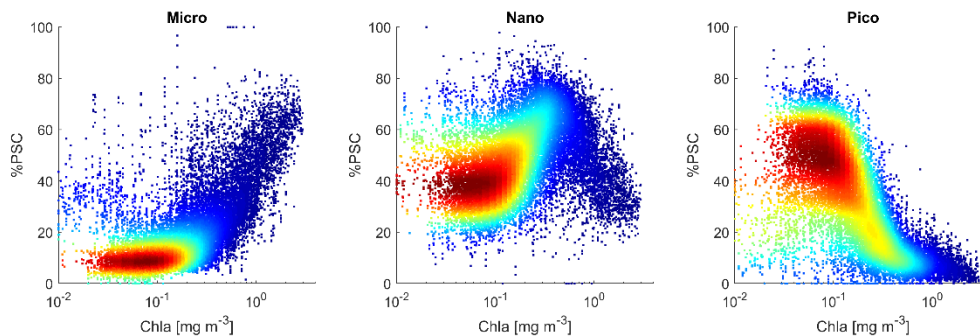
Chase et al. (2013) derived concentrations of different chlorophyll types and several accessory pigments classified in two categories: PSC and PPC s from a global data set of in situ hyper-spectral particulate absorption measurements. This work was followed by another approach (Chase *et al.*, 2017) in which they combined water leaving reflectance measurements and absorption signal to derive pigment concentrations. In both studies, they modeled the pigment absorptions by projecting the spectral signal on twelve Gaussian functions. In (Chase

et al., 2017), TChlb was estimated with an R^2 of 0.51 and PPC with an R^2 of 0.70 compared to HPLC measurements. Our SOM showed an improved retrieval of Chlb, DVChlb, Zea and Allo with an R^2 of 0.85, 0.89, 0.79 and 0.76 respectively noting that each pigment is estimated separately, not as a sum of PPC or PSC. This further indicates the robustness of our approach, characterizing the relationship between pigments and AOP (reflectance), as opposed to the IOPs used in their study. The estimation of IOP from AOP is based on an inversion model which introduces additional uncertainties.

(Bracher *et al.*, 2015) developed a method to assess pigment concentrations from continuous optical measurements. The method applied an empirical orthogonal function (EOF) analysis to remote-sensing reflectance data derived from ship-based hyperspectral underwater radiometry combined with multispectral satellite data (using the Medium Resolution Imaging Spectrometer – MERIS – Polymer product) measured in the Atlantic Ocean. Their results show a satisfactory prediction for several pigment concentrations from satellite data, with a R^2 of 0.25 for DVChla, 0.74 for 19BF, 0.68 for 19HF, 0.71 for Fuco and 0.40 for Zea. Our prediction for these pigments is more accurate recording a R^2 of 0.77, 0.79, 0.75, 0.87 and 0.79 respectively.

Hirata *et al.*, (2011) present a set of equations describing the phytoplankton size structures from Chla abundance in order to highlight the inter-pigment relationships; these authors used a global HPLC database of in-situ secondary phytoplankton pigment concentrations to derive non-linear relationships between phytoplankton size classes and Chla. These equations were also applied within the neurons of the SOM, and the results are shown in function of Chla in Figure A1.12. The underlying relationship between Micro-, Nano-, Pico-phytoplankton and Chla derived from the SOM fits perfectly this described and analyzed by Hirata *et al.*, 2011; For Micro-phytoplankton, the fractional contribution to Chla monotonically increases with increasing Chla, whereas for Pico-phytoplankton, this contribution monotonically decreases with increasing Chla while showing large variations. The fractional contribution of Nano-phytoplankton does not vary monotonically with Chla as found in Micro- and Pico-phytoplankton. Rather the percentage of Nano-phytoplankton increases as Chla increases up to approximately 0.3 mg m^{-3} but decreases as Chla further increases, resulting in a broad maximum between approximately $0.2\text{--}0.6 \text{ mg m}^{-3}$.

The comparison with other methods that have been proposed to retrieve the pigment concentrations from ocean color observations shows that the SOM method was able to model the non-linear relationship between pigment concentrations and satellite derived reflectance, SST and Chla data, and gave robust results. Meanwhile, the SOM was also able to reproduce the underlying inter-pigment relationships efficiently, while being consistent with the description of Hirata *et al.*, (2011) in their analysis performed on in-situ data.



*Figure A1.12: Inter-pigment relationships represented as percentage of phytoplankton size classes (Micro, Nano and Pico-phytoplankton) in function of Chla, calculated within the neurons of the SOM-Pigments. ($\%Micro=1.41(Fuco+Perid)/\Sigma DP$, $\%Nano=(Xn*1.27Hex+1.01Chlb+0.35But+0.6Allo)/\Sigma DP$, $\%Pico=1-(Micro+Nano)$; with $\Sigma DP=1.41Fuco+1.41Perid+1.27Hex+1.01Chlb+0.35But+0.6Allo+0.86Zea$ and Xn indicates a proportion of nanoplankton contribution in Hex (Uitz *et al.*, 2006, Hirata *et al.*, 2011)).*

5.2. Uncertainties and quality control

Besides the synoptic estimation of phytoplankton pigments, the SOM-Pigments allowed the quantification of uncertainties and to ensure a good quality control of the output product. The climatology revealed that most pigments have higher uncertainties with higher concentrations. The Fuco, 19HF and Chlb uncertainties are important in eutrophic regions and near coastal zones. In contrast, DVChla and Zea were both characterized by higher uncertainties in oligotrophic areas and this is mainly due to the non-linearity of the relationship between the pigments, the Rrs spectrum and the phytoplankton community indicated behind.

Furthermore, upon the calculation of the quality control $\overline{\Delta Rrs(\lambda)}_p$, patterns of high STD are recorded in the Southern Ocean during the blooms in winter, which indicates that predictions are highly uncertain in that region. These high STDs can be linked to different bio-optical properties which have been suggested to occur in the Southern Ocean compared to other oceanic regions (Mitchell *et al.*, 1991; Fenton *et al.*, 1994; Arrigo *et al.*, 1998). This led us to eliminate the Austral Ocean's data before the training phase. Furthermore, satellite observations are very limited in this region due to the sea ice coverage especially during the winter season. Yet, once the abnormal data are flagged, The SOM procedure produces a coherent reconstruction of the pigment variability patterns, which are mainly modulated by a global relationship between pigment composition and the satellite signal, assessed by the SOM-Pigments.

5.3. Spatio-temporal variability of phytoplankton pigments

The phytoplankton communities are unequally distributed throughout the ocean in response to several physical and biochemical factors. The latter stimulate some phytoplankton groups to prosper over others. Thus, the prediction of phytoplankton pigments via SOM-Pigments clearly represents how phytoplankton communities are clustered at the global scale. The objective of this section is not to discuss in details the retrieved phytoplankton groups' distribution but to illustrate some major features.

In upwelling and deep mixing regions, Fuco concentrations are elevated, followed by 19HF. In these regions characterized by a high input in nutrients, the variability of Fuco is largely explained by the opportunistic nature of diatoms. The latter species are the most responsive to sporadic changes in abiotic factors such as an increase in the nutrient load (Fogg, 1991) and tend to thrive in nutrient-rich and turbulent regions (Tréguer *et al.*, 1995). Furthermore, in the Southern Ocean, high concentrations of Fuco are retrieved by our method. The latter result suggests that the island mass effect mainly benefits to diatoms as evidenced by in-situ observations (Blain *et al.*, 2008; Korb *et al.*, 2008).

In temperate and subpolar environments, diatoms and nano-phytoplankton are abundant, especially during the blooming season. As stratification increases with time and heating, nutrients become depleted in the surface layer, the grazer community develops, and the

production of diatoms declines in favor of nano-phytoplankton (Holligan *et al.*, 1993; Lochte *et al.*, 1993; Iglesias-Rodríguez *et al.*, 2002). Therefore, in contrast with diatoms, nano-phytoplankton remain abundant all year long, as evidenced by the dominance of nano-phytoplankton's associated pigments (19HF, 19BF, Allo and Chlb) in both winter and summer.

Conversely, as indicated by the quasi-stationary patterns of Zea and DVChla, smaller cells such as Cyanobacteria and other Pico-phytoplankton species are favored by more constant environmental conditions mainly observed in the tropical and subtropical regions throughout the year. The equatorial Pacific is characterized by the presence of an upwelling driven by northeast and southeast trade winds. Upwelling of cool water results in a large supply of nutrients to the surface layer (Chavez and Barber, 1987; Chavez, 1996). Despite a constantly favorable light/nutrient regime, Chl biomass and primary production are relatively low considering the nutrient levels in that area (Barber and Chavez, 1991). This HNLC characteristics appears to be primarily due to iron limitation (Barber and Chavez, 1991; Coale *et al.*, 1996) and, to a lesser extent, to regulation by silicate (Dugdale and Wilkerson, 1998) and grazing (Smetacek, 1999). These environmental conditions favor a phytoplankton community dominated by small cells (Chavez, 1996; Moon-van der Staay *et al.*, 2000).

The results of the SOM can be compared to the output of the PHYSAT method (Alvain *et al.*, 2005, 2008; Ben Mustapha *et al.*, 2013). This method offers a tool to identify dominant phytoplankton functional type by exploiting satellite reflectance anomalies at several wavelengths. The global phytoplankton type patterns provided by PHYSAT shows a qualitative agreement with our Max DP analysis. The global distribution of phytoplankton groups is characterized by the dominance of Synechococcus and Prochlorococcus groups in oligotrophic tropical waters, where Zea and DVChla are the most abundant pigments. Nanoeucaryotes and diatoms prevail in the eutrophic waters of high latitudes coinciding with higher 19HF/Fuco concentrations. In the PHYSAT climatologies, diatom blooms are clearly visible in areas characterized by strong upwelling conditions. Similarly, in our study, Fuco is the most abundant pigment in these regions. Therefore, we claim that our method offers the potential to investigate the spatial and temporal patterns of the phytoplankton community, in particular the variability of its dominant groups that compose this community.

6. Conclusion

We present robust estimations of the concentrations of various phytoplankton pigments by using self-organizing maps learned on a set of satellite derived Chla, remote sensing reflectance data, surface temperature and collocated pigment concentrations. In our study, it was shown that the SOM has efficiently modeled the relationship between the phytoplankton pigment concentrations and satellite data, which enables reliable estimation of the concentration of 10 different pigments (Chla, DVChla, Chlb, DVChlb, 19HF, 19BF, Fuco, Perid, Allo, Zea). The method proves to be applicable for estimating concentrations of not only Chla but also of other pigments. Cross-validation results indicate that estimations were robust for all pigments ($R^2 > 0.75$ and an average RMSE = 0.016 mg.m⁻³). The large database used to develop and calibrate the SOM led to a satisfying estimation over a wide spatio-temporal scale. Therefore, a consistent picture of several phytoplankton pigments indicating group-specific behavior on a global scale was shown, revealing also the uncertainties associated with each pigment.

Besides long records of satellite data provided by Globcolour, the SOM-Pigments can be applied to study variability and change of overall phytoplankton and physiological responses to environmental variables by generating global images of pigment concentration with daily, weekly, monthly and climatological temporal resolution. This method can be also applied on broader areas to study the dynamic of phytoplankton at meso-scale, which involves phytoplankton pigments different than these used in this study. For that, further regional studies should be conducted to evaluate the robustness of the SOM and its capacity of reconstructing the spatio-temporal variability of the phytoplankton dynamic.

Chapitre III : La Méditerranée ; Caractéristiques Géographiques, Hydrodynamiques et Ecosystémiques

1. Caractéristiques géographiques et hydrodynamiques

La Mer Méditerranée est une mer semi-fermée entourée de grandes masses continentales : l'Europe au Nord, l'Asie à l'Est et l'Afrique au Sud. Elle s'étend sur environ 4000 km, de 6 °W à 36 °E en longitude et de 30 °N à 46 °N en latitude. Elle est connectée à l'océan Atlantique par le détroit de Gibraltar, où un flux d'eau d'un million de m³ d'eau atlantique peu salée et chaude rentre en surface (0-200 m) et un peu moins d'un million de m³ d'eau méditerranéenne levantine froide et salée sort en profondeur (200-400 m). Elle est aussi reliée à la Mer Noire par les détroits de Bosphore et de Dardanelles, et à la Mer Rouge par le canal artificiel de Suez. La Mer Méditerranée est un «modèle miniature» de l'océan global où la majorité des phénomènes responsables de la circulation globale sont présents ce qui permet une compréhension plus facile des processus complexes des océans globaux étant donné sa latitude moyenne, allant de la mésoéchelle à l'échelle du bassin (Lacombe *et al.*, 1981; Siokou-Frangou *et al.*, 2010).

La Mer Méditerranée est composée de deux principaux bassins, le bassin Occidental et le bassin Oriental, connectés par le détroit de Sicile. Chaque bassin est subdivisé en plusieurs sous-bassins caractérisés par une topographie bien marquée, particulièrement dans la partie orientale où la profondeur atteint 4982 mètres. Le bassin Occidental est relativement plat, et le bassin Oriental est caractérisé par une alternance de dépressions profondes, vallées sous-marines, pentes raides, et plus de 700 îles et îlots répartis le long de l'archipel Égée à l'Est de la Grèce (Bergamasco and Malanotte-Rizzoli, 2010).

Le climat continental est particulièrement sec, même en hiver, et provoque une évaporation intense tout au long de l'année. Le bilan hydrique est donc fortement négatif et se traduit par une perte de 0,5 à 1 mètre d'eau par an sur l'ensemble du bassin méditerranéen (Bethoux, 1980) qui fait la Méditerranée s'enrichir en sel. Les bilans en eau et en sel de la Méditerranée sont globalement équilibrés par les échanges avec l'océan Atlantique au détroit de Gibraltar où l'eau atlantique peu salée et chaude ($S \sim 36$, $T \sim 20^\circ$) pénètre en surface en Méditerranée tandis qu'une eau Méditerranéenne plus salée ($S > 38$, $T \sim 13^\circ$) s'écoule en profondeur vers l'Atlantique (Lacombe and Tchernia, 1972).

Un caractère météorologique remarquable de la Mer Méditerranée, est la présence d'évènements de vent et de précipitation de petite échelle très violents. Les montagnes sont

le principal contributeur au ruissellement d'eaux douces vers la Mer Méditerranée. Des masses d'eau considérables, chargées en nutriments et riches en alcalinité totale, sont alors déversées à la mer par des fleuves côtiers, surtout dans la partie Nord-Ouest de la Méditerranée (Llasat *et al.*, 2013).

La Mer Méditerranée est donc un bassin de concentration qui transforme l'eau atlantique entrante en une eau méditerranéenne sortante. Ce fonctionnement induit une circulation dans le plan vertical contrainte par les différences de salinité et de température. Cette circulation thermohaline anti-estuaire est engendrée par la présence d'un gradient horizontal de densité, au niveau du détroit de Gibraltar, dirigé vers la Mer Méditerranée dans la couche superficielle et vers l'océan Atlantique dans la couche profonde, ainsi que par la différence du niveau moyen des eaux entre les deux compartiments marins plus bas en Méditerranée qu'en Atlantique (Lacombe and Tchernia, 1972).

2. Régime du vent

La mer Méditerranée, étant une zone presque entièrement entourée des chaînes montagneuses, est caractérisée par plusieurs sous-bassins où le flux d'air interagit souvent avec l'orographie environnante.

Dans la mer d'Alboran, le *Vendaval* qui est un vent du Sud-Ouest se produit principalement d'octobre à novembre et de février à mars, alors que le *Levanter* provenant de l'Est souffle pendant toutes les saisons. Dans la Méditerranée centrale, le mistral vent de Nord-Ouest froid et sec (Jansá, 1987; Jiang, Smith and Doyle, 2003; Guenard *et al.*, 2005) souffle dans le golfe du Lion, et peut atteindre parfois les côtes africaines. Quand il entre dans la mer Tyrrhénienne, le mistral prend le nom de *Maestrale*, en ayant une direction plus marquée vers le Sud. La Méditerranée centrale est balayée, surtout en hiver, par le *Libeccio*, vent de Sud-Ouest et par le *Sirocco* humide et chaud, qui en automne souffle du Sud vers l'Est, induisant des tempêtes bien connues dans la Mer Adriatique, qui inondent la lagune de Venise. La *Bora* est un vent du Nord fort et froid affectant l'ensemble de la mer Adriatique (Jurčec, 1981; Smith and Smith, 1987; Pandžić and Likso, 2005). Les vents de type *Bora* existent dans d'autres régions, comme dans le Nord de la mer Egée. Dans le bassin Levantin, le *Chlouk* provient de l'Afrique au printemps (Abboud-Abi Saab, 1985), et les *Etesians* qui sont particulièrement forts en été (Ziv *et al.*, 2004).

3. Circulation globale

3.1. Circulation de l'eau Atlantique

La circulation générale de la Mer Méditerranée est une circulation thermohaline où la source chaude est le détroit de Gibraltar et la source froide les échanges océan-atmosphère avec pour conséquence la formation d'eaux profondes (Golfe du Lion, Rhodes Gyre, Mer Egée ; (Bethoux, 1980; Bethoux *et al.*, 1999)). Cette circulation thermohaline est cyclonique : l'eau Atlantique (AW) pénètre en mer d'Alboran par le détroit de Gibraltar en formant deux vastes tourbillons anticycloniques (Millot, 1999; Millot and Taupier-Letage, 2005), le Western Alboran Eddy (WAE) et le Eastern Alboran Eddy (EAE) (Vargas-Yáñez *et al.*, 2002). Puis, l'AW longe la côte Sud du bassin Occidental en formant le courant Algérien (Algerian Current, AC) qui est le siège de grands tourbillons anti-cyclonique qui peuvent envahir tout le bassin algérien : le Western Algerian Gyre (WAG) et le Eastern Algerian Gyre (EAG) (Testor, Béranger and Mortier, 2005).

L'AW traverse le canal de Sardaigne et se sépare en deux branches au niveau du canal de Sicile; La première va dans le sous-bassin Tyrrhénien et poursuit son parcours le long de la côte italienne. Puis l'AW passe le détroit de Corse et entre dans le sous-bassin Liguro-Provençal pour former le Courant Nord (ou courant Liguro-Provençal) et longe les côtes Françaises et Espagnoles. Au niveau du canal de Sicile, la deuxième partie de l'AW entre dans le bassin Oriental, rejoint la côte Libyenne, en une ou plusieurs branches: l'Atlantic Tunisian Current (ATC), et l'Atlantic Ionian Stream (AIS) (Béranger *et al.*, 2004). Dans le bassin Levantin, le courant de bord Libyo-Egyptian Current (LEC) (Hamad, Millot and Taupier-Letage, 2005; Gerin *et al.*, 2009), forme des tourbillons pouvant s'accumuler dans des zones préférentielles contrains par la topographie. Les eaux qui arrivent à l'extrémité du bassin oriental à la fin de l'été deviennent très chaudes et très salées ($T = 28^{\circ}\text{C}$, $S = 39,5$) sous l'effet du climat estival.

Ensuite, l'AW longe les côtes du Proche-Orient puis les côtes Turques où elle forme le Courant d'Asie Mineure [Asia Minor Current (AMC)]. Une partie de l'AMC passe au Sud de la Crète et une autre partie entre dans la mer Egée et retourne le long des côtes de manière cyclonique. Une branche de l'AW passe le détroit d'Otrante et circule cycloniquement le long des côtes du sous-bassin Adriatique, fortement marquée dans le Sud du sous-bassin par la présence du

gyre Sud-Adriatique [South Adriatic Gyre (SAdG)]. L'AW retourne ensuite vers le canal de Sicile bouclant ainsi un trajet cyclonique en longeant les côtes Nord de la mer Ionienne.

3.2. Circulation intermédiaire et thermohaline

En hiver, les invasions de masses d'air continentales et sèches au-dessus du bassin levantin sont fréquentes. Elles engendrent un refroidissement intense des eaux de surface et une augmentation de la salinité, qui est à l'origine de mélanges verticaux importants. Ainsi dans le Nord du sous-bassin Levantin, et plus précisément dans la zone du gyre de Rhodes, se forme l'eau Levantine intermédiaire [Levantine Intermediate Water (LIW)], à une profondeur moyenne de 200-300 m (Lascaratos, 1993; Robinson *et al.*, 2001). La LIW est identifiable dans l'ensemble de la couche intermédiaire de la Méditerranée par un maximum de salinité entre 200 et 600 m.

Une partie de la LIW sort du sous-bassin Levantin et rejoint le canal de Sicile en traversant la mer Ionienne de manière cyclonique le long du talus continental. Une fois passée dans le bassin occidental, la LIW circule principalement le long de la côte Est du sous-bassin Tyrrhénien. Une faible partie passe le détroit de Corse alors que la majorité rejoint le canal de Sardaigne. Une autre partie forme un gyre cyclonique entre les canaux de Sicile et de Sardaigne (Sorgente *et al.*, 2011) duquel s'échappent des tourbillons anticycloniques (Béranger *et al.*, 2004). Dans le bassin Algéro-Provençal, une partie de la LIW est advectée vers le Nord le long des côtes de la Sardaigne et de la Corse. Une partie se sépare sous la forme de tourbillons (Sardinian Eddies (SE)) (Testor *et al.*, 2005), qui contribuent au transport de la LIW vers l'Ouest et Gibraltar par interaction avec les tourbillons algériens de surface (Millot and Taupier-Letage, 2005). Une autre partie de la LIW a un trajet moins direct et participe à la formation d'autres masses d'eau intermédiaires et profondes méditerranéennes. En passant les détroits de l'Arc Crétois, la LIW participe à la formation de l'eau crétoise intermédiaire [Cretan Intermediate Water (CIW)] ou profonde [Cretan Deep Water (CDW)] (Theocharis *et al.*, 1999). Ces eaux crétoises ressortent parfois du sous-bassin Egée à travers les détroits de l'Arc Crétois et se retrouvent dans les sous-bassins Ionien et Levantin.

En passant le détroit d'Otrante vers le sous-bassin Adriatique, la LIW participe à la formation de l'eau profonde Sud-Adriatique [South Adriatic Deep Water (SAdDW)] dans le gyre Sud-

Adriatique. L'eau profonde Nord-Adriatique [North Adriatic Deep Water (NAdDW)] est formée au Nord de la mer Adriatique et cascade vers le Sud (Artegiani *et al.*, 1997a, 1997b). La NAdDW et la SAdDW se mélangent (Manca *et al.*, 2006), passent le détroit d'Otrante ou elles forment l'eau adriatique profonde [Adriatic Deep Water (ADW)], qui plonge vers le fond du sous-bassin Ionien. L'eau profonde Est-méditerranéenne [Eastern Mediterranean Deep Water (EMDW)] est alors principalement composée d'ADW, même si selon les périodes la CDW peut entrer dans sa composition. La partie de la LIW advectée vers le Nord du bassin Algéro-Provençal participe à la formation de l'eau profonde Ouest-méditerranéenne [Western Mediterranean Deep Water (WMDW)], dans le Golfe du Lion principalement.

Cette région est dominée en hiver par les vents continentaux froids et secs (mistral et tramontane) qui entraînent le mélange de l'eau superficielle d'origine atlantique et l'eau levantine intermédiaire. La WMDW s'écoule au fond du sous-bassin Algéro-Provençal, vers l'Ouest et le détroit de Gibraltar ainsi que vers le sous-bassin Tyrrhénien par le fond du canal de Sardaigne.

Enfin, dans mer Tyrrhénienne est formée l'eau profonde tyrrhénienne [Tyrrhenian Deep Water (TDW)], soit localement à l'Est de Bonifacio (Fuda *et al.*, 2002), soit par interactions entre la LIW et la WMDW (Millot, 1999). La TDW sort de la mer Tyrrhénienne par le canal de Sardaigne, au-dessus de la WMDW, et s'écoule vers le détroit de Gibraltar (Millot and Taupier-Letage, 2005). A Gibraltar, une eau méditerranéenne froide et salée ($T < 13^{\circ}$, $S > 38$) rejoint l'océan atlantique [Mediterranean Outflow Water (MOW)], eau qui est principalement composée de LIW, de WMDW et de TDW (Millot, 2009).

4. Caractéristiques Ecosystémiques

La concentration des eaux océaniques en nutriment constituent un indicateur critique de la production et les cycles biologiques, d'une part, et des processus d'apport (naturel et anthropique) et de transfert d'autre part (Bethoux, 1989; 1992). Pourtant, les eaux de la mer méditerranée sont parmi les plus pauvres en nutriments de l'océan mondial en surface tant qu'en profondeur (Psarra *et al.*, 2005).

La Méditerranée est classée comme bassin oligotrophe (Sournia, 1973), sa production primaire est généralement faible et la concentration de chlorophylle dans les eaux du large

dépasse rarement 2 à 3 mg.m⁻³. La dynamique saisonnière du phytoplancton suit un cycle typique de région de moyenne latitude, avec une augmentation de la biomasse à la fin de l'hiver / au début du printemps et des productions très faibles en été. Cette saisonnalité typique de moyenne latitude a été régulièrement observée à la station (DYFAMED) située dans la partie nord-ouest du bassin, au cours d'environ 20 ans de mesures (Marty *et al.*, 2002).

Par la suite, une telle tendance a été confirmée par l'analyse des observations de couleur de l'océan effectuées suivant deux périodes différentes, par les capteurs satellitaires CZCS (1979-1985) et SeaWiFS (1997-2001) (Antoine and Andr, 1995; Bosc, Bricaud and Antoine, 2004). En ce qui concerne les distributions spatiales, le trait dominant qui a émergé est un gradient est-ouest de la chlorophylle à l'échelle du bassin, ce qui renforce le paradigme d'un bassin oriental extrêmement oligotrophe et d'un bassin occidental plus productif.

Plus précisément, les résultats des études par satellite peuvent être résumés comme suit:

- Des valeurs élevées de chlorophylle-a sur de grandes surfaces sont rarement détectées dans le bassin, à l'exception d'une prolifération importante observée dans la région liguro-provençale vers la fin de l'hiver.
- Les zones où la prolifération du phytoplancton est prononcée sont géographiquement bien localisées.
- L'augmentation de la biomasse du phytoplancton semble être fortement liée au forçage physique, ce qui crée des conditions localement favorables pour la croissance du phytoplancton.
- Parmi tous les types de forçage, le vent est considéré comme le facteur le plus important pour induire l'accumulation de biomasse de phytoplancton, ainsi que la présence de structures cycloniques. La prolifération dans la mer d'Alboran à prédominance anticyclonique, où la dynamique à moyenne échelle, associée à l'afflux d'eau de l'Atlantique, joue un rôle majeur.
- Des zones de haute biomasse plus confinées sont situées près des côtes, en particulier à proximité des grandes embouchures de rivières.

Chapitre IV : Télédétection du Phytoplancton de la Méditerranée ; Approche Régionale

1. Introduction

Les applications des approches par télédétection des caractéristiques du phytoplancton par télédétection satellitale de la couleur de l'océan sont peu nombreuses en mer Méditerranée. Ce n'est que récemment que Navarro *et al.*, (2014; 2017) ont exploité la méthode PHYSAT de Alvain *et al.* (2005), en réalisant une régionalisation pour la mer Méditerranée dénomé PHYSAT-Med. PHYSAT se base sur une relation empirique entre les réflectances normalisées (réflectances débarassées de l'influence de la Chl_a) et les pigments diagnostiques d'une base de données HPLC globale. PHYSAT-Med a été validé principalement pour les nanoeucaryotes, le *Prochlorococcus*, le *Synechococcus* et les diatomées et fournit le groupe de phytoplancton dominant pour chaque pixel de l'image satellitale étudiée. Les résultats fournis par PHYSAT-Med ont montré la dominance de *Synechococcus* par rapport aux prochlorophytes tout au long de l'année au niveau du bassin Méditerranéen, bien que les nanoeucaryotes fussent plus abondants pendant les mois d'hiver. De plus, les données de PHYSAT-Med ont mis en évidence une augmentation de la biomasse d'eucaryotes (principalement des diatomées) au printemps (de mars à avril), en particulier dans les mers Ligurienne et Adriatique.

Sammartino *et al.* (2015) ainsi que Brewin *et al.*, (2011) ont utilisé des modèles empiriques globaux, basés uniquement sur la concentration en chlorophylle-a, pour obtenir la distribution de la biomasse et de la taille du phytoplancton en mer Méditerranée. Les résultats de cette approche révèlent, qu'en mer Méditerranée, la contribution principale à la chlorophylle-a provient du pico-phytoplancton, en particulier dans les environnements pauvres en éléments nutritifs. Le micro-phytoplancton domine dans les zones de fort mélange vertical d'hiver et de formation d'eau profonde / intermédiaire, et dans les zones côtières en raison de la disponibilité en nutriments provenant des apports terrigènes. La composante du nano-phytoplancton reste plutôt répandue sur l'ensemble du bassin tout au long de l'année et sa contribution à la Chl-a est de l'ordre de 30 à 40%.

Finalement, Di Cicco *et al.*, (2017) ont étudié l'assemblage du phytoplancton et sa variabilité en analysant la relation entre le contenu en chlorophylle-a et la composition pigmentaire, ceci en mer Méditerranée. En suivant une approche globale basée sur l'abondance et en sélectionnant la chlorophylle totale-a (TChl_a) comme descripteur de l'état trophique de

l'environnement, des relations empiriques ont pu être identifiées entre la concentration de TChla et sept pigments accessoires considérés comme diagnostiques pour les principaux groupes d'algues. En termes de concentration de chlorophylle, dans les zones côtières prédominent des diatomées et des haptophytes, tandis que dans les eaux ultra-oligotrophes, les procaryotes prédominent vis à vis des autres groupes, constituant la composante principale du pico-phytoplancton.

Nous proposons dans ce chapitre une double extension à la Méditerranée de l'algorithme global SOM-Pigments dédié à l'estimation de la concentration des pigments secondaires ;

- la première application est axée sur le développement d'un second algorithme régionalisé à la Méditerranée pour l'identification du groupe de phytoplancton dominant à partir des images satellitaires en utilisant les sorties de l'algorithme SOM-Pigments
- la deuxième application servira à observer les différentes classes de taille du phytoplancton PSC sur la Méditerranée, en se basant sur les sorties de l'algorithme SOM-Pigments et des relations qui lient entre pigments et classes de taille. Nous proposons à la fin de ce chapitre une validation in-situ vis-à-vis des données in-situ de comptage de phytoplancton en tentant de décrire la relation entre chaque fraction de la Chla et les comptages de phytoplancton.

Partie 1 : Télédétection du groupe de phytoplancton dominant au niveau de la Méditerranée en utilisant des cartes auto-organisatrices (SOM)

Résumé :

Au cours des deux dernières décennies, plusieurs algorithmes satellitaires ont été proposés pour extraire des informations sur les groupes de phytoplancton à l'aide des données de couleur de l'océan.

Dans cette étude, une nouvelle méthode sera présentée en utilisant les cartes auto-organisées (SOM), qui sont des classificateurs de réseaux neuronaux, pour identifier les types fonctionnels du phytoplancton (PFT) et pour reconstruire la variabilité du phytoplancton avec une performance significative. Nous avons d'abord estimé les pigments de phytoplancton secondaires à partir des données satellite GlobColour et de l'AVHRR SST en utilisant une SOM dédié (SOM-Pigments), puis les groupes de phytoplancton dominants ont été identifiés à l'aide d'un autre SOM (SOM-PFT). L'étude a été concentrée sur la mer Méditerranée. Nous avons identifié six groupes de phytoplancton qui sont repartis en 4 classes et 2 genres : Bacillariophycées (=Diatomées), Haptophycées Chlorophycées, Cryptophycées, et deux genres appartenant au Cyanophycées: *Synechococcus* et *Prochlorococcus*. Ces groupes de phytoplancton ont montré une variabilité particulière due aux caractéristiques hydrophysiques et biogéochimiques complexes de la mer Méditerranée; Les Haptophytes et les Chlorophytes ont dominé pendant l'hiver et principalement dans le bassin ouest de la Méditerranée, tandis que les *Synechococcus* et les *Prochlorococcus* ont lutté pendant l'été. La dominance des diatomées a été principalement observée au printemps dans la mer des Baléares en réponse à des phénomènes de convection en eau profonde et près du littoral et des estuaires en raison d'importants apports continentaux de nutriments. Nous avons constaté que les Cryptophycées ont une faible présence dans la mer Égée en automne. Les tests de validation effectués sur des correspondances *in-situ* ont donné des résultats satisfaisants et ont démontré la capacité de la méthode à reconstruire efficacement les patterns spatio-temporels de groupes de phytoplancton en mer Méditerranée.

Phytoplankton diversity in the Mediterranean Sea from satellite data using Self-Organizing Maps

Roy El Hourany ^{1,2*}, Marie Abboud-Abi Saab ³, Ghaleb Faour ², Carlos Mejia ¹, Michel Crépon ¹ & Sylvie Thiria ^{1,4}

¹ IPSL/LOCEAN, Sorbonne Université (Université Paris VI, CNRS, IRD, MNHN), Paris, France,

² National Center for Remote Sensing, National Council for the Scientific Research, Beirut, Lebanon,

³ National Center for Marine Sciences, National Council for the Scientific Research, Batroun, Lebanon,

⁴ Versailles Saint-Quentin-en-Yvelines University, Versailles, France

Corresponding author: Roy El Hourany (roy.elhourany@locean-ipsl.upmc.fr)

Abstract

We present a new method to identify phytoplankton functional types (PFT) in the Mediterranean Sea from ocean colour data (Globcolour data in the present study) and AVHRR SST. The principle of the method is constituted by two very fine clustering algorithms one mapping the relationship between the satellite data and the pigments, the other between the pigments and the PFTs. The clustering algorithms are constituted of two efficient Self-Organizing Maps which are neural network classifiers. We were able to identify six phytoplankton functional types: Haptophytes, Chlorophytes, Cryptophytes, *Synechococcus*, *Prochlorococcus* and Diatoms. We found that these PFT present a peculiar variability due to the complex physical and biogeochemical characteristics of the Mediterranean Sea: Haptophytes and Chlorophytes dominate during winter and mainly in the western Mediterranean basin, while *Synechococcus* and *Prochlorococcus* thrive during summer. The dominance of diatoms was mainly observed in spring in the Balearic Sea in response to deep water convection phenomena and near the coastline and estuaries due to important continental inputs. Cryptophytes present a weak concentration in the Aegean Sea in autumn. The validation tests performed on in-situ matchups showed satisfying results and proved the ability of the method to reconstruct efficiently the spatio-temporal patterns of phytoplankton groups in the Mediterranean Sea. The method can be applied to other oceanic regions.

1. Introduction

The Mediterranean Sea is a semi-enclosed basin, covering approximately 0.8% of the world's ocean surface area. Although it has a limited geographical dimension, it is considered as one of the most complex marine environment with comparable features present in the global ocean such as a thermohaline circulation, deep-water convection events, complex geomorphology and bathymetry... (Bethoux et al., 1999).

Hydrological differences along the basin cause the presence of an increasing oligotrophic gradient from west to east. It results a west to east decrease of surface Chlorophyll-a concentration (Turley et al., 2000) that was observed *in-situ* and from space (Antoine et al., 1995), with the Eastern Mediterranean Levantine waters exhibiting highly oligotrophic conditions (Dowidar, 1984; Abdel-Moati, 1990; Yacobi *et al.*, 1995). However, the seasonal evolution of Chlorophyll-a (Chla in the following) distribution still follows the typical succession of temperate regions, characterized by a phytoplankton abundance increase in late winter/early spring, a decrease during the summer season and a second smaller phytoplankton bloom in autumn (Siokou-Frangou *et al.*, 2010; Sammartino *et al.*, 2015).

Phytoplankton community in oligotrophic areas throughout the world ocean is mainly composed by picoplankton (Yogev *et al.*, 2011). Nevertheless, the Mediterranean phytoplankton community reveals a considerable variability over both temporal and spatial scales, and large dissimilarities in phytoplankton assemblage composition (Siokou-Frangou et al., 2010). Many studies have pointed the dominance of picoplankton in the Mediterranean Sea due to its oligotrophy, but the occurrence of regional phytoplankton blooms causes the coexistence of numerous microalgal groups (Siokou-Frangou et al., 2010).

An extensive amount of observations of the phytoplankton community along the Mediterranean coastline is available. On the contrary, longitudinal data based on large-scale investigations in open sea waters are scarce in the literature (Ignatiades *et al.*, 2009). This lack of measurements can be partly overcome by using new tools, such as remote sensing techniques. Remote sensing of surface optical properties has provided synoptic views of the abundance and distribution of sea surface constituents. Satellite ocean color sensors have been an effective platform to estimate Chla in the surface waters, providing synoptic

measurements over the world ocean (Longhurst *et al.*, 1995; Antoine, André and Morel, 1996; Behrenfeld and Falkowski, 1997; Behrenfeld *et al.*, 2005; Westberry *et al.*, 2008).

The spatio-temporal distribution and identification of remote sensing derived phytoplankton groups give powerful insights on the dynamics of the marine food web and the ocean's role in climate regulation in the context of the global change (Bracher *et al.*, 2017b; Kostadinov *et al.*, 2017). It has been for several decades recognized that detection of phytoplankton from remote sensing images was a major challenge in ocean optics (Platt *et al.*, 2006). Therefore, several remote sensing algorithms have been developed to characterize the global distribution patterns of phytoplankton functional types (PFT) or size classes (PSC) (Alvain *et al.*, 2005, 2008; Ciotti and Bricaud, 2006; Uitz, Claustre, Morel and Stanford B. Hooker, 2006; Hirata *et al.*, 2008; Aiken *et al.*, 2009; Brewin *et al.*, 2010; Sathyendranath *et al.*, 2014). This was done by using relationship between Chla and PSC (Bricaud *et al.*, 2006; Uitz, Claustre, Morel and Stanford B Hooker, 2006; Hirata *et al.*, 2008; Mouw and Yoder, 2010) and by taking into account the spectral information (Alvain *et al.*, 2005; Ben Mustapha *et al.*, 2013; Sathyendranath *et al.*, 2014). Some of these algorithms are based on spectral features, such as backscattering (Kostadinov, Siegel and Maritorena, 2009), absorption (Ciotti and Bricaud, 2006; Bracher *et al.*, 2009; Mouw and Yoder, 2010; Roy *et al.*, 2013) or a hybrid between absorption and backscattering (Fujiwara *et al.*, 2011) while other algorithms exploit second-order anomalies of water leaving radiance such as PHYSAT (Alvain *et al.*, 2005; 2008; Ben Mustapha *et al.*, 2013; Navarro *et al.*, 2014; 2017). The PHYSAT method is based on the identification of specific signatures in the normalized water leaving radiance spectra measured by an ocean color sensor after removal the impact of Chla variations. Applied to oceanic Case I water, this method is designed to detect satellite pixels in which the dominant groups are nanoeukaryotes (and separately Phaeocystis-like and coccolithophores), two types of picoplankton (Prochlorococcus and Synechococcus-like cyanobacteria) and diatoms (Alvain *et al.*, 2008). Furthermore, a regionalized version of the PHYSAT method has been specifically developed for the Mediterranean Sea due to the peculiarities of phytoplankton assemblages and succession than can be found in this basin and its particular optical properties (Navarro *et al.*, 2014; 2017).

We have developed an innovative approach using self-organizing maps (SOM) (Kohonen, 2013) in order to evidence the relationship between satellite data measured at the ocean

surface and the phytoplankton assemblage. The SOM are unsupervised neural classifiers commonly used when dealing with environmental studies (Cavazos and Cavazos, 1999; Hewitson and Crane, 2002; Richardson, Risien and Shillington, 2003; Liu and Weisberg, 2005; Liu, Weisberg and Mooers, 2006; Niang *et al.*, 2006; Reusch, Alley and Hewitson, 2007). In El Hourany *et al.*, (2019), a global approach was built to estimate phytoplankton pigment HPLC data using SOM (SOM-Pigments) and optical satellite data. This approach was calibrated and efficiently validated using a global database of HPLC measured pigments collocated with Globcolour satellite data (Chla, Rrs at four wavelengths; 412, 443, 490 and 555 nm) alongside the AVHRR SST. The output of this method provided 10 types of pigment that can be associated with phytoplankton groups (Vidussi *et al.*, 2001).

In the present work, we propose a new SOM (the so-called SOM-PFT) to identify dominant PFTs from satellite data. This algorithm is regionalized and centered on the oligotrophic, yet complex, ecosystem of the Mediterranean Sea and allowed us to track specific features of phytoplankton assemblage along with their associated bio-optical properties. While observing the availability of in-situ HPLC measurements in the Mediterranean Sea, the heterogeneity of the sampling locations, which are significantly more abundant in the Western basin, limits the development of a regional algorithm. Therefore, our method was based on estimating the phytoplankton pigments composition of the Mediterranean Sea from satellite data by using the SOM-Pigments developed by El Hourany *et al.*, (2019). Then the dominant PFTs were estimated by using a dedicated SOM (the SOM-PFT), which was calibrated on the output of the SOM-Pigments. The performances were evaluated by using in-situ data sampled in the Mediterranean Sea.

The paper is articulated as follows. In section 2, we briefly describe the SOM-Pigments outputs, the Satellite Data used to estimate the pigment concentrations and the in-situ HPLC samples used to validate the results in the Mediterranean Sea. The SOM algorithm and the calibration procedure of SOM-PFT are described in section 3. In section 4 and 5, we present the results and the validation of the method, while discussing the spatio-temporal variability of dominant PFTs in the Mediterranean Sea.

2. Materials

2.1. The SOM-Pigments:

The SOM-Pigments (El Hourany *et al.*, 2019) allows the estimation of phytoplankton pigment concentrations in oceanic waters from satellite ocean color data and SST. The global database used to calibrate and cross-validate the SOM-Pigments gathers several oceanic campaigns and published databases of HPLC measurements which include 10 phytoplankton pigment concentrations : Chlorophyll-a (Chla), Divynil-Chlorophyll-a (DVChla), Chlorophyll-b (Chlb), Divynil-Chlorophyll-b (DVChlb), 19'Hexfucoxanthin (19HF), 19'Butfucoxanthin (19BF), Fucoxanthin (Fuco), Peridinin (Perid), Alloxanthin (Allo), Zeaxanthin (Zea). The SOM-Pigment was calibrated on a database comprising 12000 in-situ samples, which were collocated to remote sensing reflectance data at 4 wavelengths (412, 443, 490 and 555 nm) and Chla GlobColour data, along with SST AVHRR data described in the following. The results of the cross-validation procedure scored a coefficient of determination (R^2) of 0.75 and an average RMSE of 0.016 mg.m⁻³.

The SOM-Pigment allowed us to estimate the 10 phytoplankton pigment concentrations cited above from satellite data on the global scale.

The output of such method regroups images (4km, gridded) of 10 pigment concentrations. Using these pigment concentrations, it is possible to determine several PFT. For example, Zea, DVChla and DVChlb are associated with cyanobacteria (Guillard *et al.*, 1985, Dandonneaux *et al.*, 2004). Fuco is the principal marker of diatoms (Jeffrey, 1980). For nanoplankton quantification, Allo is a pigment typical of the cryptophytes (Gieskes and Kraay, 1983); 19HF, whose concentration is related to Haptophytes and chromophytes nanoflagellates (Wright and Jeffrey, 1987); 19BF a typical marker of chromophytes nanoflagellates (Wright and Jeffrey, 1987) and Chlb for Chlorophytes nanoflagellates and last, we also used the Perid, which appears in dinoflagellates (Jeffrey and Hallegraeff, 1987).

2.2. Satellite Data

In the following, we present the satellite data used to estimate the phytoplankton pigment concentrations in the Mediterranean Sea with the SOM-Pigments.

- GlobColour Data:

To extend existing time series beyond that provided by a single satellite sensor, the ESA initiated the GlobColour project (<http://www.globcolour.info/>) to develop a satellite based ocean colour data set to support global carbon-cycle research. It aims at satisfying the requirements of scientists for a long (over a decade) time-series of consistently calibrated global ocean colour information with the best possible spatial coverage. This has been achieved by merging data from SeaWiFS, MODIS, VIIRS, MERIS and OLCI satellite sensors.

The GlobColour project provides a continuous data set of merged level 3 (Mapped, 4km) daily Remote sensing reflectance (Rrs). This product is generated for each instrument, using the corresponding Level 2 data. The merged Rrs is then computed as the weighted average of all the single-sensor products. Meanwhile, the Chla was estimated via the OC5 algorithm (Gohin, 2011).

- AVHRR SST Data

Characterized by a defined annual cycle, the use of SST to calibrate SOM-Pigments permitted to better fit the relationship between the *in-situ* HPLC and satellite data, and therefore facilitated the identification of thermophile phytoplankton groups.

For that, the SST data were generated at 4 km resolution and at a daily frequency using AVHRR instruments aboard NOAA polar-orbiting satellites. Retrieval algorithms for sea surface temperature from AVHRR are mainly based upon the multi-channel sea surface temperature (MCSST) algorithm (McClain *et al.*, 1985).

2.3. Med HPLC pigment Dataset (Med-Pigments)

The in situ HPLC database (Med-Pigments) used to validate the method is a compilation of different datasets and campaigns performed in the Mediterranean Sea (Table A2.1, Figure A2.1). A total of 1772 HPLC samples were compiled in this database, limited to the first optical depth which is about 15-35m on average (D'Ortenzio and Ribera d'Alcalà, 2008). The spatial availability of the data is heterogeneous; more than 70% of the data were measured at a specific station, the Boussole/DYPHAMED stations anchored in the Ligurian Sea, which are dedicated to the study of biogeochemical and physical processes. Thus, Med-Pigments can drive a limitation to develop a representative algorithm for the whole Mediterranean and therefore only used to validate the output of the proposed method.

Table A2.1: Available in-situ HPLC inventory in the Mediterranean Sea.

	Cruises	Location	Périod	N	%
Calibration/Cross-Validation	Prosope	Western Basin	9/4/1999-4/10/1999	59	3
	SODYFT	Ligurian Sea	02/25/2002–12/19/2005	160	9
	SESAME_IT1_BOT	Adriatic Sea	02/16/2008–02/25/2008	21	18
	SESAME_IT2_BOT	Ionian Sea	03/02/2008–03/08/2008	16	
	SESAME_IT3_BOT	Ionian Sea	03/17/2008–03/18/2008	16	
	SESAME_IT4_BOT	Western Basin	03/20/2008–04/05/2008	24	
	SESAME_IT5_BOT	Ionian Sea	09/16/2008–09/26/2008	20	
	SESAME_IT6_BOT	Ionian Sea	09/20/2008–09/23/2008	11	
	SESAME_IT7_BOT	Adriatic Sea	10/09/2008–10/13/2008	21	
	SESAME_SP1_BOT	Alborán Sea	04/08/2008–04/11/2008	5	
	SESAME_SP2_BOT	Alborán Sea	09/20/2008–09/27/2008	39	
	SESAME_Turkey_BOT	Levantine Basin	03/20/2008–04/26/2008	34	
	SESAME_Turkey_BOT	Levantine Basin	09/19/2008–10/19/2008	10	
	SESAME-IS-1_BOT	Levantine Basin	04/08/2008–04/09/2008	6	
	SESAME-IS-2_BOT	Levantine Basin	09/07/2008–09/12/2008	14	
	SES_UNLUATA_gr1	Levantine Basin, Aegean Sea	03/27/2008–04/06/2008	12	
	SES_UNLUATA_gr2	Levantine Basin, Aegean Sea	08/29/2008–09/07/2008	12	
	BOUM_bot	Mediterranean Sea	06/21/2008–07/18/2008	64	
	Tara_oceans	Mediterranean Sea	9/20/2009-10/27/2013	115	
Validation	BOUSSOLE	Ligurian Sea	07/22/2001–11/10/2016	1113	
	Total			1772	

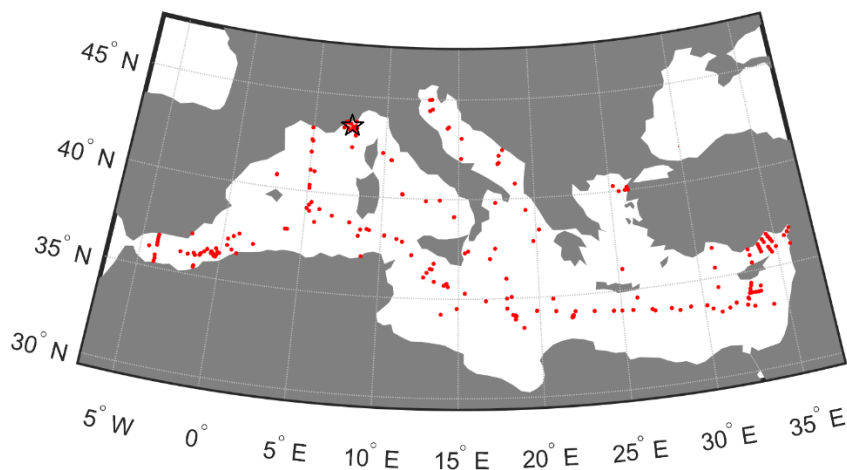


Figure A2.1: Localities of the HPLC samples in the Mediterranean Sea (Boussole station is represented with a star).

3. The Proposed Method

3.1. Self-organizing maps (SOM), general concept

The SOM algorithms (Kohonen, 2013) are nonlinear unsupervised classification methods. Such methods are able to cluster vectors of a multidimensional database into classes represented by a fixed network of neurons (the SOM map). The self-organizing maps are defined by a web-like graph, usually a rectangular grid of dimension $p \times q$ where each node is a neuron. This graph structure is used to define a discrete distance (denoted by δ) between the neurons of the map which represents the length of the shortest path between two neurons. Moreover, the SOM enables the partition of the initial database in such a way that each cluster is associated with a neuron of the map and is represented by a synthetic multidimensional vector (the referent vector w). Each vector z_i of the initial database will be assigned to the neuron whose referent w is the closest, in the sense of the Euclidean Norm, and will be called the projection of the vector v on the map.

The cost function can be written as:

$$J_{SOM}^T(\chi, W) = \sum_{z_i \in SOM} \sum_{c \in SOM} K^T(\delta(c, \chi(z_i))) \|z_i - w_c\|^2 \quad (1)$$

where $c \in SOM$ indices the neurons of the SOM map, χ is the allocation function that assigns each element z_i of the data to its referent vector w_c , $(c, \chi(z_i))$ is the discrete distance on the map SOM between a neuron c and the neuron allocated to observation z_i , and K^T a kernel function parameterized by the temperature T that weights the discrete distance on the map and decreases during the minimization process. This cost function takes into account the proper inertia of the partition of the data set and ensures that its topology is preserved. In the same way, two close neurons on the SOM-map represent close data of the initial database. The estimation of the referent vectors w of a SOM and the topological order is made through the Training phase.

The referent vectors are obtained by a weighted average of the initial data's vectors assigned to each neuron and their neighborhood on the map.

We introduce in the following a dedicated SOM map specialized in identifying dominant PFTs on the Mediterranean Sea (SOM-PFT).

3.2. SOM-PFT: calibration and validation procedure

Step 1: Pigment database for retrieving the PFTs in the Mediterranean Sea

The PFTs can be retrieved from the secondary pigments. The Med-Pigments dataset being sparse in time and space, we built a high coverage secondary pigment database by processing the globcolour and AVHRR SST data with the SOM-Pigments as did El Hourany et al (2019). Level 3 mapped 4km daily images of SST, Chla and Rrs at four wave-length (412, 443, 490 and 555 nm) were used as input of SOM-Pigments. We processed 7600 daily images of Globcolour and AVHRR from 1997 up to 2018 generating more than 3 million pixels sampling the Mediterranean. The outputs are thus constituted by 9-dimensional vectors, each vector representing nine secondary pigments (DVChla, Chlb, DVChlb, 19HF, 19BF, Fuco, Perid, Allo and Zea) on a 4kmX4km pixel,. The performances of the SOM-Pigments in the Mediterranean Sea were check with a cross-validation algorithm, and on the Boussole data which were not learned in the SOM-Pigments calibration.

We then computed for each pixel the relative values of the nine pigment concentrations (P_r : $DVChla_r$, $Chlb_r$, $DVChlb_r$, $19HF_r$, $19BF_r$, $Fuco_r$, $Perid_r$, $Allo_r$ and Zea_r) using the following equation:

$$P_r = \frac{P}{\sum P} \quad (2)$$

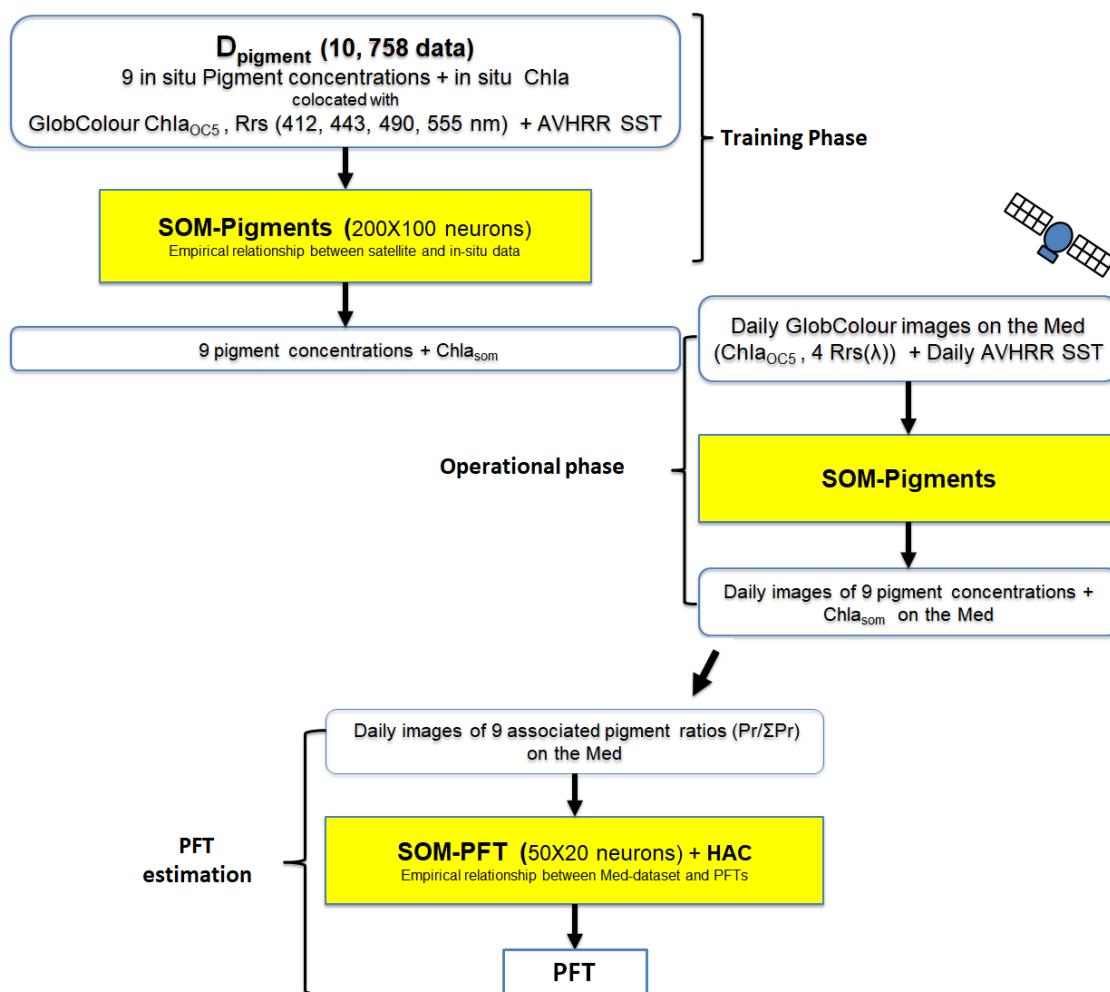
Where P is the estimated phytoplankton pigment concentration and $\sum P$ is the sum taken over the 9 P :

$$\sum P = DVChla + Chlb + DVChlb + 19HF + 19BF + Fuco + Perid + Allo + Zea$$

Each of the 3 million pixels is now represented by its nine P_r ; we gathered all the pixels in a specific database denoted Pr-MED.

Step 2: identification of PFTs in the Med

We then used Pr-MED to determine specific PFT classes in the Mediterranean Sea from the 7600 images under study. For that, we trained a dedicated SOM (SOM-PFT), which is composed of $50 \times 20 = 1000$ neurons, on Pr-MED. The 1000 neurons represent a fine description of the initial database. Each satellite pixel of Pr-MED is thus assigned to a particular neuron, which represents a grouping of similar pigment assemblage. In order to



A2.2: Flowchart summarizing the methodology to identify dominant PFTs from satellite data using SOM-Pigments and SOM-PFT on the Mediterranean Sea.

reduce the number of clusters, we then applied a hierarchical clustering algorithm (HAC) using the Ward dissimilarity (Jaines and Dubes, 1998), which regroups the neurons presenting statistical similarities into classes. The problem is thus to choose a small number of classes that “best” characterize specific phytoplankton pigment assemblages associated with specific PFTs. We found that six classes give a good insight on the Mediterranean biological diversity. In the following, each pixel is thus assigned to a given class. Therefore, each of the 7600 daily images are segmented using the six HAC classes.

Each class is thus constituted by the union of particular pigment compositions associated with a specific dominant PFT. Figure A2.2 summarizes the steps described above in a simplified flowchart.

Step 3: Validation of the SOM-PFT using Mediterranean in-situ HPLC Data

We chose to validate the SOM-PFT classification via two procedures:

- We first used the PFT labeling proposed by Alvain *et al.*, (2005) denoted Alvain criteria in the following. We applied the relative pigments ratios (P_{rel}) to identify the PFTs:

$$P_{rel} = \frac{P}{Chla + DVChla} \quad (3)$$

Alvain *et al.*, (2005, table 4) provided a table presenting threshold values on the different P_{rel} whose combinations are related to specific phytoplankton groups (Hapto, Proc, SLC, Diatoms and Dinoflagellates). Therefore, one of the tests consists in analyzing the coherence of the SOM-PFT output with the PFT given by the Alvain criteria (Table A2.2). For that, we estimated the PFT associated with the Med-Pigments in situ measurements by using the Alvain criteria. A match up was then defined by collocating each in situ measurement to the Globcolour and AVHRR data and by extracting the nearest pixel in a 3X3 pixel box surrounding the in situ measurement site. The secondary pigments of that pixel were estimated via SOM-Pigments and are presented as input to the SOM-PFT to identify the PFT. The two PFTs were then compared.

- As a second test, we validated our PFT retrieval algorithm (SOM-Pigments → SOM-PFT+HAC) on the Boussole measurements. A 10x10 pixel box was extracted from Globcolour and AVHRR satellite data around the Boussole station (Fig. A2.1) from July 2001 till November 2016. At each time, the PFT retrieval algorithm was applied on the pixel box and the ratio of each PFT class given by the SOM-PFT were calculated. The validation consists in analyzing the correlations between the time series of each PFT class ratio with the time series of the corresponding in situ P_{rel} . This experiment allowed us to observe and analyze the temporal coherence of this method compared to a long in-situ time series.

Table A2.2: Alvain Criterium. Thresholds used in this study to associate a relative pigment concentration inventory to a specific phytoplankton group (Alvain et al., 2005, Table 4).

	Pheo a	DVChla	Perid	Fuco	19HF	Zea
Diatoms	<0.30	<0.40	<0.10	>0.18	-	<0.20
Prochlorococcus	<0.30	>0.40	<0.10	-	-	>0.35
Haptophytes	<0.30	<0.40	<0.10	-	>0.14	<0.20
SLC*	<0.30	<0.40	<0.10	-	-	>0.20
Dinoflagellates	<0.30	<0.40	>0.10	-	-	<0.20
* <i>Synechococcus</i> -like Cyanobaterias Values in bold correspond to the primary biomarkers shown						

4. Results

The cross-validation of the SOM-Pigments showed a satisfactory result endorsing the validity of the SOM-Pigments database in the Mediterranean Sea. We found a $R^2 > 0.62$ (with a minimum $R^2 = 0.62$ for Allo concentrations, and the highest $R^2 = 0.85$ for Fuco concentrations) and a low average RMSE of 0.015 mg m^{-3} (Table A2.3, Figure A2.3)

This result allowed us to proceed with the reconstruction of the secondary pigment concentrations images on the Mediterranean as described in section 3.2, step 1. We found a coherent spatio-temporal variability of each pigment. Figure A2.4 shows a monthly climatology of the pigment concentrations and the Chla OC5 (1997-2018). It can be seen that most of the diagnostic pigment variability follows that of Chla, since it is the major proxy of the total phytoplankton biomass, except DVChla, DVChlb and Zea which present an opposed behavior.

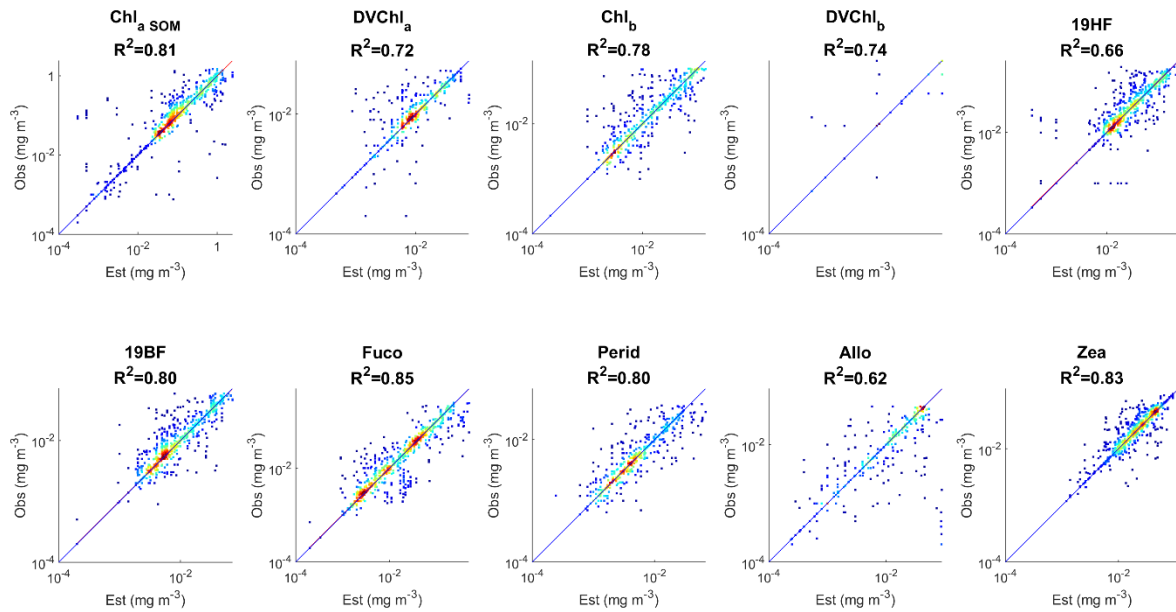


Figure A2.3: Scatter plots representing the cross-validation results of the SOM-Pigments using the Med HPLC Dataset.

Table A2.3: Statistical results of the cross-validation procedure performed by the SOM-Pigments using the Med HPLC Dataset:

Pigment	R2	RMSE [mg m-3]	N Obs
Chl _a _{SOM}	0.81	0.21	1113
DVChl _a	0.72	0.007	663
Chl _b	0.78	0.015	858
DVChl _b	0.74	0.0005	79
19HF	0.66	0.023	1030
19BF	0.80	0.006	1096
Fuco	0.85	0.044	1133
Perid	0.80	0.006	890
Allo	0.62	0.014	579
Zea	0.82	0.008	1241

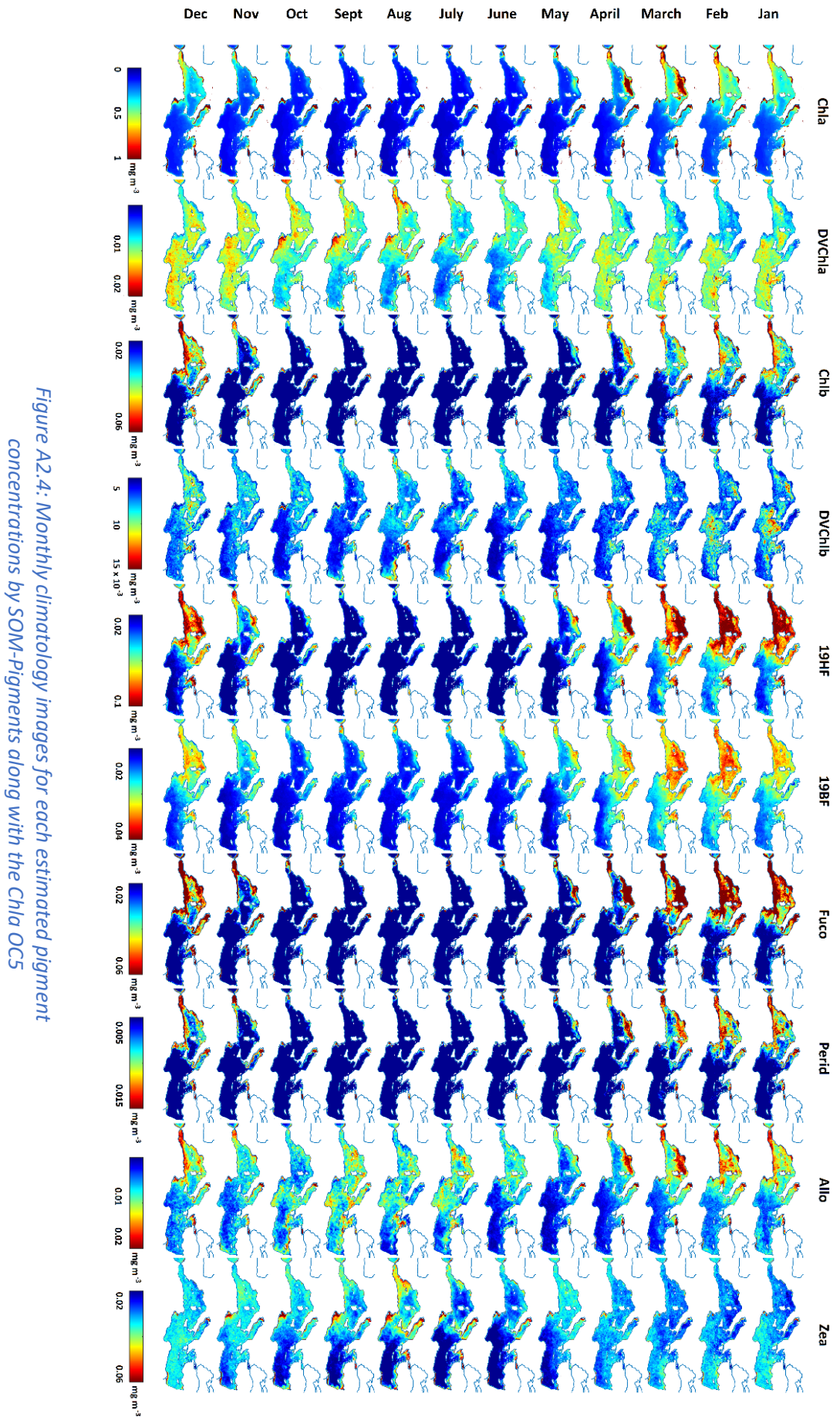


Figure A2.4: Monthly climatological images for each estimated pigment concentrations by SOM-Pigments along with the Chla OCS

The results of the validation of the Pr-Med dataset with respect to Med-Pigments one and the spatio-temporal reconstruction of the different phytoplankton secondary pigments proves the coherence of Pr-Med dataset used to train SOM-PFT. In the following, an analysis of SOM-PFT is presented along with the results of the validation experiments.

4.1. Analysis of the SOM-PFT

Figure A2.5 represents a two-dimensional map of the neurons of the SOM-PFT after the training phase. It is seen that the neurons are coherently organized with respect to the intensity of the different Pr . The different Pr distribution are well distinctive on the SOM grid and poorly correlated except 19HF and 19BF images which somewhat overlaps. High $DVChl_a$ concentration ($Pr > 40\%$) occupied the upper left corner of the neuron's map, whereas high Zea_r concentration ($Pr > 50\%$) were clustered at the left corner extending to the top. This means that the left corner of the SOM-PFT is specialized in Pico class detection. Whereas the $19HF_r$ and $19BF_r$ pigment ratios occupy the bottom side of the SOM-PFT with ratios $>40\%$ and $>15\%$ respectively. High Chl_b_r concentration ($Pr > 50\%$) are found near the bottom right corner, and most of the $Allo_r$ information is found at the top right corner. The bottom right corner of SOM-PFT contains information on $19HF_r$, $19BF_r$, $Allo_r$ and Chl_b_r , therefore this SOM-PFT part is specialized in the Nano sized class detection. Whereas for the $Fuco_r$, high ratios $>50\%$ are found near the top right corner. The $Perid_r$ seem to overlap the $19HF_r$.

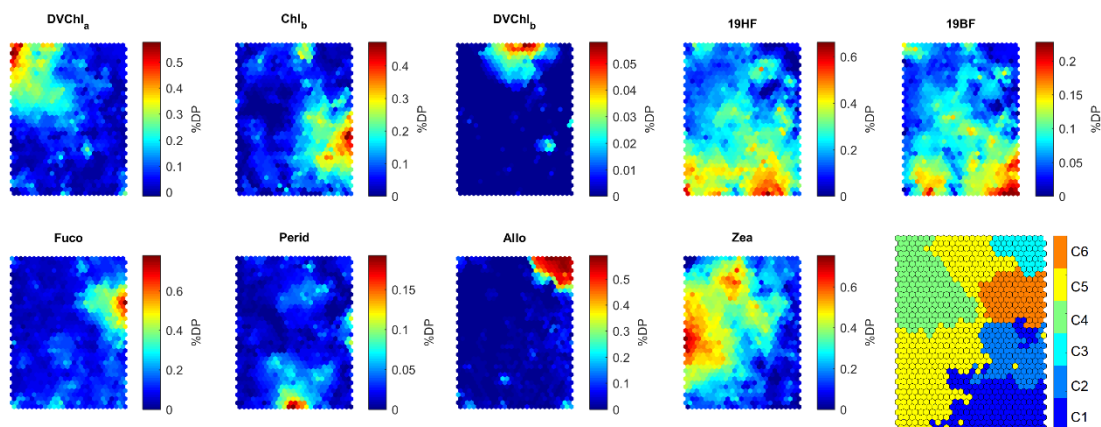


Figure A2.5: Discrete representation of each Pr on the SOM-PFT and the hierarchical clustering in 6 classes (bottom right panel)

The six classes defined by the HAC are represented on the SOM-PFT grid in Figure A2.4 last panel. Figure A2.6 shows the boxplots of the pigment ratio composition regrouped in each class. The segmentation of the SOM-PFT into classes further highlights the specialization of each neuron to identify a specific pigment ratio dominance indicating a PFT: C1 is characterized by high 19HF_r and indicates dominant Haptophytes nanophytoplankton (Hapto). High 19HF_r and Chlb_r characterize C2 and therefore this class can be attributed to Chlorophytes nanophytoplankton (Chloro). C3 is characterized with high Allo_r indicating a class of Cryptophytes nanophytoplankton (Crypto). Meanwhile, C4 regrouped neurons with high DVChla_r and Zea_r indicating the dominance of *Prochlorococcus* picophytoplankton (Proc), whereas C5 is characterized with high Zea_r and can be attributed to dominant *Synechococcus* picophytoplankton (Syne). Finally C6 is highlighted by high Fuco_r and denotes a dominance of Diatoms microphytoplankton (Diat).

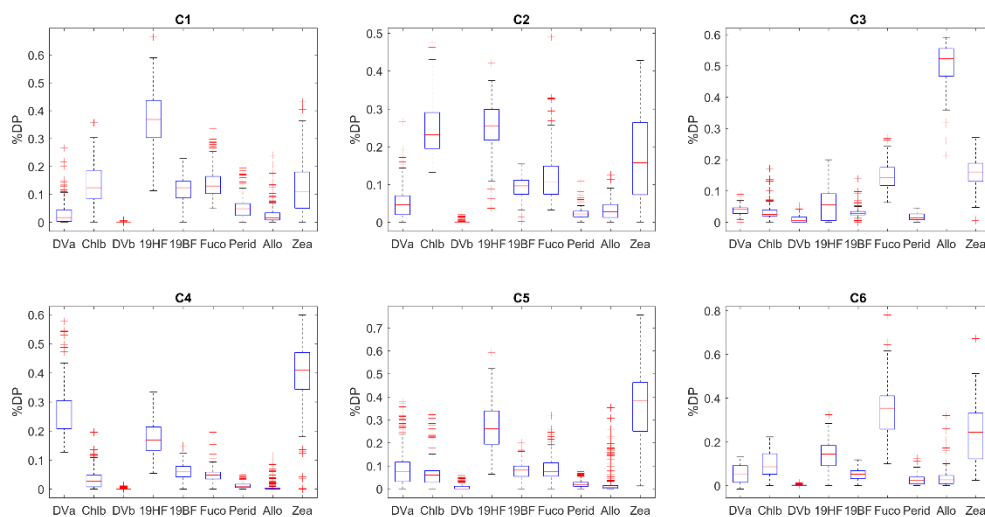


Figure A2.6: Boxplots of the nine pigment ratios for each class; C1: Haptophytes (Hapto), C2: Chlorophytes (Chloro), C3: Cryptophytes (Crypto), C4: Prochlorococcus (Proc), C5: Synechococcus (Syne), C6: Diatoms (Diat)

4.2. Validation experiments

4.2.1 Experiment 1: Validating Med-Pigments.

As mentioned above, the Alvain *et al.*, (2005) criteria (section 3.2) considers 5 PFT (Hapto, Proc, Syne, Diat, Dino) while the SOM-PFT allows to consider 6 PFT whose only 4 are in common (Hapto, Proc, Syne, Diat). The labeling thresholds of Alvain *et al.*, (2005) take into account 6 pigment ratios: Pheopigments, Fuco, 19HF, Perid, Zea and DVChla. To proceed with

the labeling, we selected from Med-Pigments 764 in situ measurements with no missing values.

In table A2.4, the confusion matrix is presented. It is obtained by comparing the results of the SOM-PFT with the Alvain *et al.*, (2005) criteria for the Med-Pigments dataset. Accordingly, 380 matchups were labeled as Hapto, 3 as Proc, 253 as Syne, 37 as Diatoms, 35 Dinoflagellates and last 56 samples were not assigned to any category. The SOM-PF presents a performance of 70% (265 in-situ versus 380 satellite matchups) of the Hapto class, of 75.5% (191 in-situ versus 253 satellite matchups) for the Syne and finally of 94.6% (35 in-situ versus 37 matchups) for the Diatoms.

Misplacements were found mostly when identifying Hapto and Syne dominated samples, where 23% (88 out of 380 matchup) of Hapto were identified as Syne, and 11.9% (30 out of 253 matchups) of Syne were assigned as Hapto by the SOM-PFT. Meanwhile, the Dino labeled samples and the samples that did not comply with any of the proposed thresholds in Alvain *et al.*, (2005) were also presented in table A2.4. We found that 54% of the Dino samples (19 out of 35) were assigned by SOM-PFT to Hapto and 28.6% (10 out of 35) to Diatoms. And at last, 30.4% (17 out of 56) of the samples unidentified by Alvain *et al.*, (2005) were assigned to Hapto by the SOM-PFT, while 44.6% (25 out of 56) were classified as Crypto or Chloro by the SOM-PFT.

Table A2.4: Confusion matrix showing the results of the validation test performed by comparing the SOM-PFT classes with the in-situ HPLC PFT computed with the Alvain et al., (2005) criteria.

Method		Alvain <i>et al.</i> , 2005					Unidentified
		Hapto	Proc	Syne	Diat	Dino	
SOM -PFT	Hapto	70% (265)	-	11.9% (30)	5.4% (2)	54.3% (19)	30.4% (17)
	Proc	1.3% (5)	33.3% (1)	7.5% (19)	-	2.9% (1)	3.6% (2)
	Syne	23.2% (88)	66.7% (2)	75.5% (191)	-	8.5% (3)	14.3% (8)
	Diat	4.5% (17)	-	3.9% (10)	94.6% (35)	28.6% (10)	7.1% (4)
	-	-	-	-	-	-	-
	Crypto/Chloro	1.3% (5)	-	1.2% (3)	-	5.7% (2)	44.6% (25)
N Obs		380	3	253	37	35	56

4.2.1. Experiment 2: Validation by using the Boussole time-series

In order to validate the coherence of the PFT reconstruction by the SOM-PFT, we analyzed a 10x10 pixels box around the location of the Boussole site (Fig. A2.1) from July 2001 till November 2016. The results are given in Figure A2.7(a, b, c, d) which shows the daily estimation of the satellite PFT percentage in this box. To simplify the analysis, we regrouped 3 PFT classes (Hapto, Chloro, Crypto) in a single one that represents nanoplankton size fraction (Nano).

The Spearman correlation (R_s) between the satellite PFT and their corresponding in situ pigment ratios (Table A2.5) was estimated for Diatoms, Nano, Proc and Syne. We found a good agreement the two set of measurements. The R_s values (Table A2.5) are of 0.47, 0.53, 0.11 and 0.54 for Diatoms, Nano, Proc and Syne respectively. A comparison between the in situ Chla measurement (whose concentration is maximal in winter) and satellite PFTs emphasizes the contribution level of each PFT class to Chla at the Boussole site. A relatively high R_s of 0.58 was noted for the Diatoms, 0.46 for Nano, while for the Pico classes, Proc and Syne showed a high inverse correlation of -0.65 and -0.79 respectively.

During the spring bloom period, the Diatoms and the associated $Fuco_{rel}$ reach their maximum percentage (Figure A2.7.b) when the mixed layer depth is maximum as mentioned by Marty *et al.*, (2002). The Hapto and Chloro class distribution present maxima percentages during winter, usually around January, while Crypto are peaking in autumn (Figure A2.7(c)). The seasonal pattern of the three nanophytoplankton classes coincides with the variability of in situ $19HF_{rel}$, $Chlb_{rel}$ and $Allo_{rel}$. Meanwhile, *Prochlorococcus* show maximum values in autumn during several years, at the end of the stratification period (Figure A2.7 (d)). The maxima

Table A2.5: Spearman correlation results performed on the Boussole time-series obtained by comparing the dominant satellite PFT percentage to the in-situ pigment ratios (Prel)

R_s	P_{rel}	Chla <i>in-situ</i>	N Obs
Diat	0.47*	0.58*	392
Nano	0.53*	0.46*	374
Proc	0.11	-0.65*	209
Syne	0.54*	-0.79*	393
*Significant at the 0.001 level			

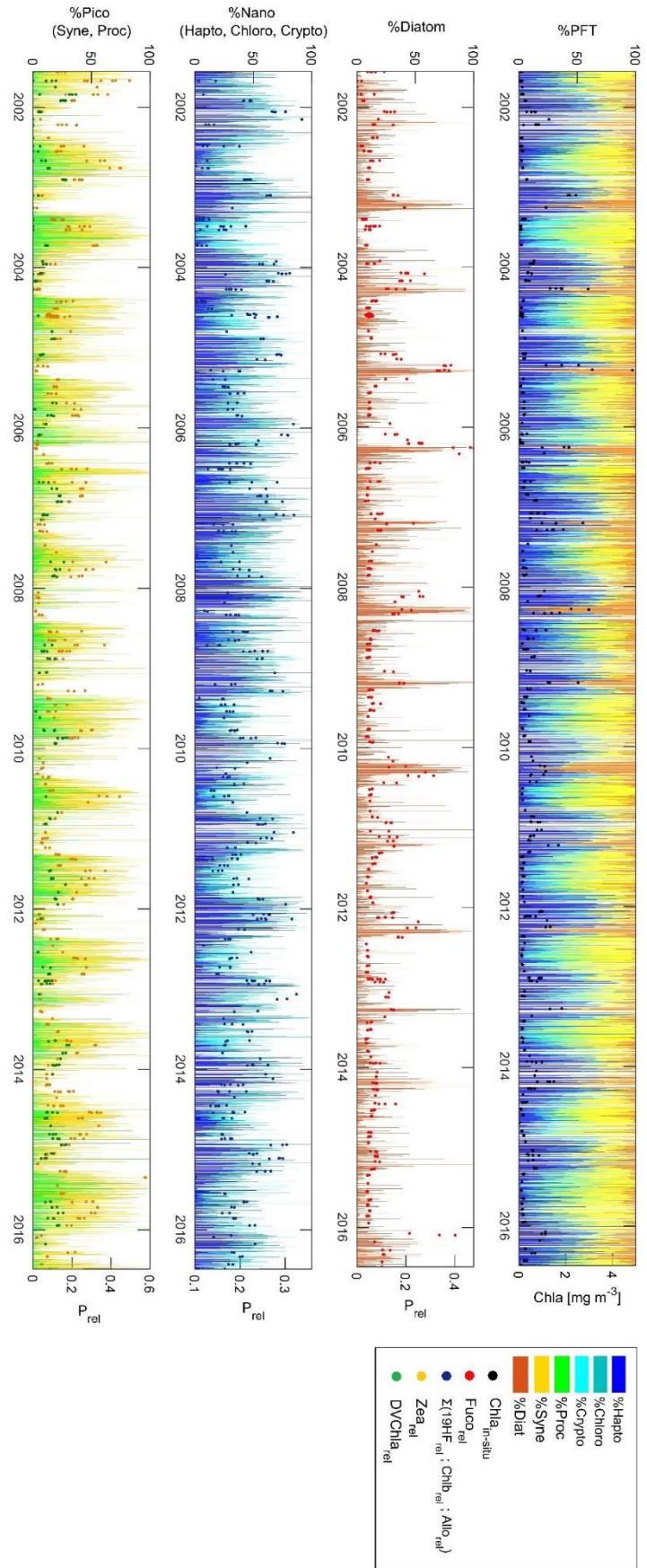


Figure A2.7: Daily percentage of dominant PFT computed from satellite data around the Boussole Station, Ligurian Sea, from 2001 till 2016. a) representation of the 6 PFT Classes along with the in-situ Chla (black dots), b) Diatoms (red) frequency overlapped with in-situ Fuco rel (red dots), c) Hapto (navy blue), Chloro (blue) and Crypto (cyan) frequency along with the corresponding in-situ P rel (19HF rel, Chlb rel, Allo rel) (blue dots), d) Synechococcus (yellow) and Prochlorococcus (green), overlapped with in-situ Zea rel (yellow dots) and DVChla rel (green dots)

found by the SOM-PFT are in close agreement with the maxima in the concentrations of the pigment ratio for DVChla measured in situ. These results agree with the pattern reported by Vaultot *et al.*, (1990) and Marty *et al.*, (2002) estimated by flow cytometry and HPLC analysis, respectively. Also coinciding with the stratification period during summer, the dominant group is *Synechococcus*, and a maximum of Zea_{rel} is observed (Figure A2.7 (d)).

4.3. Distribution of dominant PFTs in the Mediterranean Sea

The Mediterranean Sea shows a well-marked spatio-temporal variability of the six PFT classes. Diatoms are mainly found in coastal zones, where a high occurrence of this class was found. Besides, a dominance of Diatoms was also found during April in the Balearic Sea, in association with the deep convection phenomena. Simultaneously, Nanophytoplankton

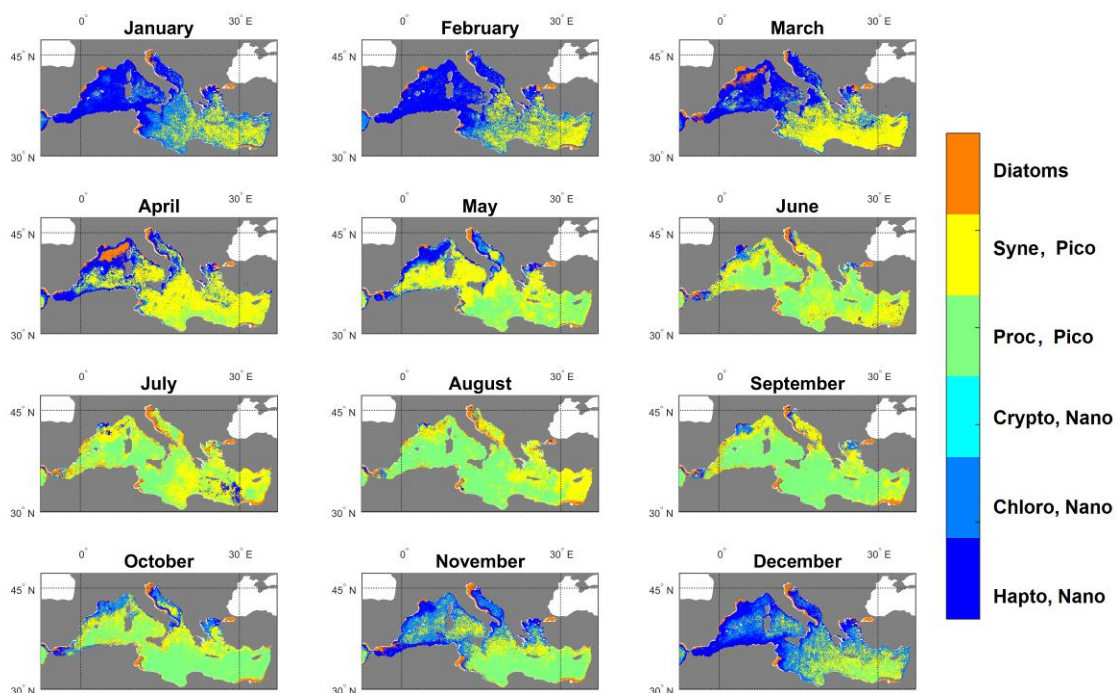


Figure A2.8: Monthly climatology images of the six dominant PFT classes identified using SOM-PFT

classes (Hapto and Chloro) dominate in the Mediterranean Sea during winter and early-spring season (December up to April). Meanwhile, the Pico size fraction (Syne and Proc) mostly dominate in the summer season. Figure A2.8 clearly indicates that the dominance of Diatoms, Hapto and Chloro in the Western basin is more frequent than in the Eastern basin, whereas Pico show an antagonist behavior with a quasi-dominance in the eastern Mediterranean.

Let us now consider the monthly patterns of the PFTs in each sub-basin (Figure A2.9). Diatoms present significant blooms in winter-spring in the Alboran and Balearic Seas. In the Adriatic Sea, Diatoms show dominance in mid-summer. Meanwhile in the open waters of the eastern Mediterranean, this class presents weak blooms.

Besides, the Nanophytoplankton classes (Hapto and Chloro classes) seem to co-vary, showing a pronounced abundance during the November-April period. In the Balearic Sea, a sharp increase of Diatoms is noted to the detriment of the nanophytoplankton classes in April,

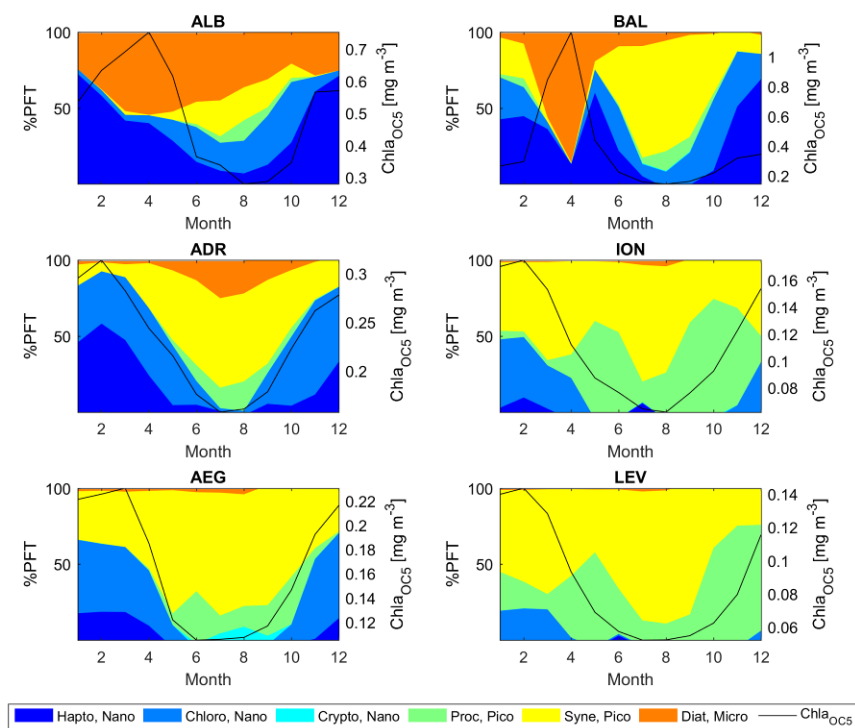


Figure A2.9: Monthly Climatology of dominant PFT frequency alongside the Chla OC5 at the Alborian, Balearic, Adriatic, Ionian, Aegean and Levantine basins

which mainly is a consequence of the convection phenomena. In the western basin, Hapto and Chloro co-dominate, while in the eastern basin, Chloro tend to dominate the Nano size

fraction during the winter bloom. The Crypto class is absent in most of the basins, except in the Aegean Sea, where a bloom of Crypto is observed in August.

In parallel, the picophytoplankton represented by both *Synechococcus* and *Prochlorococcus* classes shows an antagonist monthly cycle. Unlike the other classes, Syne and Proc are the most abundant class during the summer season, mainly dominated by Syne. In the Adriatic and the Balearic Sea, the Syne dominated in July-September, while in the eastern part of the Mediterranean Sea, it quasi-dominated all year long. The Proc class is significantly present in the eastern Mediterranean, showing small antagonist variability in favor of the Syne.

5. Discussion

The present approach for identifying dominant phytoplankton groups from satellite data, which is based on the use of the SOMs, is innovative and gave good performances in the validation tests. The reconstruction of the temporal variability of the PFTs at the Boussole station clearly shows that our approach efficiently reproduces the annual variability of the PFTs at least in the Ligurian Sea.

Comparing the PFTs given by the present method to those obtained by processing in situ data with the identification criteria proposed by Alvain *et al.*, (2005), we found some differences in identifying Nano and Syne classes. These errors which are of 23% and 13% respectively can be associated with cumulative uncertainties during the process of the method; some errors can be due to the pigment estimation in the first step of the approach when using the SOM-Pigments (El Hourany *et al.*, 2019), where 19HF or Zea concentrations can be mis-estimated with high uncertainties. The use of SOM-Pigment highly depends on the quality of the satellite data (Globcolour and AVHRR) and therefore controls the pigment estimation errors. Other causes may interfere also, such as uncertainties in the thresholds of Alvain *et al.*, (2005) and the classification criteria via SOM-PFT. Meanwhile, the SOM-PFT did not detect any Perid predominance, due to the low values of these pigments compared to that of 19HF or Zea. Therefore, we were unable to detect any dominance of Dinoflagellates in the Mediterranean Sea. In parallel, the Crypto and Chloro classes were not validated using Alvain *et al.*, (2005) thresholds. Yet, an undeniable coincidence of 44.5% is noted while assigning these two classes by SOM-PFT to the in-situ samples that were unidentifiable via the thresholds of Alvain *et al.*, (2005).

Navarro *et al.*, (2014; 2017) developed a regionalized version of PHYSAT (PHYSAT-Med) which is based on the OC-CCI merged ocean color data. Originally, PHYSAT identifies dominant phytoplankton types using normalized reflectance anomalies. PHYSAT-Med was validated using the Mediterranean in situ HPLC dataset and the threshold classification method of Alvain *et al.*, (2005). The results of this validation gave an agreement of 74% for Nano and 60% for Syne with the in situ matchups (Navarro *et al.*, 2014). Our results showed a close agreement with Physat-Med for dominant Nano identification (70%) and an improvement of the Syne identification up to 76%.

5.1. Phytoplankton dynamic in the Mediterranean Sea

The oligotrophic state of the Mediterranean has been recognized for a long time (Sournia, 1973). A general decreasing gradient from west to east is typically observed in ocean color data of Chla (Bosc, Bricaud and Antoine, 2004; Barale, Jaquet and Ndiaye, 2008), to reach an ultraoligotrophic environment in the most eastern basin, the Levantine basin.

In this study, the detected winter bloom of Nano PFT classes (Hapto and Chloro classes) in most of the Mediterranean open water is induced by the nutrient availability generated by the water mixing in winter (Marty *et al.*, 2002; Uitz *et al.*, 2012). Such water mixing is due to the decrease of the SST and the resulting homogenization of the water column (Krom, Herut and Mantoura, 2004; Krom *et al.*, 2005; Krom, Emeis and Van Cappellen, 2010). The bloom stops shortly after the surface waters run out of nutrient.

Major hydrodynamic events play important roles in the mechanism of the bloom. In our study, the dominance of Diatoms was observed in the northernmost part of the western Mediterranean basin. This region is characterized by the occurrence of an annual spring bloom of big phytoplankton triggered by nutrient enrichment of the euphotic zone due to deep winter convection events which mix the whole water column (Marty and Chiaverini, 2010). In addition to this pattern, the dominance of the Hapto and Chloro classes were also observed in this study all over the western basin in the winter-spring period. These waters are characterized by complex physical and chemical processes, such as water mass circulation, mesoscale hydrological structures such as anticyclonic eddies along the Algerian coast, and the intrusion of the Atlantic waters, which have the potential of alleviating the general nutrient limitation (Siokou-Frangou *et al.*, 2010).

According to our findings, several eutrophic regions were marked by the recurrence of Diatom dominance, such as the continental shelf of the Gulf of Lyon in the South of France, the Gulf of Gabes in the Tunisian coast, the Gulf of Venice in the north of the Adriatic Sea, the Nile Delta in the Levantine basin, and the northern edge of the Aegean Sea. These regions are characterized by their nutrient input originating from several sources such as the presence of a large and shallow continental shelf where the majority of the oceanic burial of organic carbon occurs and nutrients can be put in suspension by small scale hydrodynamic phenomena (Premuzic *et al.*, 1982; Hedges and Keil, 1995). Anthropogenic activities and riverine inputs, such as the Ebro (Eastern Spanish coast), the Rhone (Gulf of Lyon), the Po (Gulf of Venice) and the Nile (North of Egypt) also participate to nutrient enrichment. In such nutrient rich zones, the diatoms mainly develop due to the presence of essential elements for their growth including the silicon that is used to build their silica frustules. Unlike other minerals, the requirement for silicon is unique to diatoms. In the open ocean, the diatom bloom is typically ended by a shortage of silicon (Egge and Aksnes, 1992).

Moreover, the observed dynamic of the *Prochlorococcus* and *Synechococcus* classes is mainly due to their metabolic capacities, which make that these Pico are able to bloom in the summer season and then accordingly recycle dissolved organic matter in a very efficient manner under competitive circumstances such as limiting nutrients and predators (Sieburth, Smetacek and Lenz, 1978). Furthermore, the abundance of Pico reaches a maximum in summer when the Chla concentration are at their lowest values. This fact can be explained by the lower photosynthetic pigment content in the Pico cells under higher irradiances and longer daylight periods during summer (Calvo-Díaz, Morán and Suárez, 2008). During this season, primary production due to Pico exhibits a maximum value (Uitz *et al.*, 2012). It is well known that thanks to their high surface/volume ratio, *Synechococcus* and *Prochlorococcus* can cope optimally with nutrients-impooverished environments (Le Quéré *et al.*, 2005). The presence of *Synechococcus* in the Levantine basin has been widely reported as a dominant group where ultra-oligotrophic conditions are present and particularly during summer (Siokou-Frangou *et al.*, 2010; Uitz *et al.*, 2012). The spatio-temporal variability of the dominant PFTs in the Mediterranean Sea observed in this study highlights the dominance of Hapto and Chloro in winter-spring seasons essentially in the Western basin, whereas *Synechococcus* dominate most of the Mediterranean waters during summer. These results

agree with previous studies on the Mediterranean Sea e.g: Vidussi *et al.*, (2000; 2001); Uitz *et al.*, (2012); Samartino *et al.*, (2015); Navarro *et al.*, (2017).

6. Conclusion

The clustering method we have developed was very efficient to identify the PFT from satellite measurements and to reconstruct the phytoplankton variability with significant performances. The approach was based on the optical characteristics of phytoplankton pigments and on the efficiency of the SOMs to make robust classifications. It permitted to retrieve the regional distribution of the PFT assemblage in the Mediterranean Sea and to estimate their pigment composition, which is innovative compared to most of the other methods inverting satellite observations. These PFTs showed a well-marked variability due to the complex physical and biogeochemical environment: Nanophytoplankton such as Haptophytes and Chlorophytes mainly dominated during winter in the Western Mediterranean basin, while *Synechococcus* and *Prochlorococcus* dominated during summer. The dominance of Diatoms was mainly observed in spring in the Balearic Sea in response to deep-water convection phenomena and near the coastline and estuaries, due to important continental inputs and river discharges. Besides we were able to observe the dynamics of several new phytoplankton types in the Mediterranean Sea, such as Chlorophytes and Cryptophytes that were poorly documented so far. SOMs may highlight other important phytoplankton types.

The validation tests, which were performed on in situ matchups, showed satisfying performances and proved that our approach was suitable to study the phytoplankton diversity in the Mediterranean Sea. This work shows that the use of fine clustering methods such as the SOM chain we developed (SOM-Pigments → SOM-PFT+HAC), can accurately retrieve phytoplankton pigments from satellite observations, then convert pigment assemblages into well-defined phytoplankton communities, which can be identified to PFTs. The method is generic and could be applied to other oceanic regions.

Partie 2 : Reconstruction de la variabilité spatio-temporelle des PSC au niveau la Méditerranée.

1. Introduction

Les différences hydrologiques dans le bassin méditerranéen entraînent une augmentation de l'oligotrophie lors que l'on va de l'ouest vers l'est. Il en résulte une diminution de la concentration de Chla de surface de l'ouest vers l'est (Turley *et al.*, 2000) observée depuis l'espace (Antoine *et al.*, 1995), les eaux du bassin Levantin de la Méditerranée orientale présentant des conditions très oligotrophes (Dowidar, 1984; Abdel-Moati, 1990; Yacobi *et al.*, 1995; Krom, Herut and Mantoura, 2004; Krom *et al.*, 2005; Krom, Emeis and Van Cappellen, 2010).

Cependant, l'évolution saisonnière de la répartition de Chla suit toujours le cycle typique des régions tempérées, caractérisée par une augmentation de l'abondance du phytoplancton à la fin de l'hiver / au début du printemps, une diminution lors de la saison estivale et une deuxième prolifération moins intense de phytoplancton en automne (Siokou-Frangou *et al.*, 2010; Sammartino *et al.*, 2015).

La structure de la communauté phytoplanctonique dans les zones oligotrophes de l'océan mondial est principalement composée de picoplancton (Li *et al.*, 1993; Brunet *et al.*, 2006; Dandonneau *et al.*, 2006). Néanmoins, la structure des communautés phytoplanctoniques méditerranéennes révèle une grande variabilité tant à l'échelle spatiale que temporelle, ainsi que de grandes disparités dans la composition des assemblages de phytoplancton (Siokou-Frangou *et al.*, 2010). De nombreuses études ont mis en évidence la dominance du picoplancton dans la mer Méditerranée, conséquence de son oligotrophie. Cependant la prolifération du phytoplancton dans certaines régions entraîne la coexistence de nombreux groupes de microalgues (Siokou-Frangou *et al.*, 2010). Le modèle empirique satellitaire appliqué par Sammartino *et al.*, (2015) englobe cette structure de communauté inhabituelle et complexe en mer Méditerranée et a permis d'évaluer la variabilité spatio-temporelle des trois PSC (micro, nano et pico-plancton) pendant toute la période de SeaWiFS (1998-2010).

Dans cette partie de notre étude, les différentes classes de PSC ont été estimées via l'application de l'algorithme de SOM-Pigments à la Méditerranée. Des équations simplifiées

(Dandonneaux *et al.*, 2004) ont été utilisées pour passer de la composition pigmentaire au pourcentage en PSC. Ensuite, une tentative de comparaison a été effectuée entre les différents résultats avec des longues séries temporelles de comptage de phytoplancton in-situ, tant du point de vue de la composition pigmentaire, pourcentage de PSC et comptage phytoplanctonique in-situ.

2. Matériels et méthodes

Etant données les bonnes performances de la validation de la SOM-Pigments vis-à-vis des données HPLC in-situ en Méditerranées, on a pu générer avec SOM-Pigments une nouvelle base de données de concentrations des pigments secondaires à la fréquence journalière à partir des données satellitaires GlobColour.

Les différents classes de tailles de phytoplancton sont estimées à partir des pigments par des équations donnant le pourcentage de Micro, Nano et Pico (Dandonneaux *et al.*, 2004):

$$\%Micro = (Fuco+Perid)/ \Sigma DP, \quad (1)$$

$$\%Nano = (19HF+19BF+Chlb+Allo)/ \Sigma DP, \quad (2)$$

$$\%Pico = (Zea+DVChlb)/ \Sigma DP, \quad (3)$$

$$\text{Où } \Sigma DP = Fuco+Perid+19HF+19BF+Chlb+Allo+ Zea+DVChlb \quad (4)$$

Ces équations ont été appliquées aux concentrations de pigment produites à l'aide de SOM-Pigments. Une climatologie mensuelle des trois PSCs a été réalisée à la fin de ce processus.

Par la suite, nous avons comparé les résultats avec deux longues séries temporelles de comptage de phytoplancton mesurés en deux stations distinctes, la première dans le Golfe de Naples, Italie (Sarno and Zingone, 2008), la deuxième au Sud du Royaume-Unis (Tarran, 2011). A la première station, seuls les Micro et Nano ont été identifiés et comptés étant donné le manque de Pico en Méditerranée. En admettant que la méthode est applicable en différentes zones du globe, nous avons eu recours à une série temporelle de Pico et de Nano mesurée dans la deuxième station (Tableau IV.1, Figure IV.1).

Une moyenne mobile a été appliquée sur ces séries temporelles pour minimiser l'amplitude des données « aberrantes ».

A ces deux stations, les 8 pigments diagnostiques (DP : Fuco, Peri, 19HF, 19BF, Chlb, Allo, Zea et DVChlb) ont été récupérés à la fréquence journalière depuis les images de pigments estimés par SOM-Pigments et sur une échelle spatiale de 3x3 pixels. Afin d'éliminer le bruit et de combler les données manquantes, une moyenne mobile de 9 jours a été appliquée sur les séries de pigments estimés. Les équations 1 à 4 ont été appliquées sur ces séries pour estimer le pourcentage de PSC au niveau des 2 stations. Les pourcentages de PSC sont relatifs ; pour cela nous avons exprimé les résultats des PSC estimés en terme de Chla relative à chaque PSC, soit :

$$Chla_{Micro} = Chla * \%Micro (5) ; Chla_{Nano} = Chla * \%Nano (6) ; Chla_{Pico} = Chla * \%Pico (7)$$

Table IV.1: Description des stations de mesure de phytoplancton

Station	Location	Type of Phytoplankton	Method	Frequency	N Obs	Evaluated Period	Ref
1	Gulf of Naples, Italy	Micro- and Nano-phytoplankton	Quantification using microscope	Weekly	438	1997-2006	(Sarno and Zingone, 2008)
2	Western English Channel	Nano- and Pico-phytoplankton	Flow Cytometer	Weekly	260	2007-2013	(Tarran, 2011)

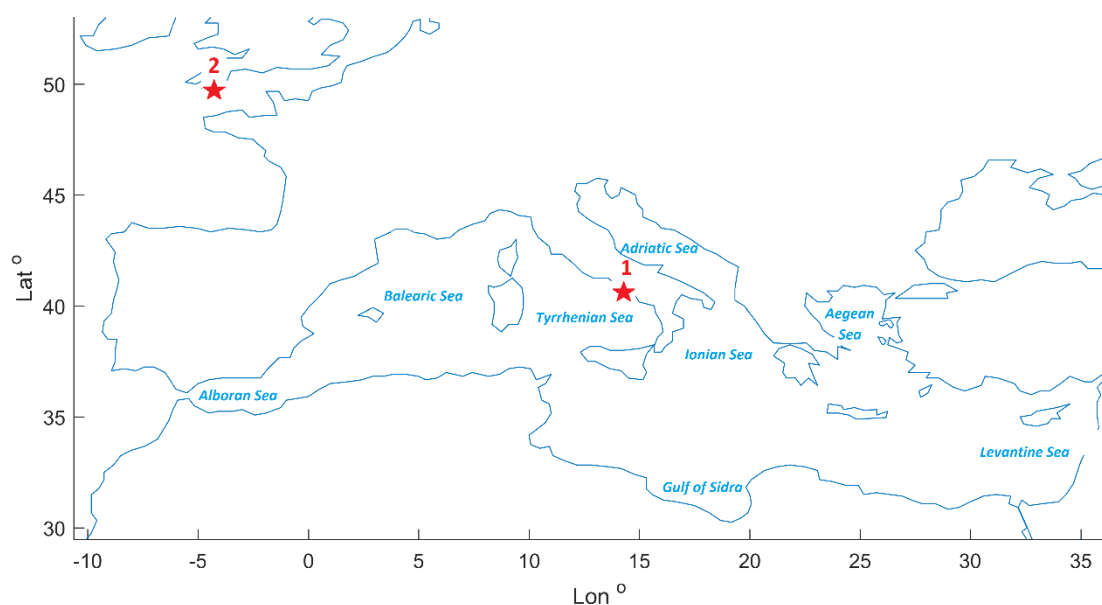


Figure IV.1 : localisation des stations à série temporelle de comptages de phytoplancton, 1) Gulf of Naples, 2) Western English Channel.

3. Résultats

3.1. Analyse de la variabilité des PSC

La Méditerranée présente une variabilité spatio-temporelle particulière des trois classes de taille : l'abondance du Micro va de 2% dans les eaux du large à 75% dans les zones eutrophes. Le Micro dénote une affinité pour les zones côtières eutrophes. En effet, le mécanisme du bloom du Micro semble varier d'un bassin à un autre ; en Méditerranée occidentale, le bloom de Micro est observé en Décembre-Mars, alors qu'en Méditerranée orientale, nous observons 2 blooms, le premier entre Décembre et Mars, et le deuxième pendant la période automnale. De plus, des pourcentages élevés de Micro sont observables dans la mer des Baléares qui est caractérisée par des phénomènes de convection profonde. Simultanément, le Nano domine en répartition spatiale la Méditerranée durant les mois d'hiver et de printemps avec des pourcentages fluctuants entre 30% et >80% durant le bloom. Par contre, le Pico domine en été. La figure IV.2 montre l'abondance importante des classes Micro et Nano dans le bassin occidental, contrairement au bassin oriental. Le Pico présente un comportement antagoniste, où des abondances importantes sont observées dans le bassin oriental et pendant l'été.

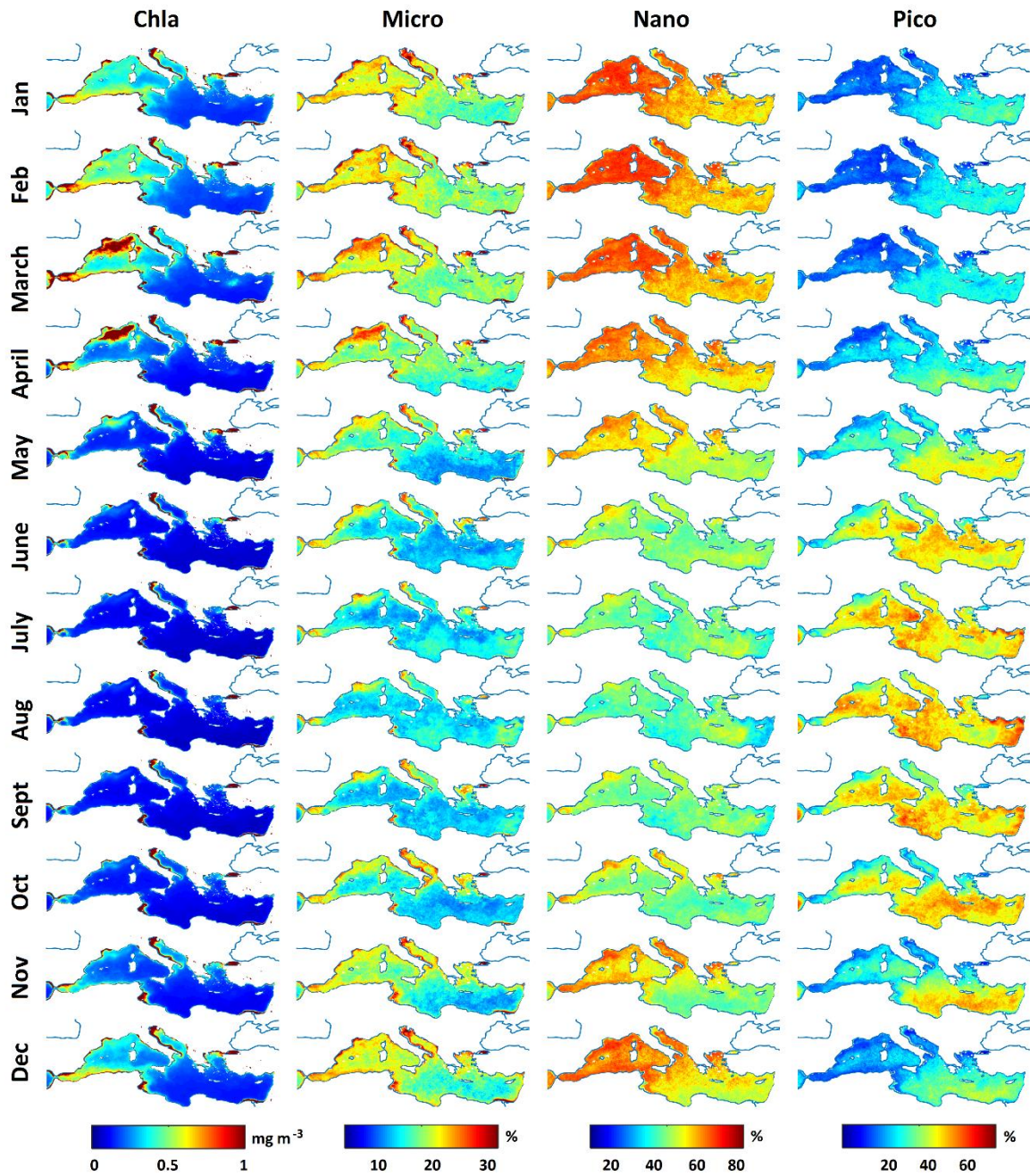


Figure IV.2: Climatologies mensuelles des 3 classes de taille de phytoplancton et la Chla GlobColour

En examinant le cycle des PSC dans chaque sous-bassin, nous avons observé la même variabilité spatio-temporelle dans les données climatologiques (Figure IV.3). En d'autres termes, la classe Micro a montré un cycle unimodal dans les mers d'Alboran, des Baléares, de la Tyrrhénienne, de l'Adriatique, de la mer Ionienne et de la mer Égée, avec une floraison observée durant les mois d'hiver-printemps. Dans le même temps, le MICRO présentait un cycle bimodal dans le bassin Levantin et dans le golfe de Sidra, avec une première floraison en hiver-printemps suivie d'une seconde en fin d'été. Les mers d'Alboran, des Baléares et l'Adriatique présentent le pourcentage le plus élevé de micro-organismes, atteignant un pourcentage d'environ 25% pendant la saison de floraison, tandis que le bassin levantin présente le pourcentage le plus faible, atteignant 17% au cours de la même saison. Nano et Pico ont montré les mêmes cycles opposés avec une variabilité spatiale bien marquée, où le Nano domine dans les eaux de la mer d'Alboran et des Baléares pendant toute l'année, tandis que tous les autres sous-bassins montrent une prédominance du Pico pendant la saison estivale. De plus, la dynamique entre Micro et Pico semble montrer un mécanisme intéressant. Dans les sous-bassins occidentaux, dans la mer d'Alboran, des Baléares, de la Tyrrhénienne et de l'Adriatique, la classe Micro a dominé le Pico lors de la floraison hivernale-printanière, avant que ce dernier ne prenne la tête en été. Pendant ce temps, dans les sous-bassins est, les mers Ionienne, Égée et Levantine, le Pico domine le Micro tout au long de l'année.

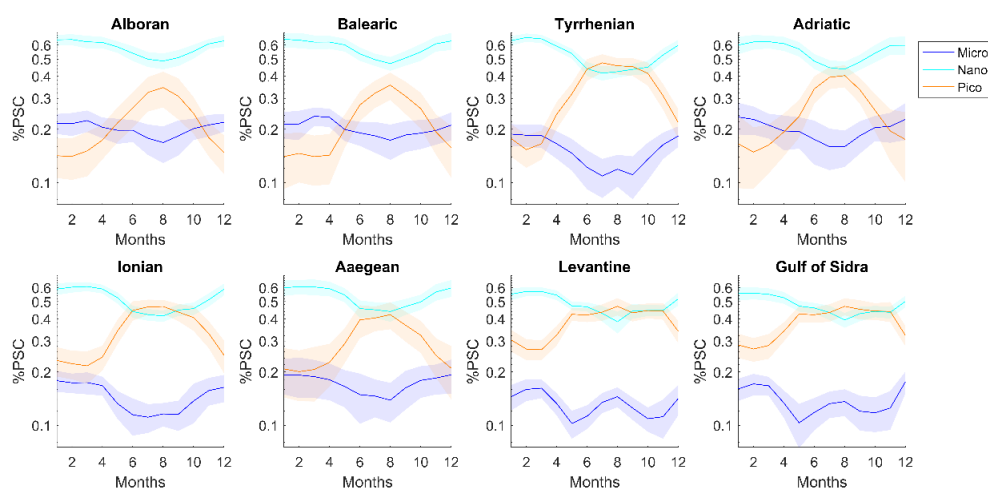


Figure IV.3: Monthly climatology of the 3 estimated PSC's percentage for each of the Mediterranean sub-basins. The full line represents the climatology mean with a

3.2. Comparaison avec des séries temporelles de comptage de phytoplancton

Selon les équations 4, 5 et 6, les Micro sont représentés par la somme Fuco + Perid, le Nano avec la somme Chlb + 19HF + 19BF + Allo, et le Pico par la somme Zea + DVChlb. Les figures IV.4, IV.5, IV.6 et IV.7 montrent les comptages in-situ de micro- et de nano-phytoplancton ainsi que les DP correspondants à chacune des 2 stations. Pour la station du Golfe de Naples, les comptages Micro vs. [Perid et Fuco] dénotent un coefficient de corrélation de Spearman (R_s) significatif de 0.63 ($pval < 0.0001$) et un R_s de 0.51 ($pval < 0,0001$) a également été noté pour les comptages Nano vs. [Chlb 19HF, 19BF et Allo]. Et à la station de la Manche, un R_s de 0.36 ($pval < 0,0001$) a été noté pour les comptages de nano vs [Chlb 19HF, 19BF et Allo] et un R_s de 0,49 ($pval < 0,0001$) pour les comptages de *Synechococcus* vs. [Zea].

Après avoir estimé la part de chaque PSC vis-à-vis de la Chla suivant les équations 8, 9 et 10, il en est résulté une série chronologique périodique avec un cycle annuel bien défini. Le R_s le plus élevé a été observé à la station du golfe de Naples avec un R_s significatif de 0.72 pour les comptages de nano par rapport à $Chla_{Nano}$ et de 0,69 pour les comptages de Micro par rapport à $Chla_{Micro}$. Enfin, la station de la Manche suit avec une corrélation modérée, avec un R_s significatif de 0.66 pour les comptages *Synechococcus* vs. $Chla_{Pico}$ et de 0.53 pour les comptages Nano vs. $Chla_{Nano}$ correspondant (Tableau IV.2).

En comparant les résultats de corrélation des comptages de phytoplancton vs. Chla OC5 avec ceux de $Chla_{PSC}$ à chacune des stations, les $Chla_{Micro}$ et $Chla_{Nano}$ ont montré une amélioration de 1 à 3% des valeurs de R_s par rapport à la Chla OC5. Pour le *Synechococcus* à la station de la Manche, $Chla_{Pico}$ vs Syne a montré une amélioration de 40% de la corrélation par rapport au Chla OC5 vs Syne.

Nous pouvons noter que l'utilisation des DP dans l'estimation de la contribution de chaque PSC à la Chla a introduit une information spécifique relative à la dynamique temporelle de PSC. Cette constatation est bien mise en évidence par les résultats de la contribution de *Synechococcus* au Chla, tout en extrayant efficacement le cycle. En outre, ce résultat aide également à indiquer que le Nano et le Micro ont été les principaux contributeurs à la Chla dans ces 2 stations, montrant de fortes corrélations avec $Chla_{OC5}$ et $Chla_{PSC}$.

Table IV.2: Resultats des correlations Spearman entre les differents DP, Chl_{PSC} estimées et $Chla$ Satellitaire vs. comptage in-situ de phytoplancton sur les differentes stations

Station	PSC	R_s DP vs. #PSC	R_s $Chla_{PSC}$ vs. #PSC	R_s $Chla_{Sat}$ vs #PSC	pval
1	Micro	0.63	0.69	0.66	<0.001
	Nano	0.51	0.72	0.69	<0.001
2	Nano	0.36	0.53	0.54	<0.001
	Pico	0.49	0.66	0.26	<0.001

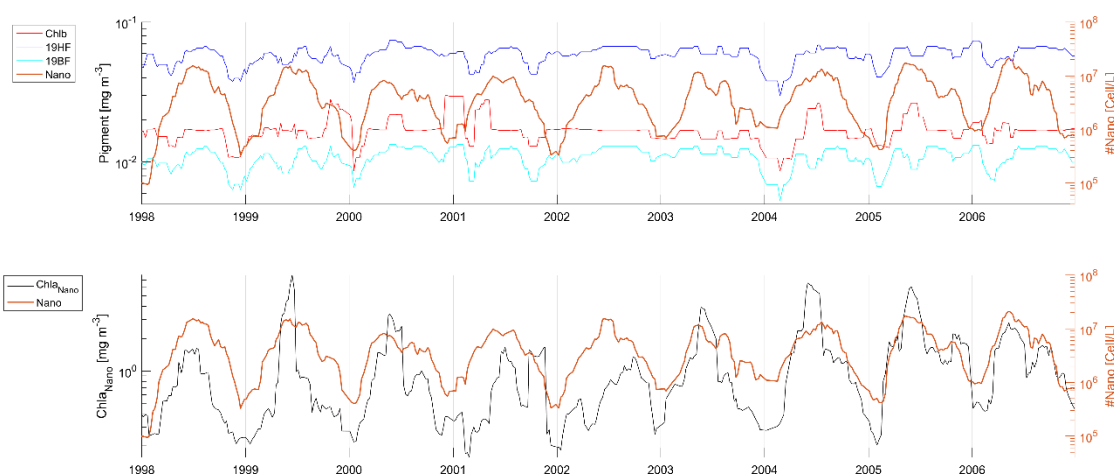


Figure IV.4: Série temporelle de comptage in-situ de Nano, vis-à-vis des différents pigments secondaires (a) et de la $Chla$ Nano (b) au niveau de la station du Golfe de Naples.

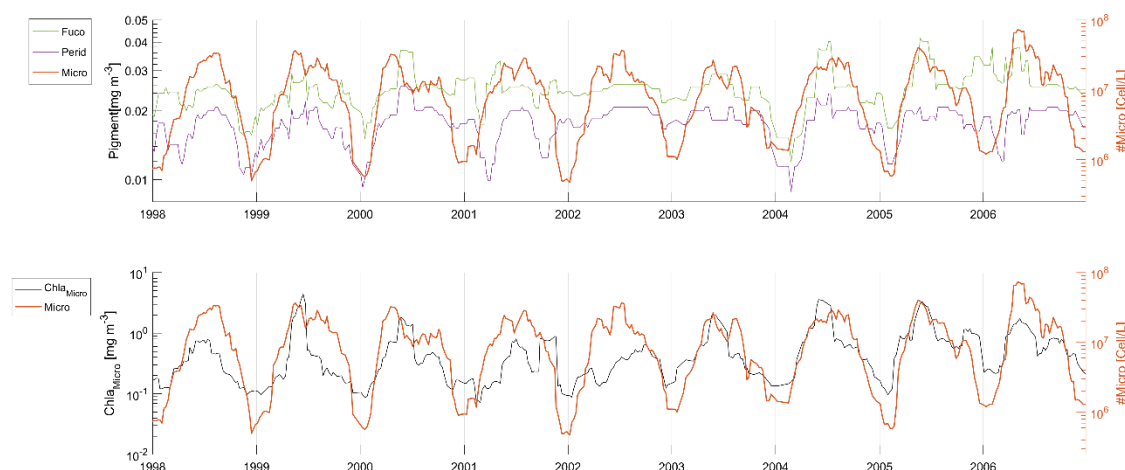


Figure IV.5: Série temporelle de comptage in-situ de Micro, vis-à-vis des différents pigments secondaires (a) et de la $Chla$ Micro (b) au niveau de la station du Golfe de Naples.

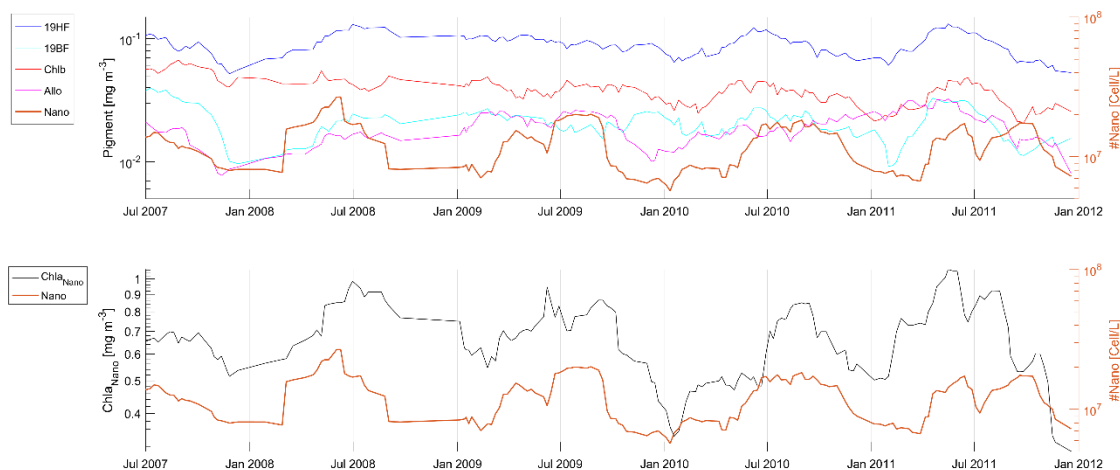


Figure IV.6: Série temporelle de comptage in-situ de Nano, vis-à-vis des différents pigments secondaires (a) et de la Chla Nano (b) au niveau de la station de la Manche.

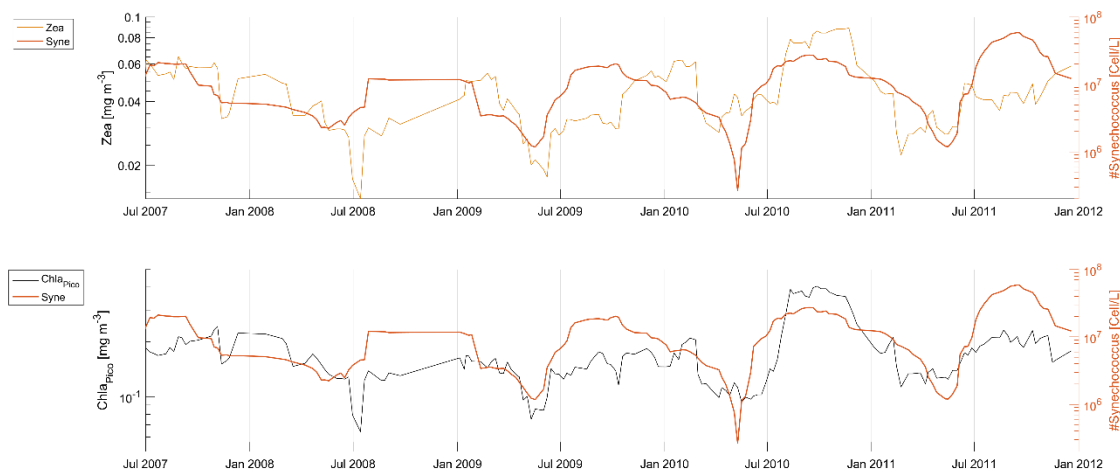


Figure IV.7: Série temporelle de comptage in-situ de Synechococcus, vis-à-vis des différents pigments secondaires (a) et de la Chla Pico (b) au niveau de la station de la Manche.

Afin de mieux caractériser la relation entre Chla_{PSC} estimées et le comptage de phytoplancton, nous avons appliqué une régression linéaire sur les données (en log10) de chaque station (Figure IV.8 et Figure IV.9). Les coefficients de régression observés varient entre 0.24 et 0.52 (Tableau IV.3). Cependant, les pentes observées sont tous significativement positives fluctuant entre 0.1 et 0.33. La variabilité de ces coefficients est due en premier lieu au type de PSC et à la station étudiée.

Les valeurs de la Chla_{PSC} après transformation logarithmique, tendent à avoir une distribution normale au niveau des 2 stations. Mais, ceci ne peut pas être généralisé sur les données de

comptage correspondantes, et cela prouve la non-linéarité de la relation entre ces composantes.

Et si nous observons les valeurs de la Chla simultanément avec les 'scatter-plots' de $Chla_{PSC}$ vs. comptage de phytoplancton, nous remarquons que les classes de Micro et Nano suivent un gradient ascendant de Chla, alors que le Pico ne présente aucune relation. Ce qui nous amène à dire que les Nano et Micro sont les contributeurs majeurs à la Chla du milieu.

Tableau IV.3: résultats de la régression log-linéaire entre $Chla_{PSC}$ et Chla Satellitaire vs. comptage in-situ au niveau des différentes stations

Station	N Obs	PSC	Intercept	Slope	R ² #PSC vs. Chla PSC	R ² #PSC vs. Chla OC5
1	438	Micro	-2.70	0.33*	0.46	0.41
		Nano	-1.29	0.17*	0.47	0.41
2	260	Nano	-1.6	0.23*	0.30	0.30
		Pico	-2.22	0.19*	0.38	0.05

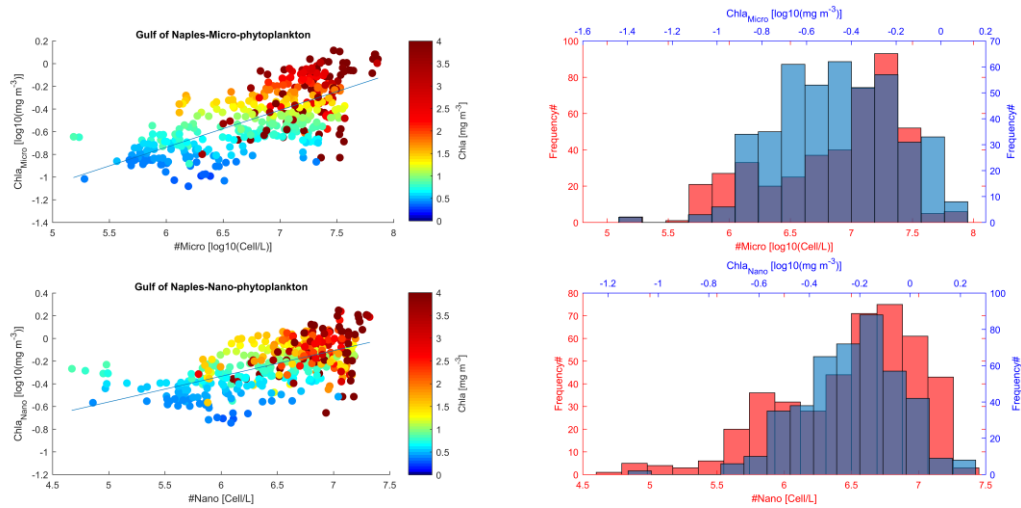


Figure IV.8: Diagramme de dispersion des comptages in-situ de Micro et de Nano vs $Chla_{PSC}$ au niveau de la station de Golfe de Naples.

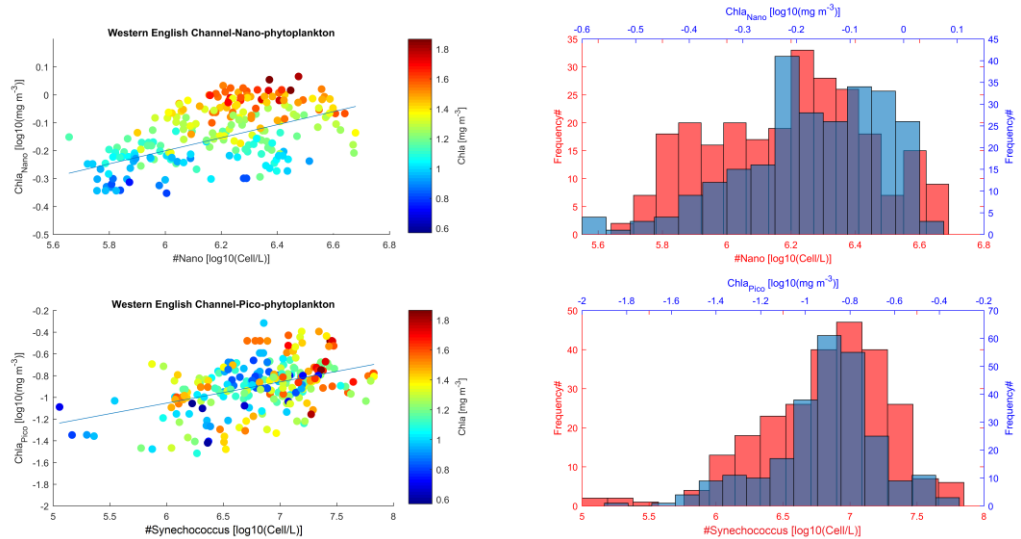


Figure IV.9: Diagramme de dispersion des comptages in-situ de Nano et Pico vs $Chla_{PSC}$ au niveau de la station 2.

4. Discussion

4.1. La dynamique du phytoplancton dans la Méditerranée:

La Méditerranée est caractérisée par son état oligotrophe (Sournia, 1973). Un gradient négatif est observé de l'ouest vers l'est sur les images de Chla de couleur de l'océan (Bosc, Bricaud and Antoine, 2004; Barale, Jaquet and Ndiaye, 2008), pour aboutir à un environnement ultraoligotrophe dans le bassin Levantin.

Dans cette étude, les blooms hivernaux de Micro et Nano dans la majorité des eaux du large de la Méditerranée sont contrôlés par la présence des nutriments résultant du mélange de la colonne d'eau en hiver (Marty *et al.*, 2002; Uitz *et al.*, 2012). Ce mélange résulte de l'homogénéisation de la colonne d'eau en densité due aux pertes de chaleur provenant des échanges océan-atmosphère intenses en hiver (Krom, Herut and Mantoura, 2004; Krom *et al.*, 2005; Krom, Emeis and Van Cappellen, 2010). Le bloom cesse dès l'épuisement des nutriments en surface. Cependant, le bloom de Micro observé dans cette étude en automne au niveau du bassin oriental peut être expliqué par la régénération des nutriments après installation de la thermocline saisonnière profonde. En réponse à cet événement, les groupes de phytoplancton opportunistes et thermophiles tels que les Dinoflagellés (Micro) prolifèrent.

En Méditerranée occidentale, les processus hydrodynamiques jouent un rôle important dans la régulation des blooms. Des pourcentages élevés de Micro sont observés dans les eaux au large du Golfe de Lion. Cette région est caractérisée par l'occurrence de phénomènes de convection profonde en hiver, engendrant la prolifération des classes de tailles relativement grandes, tel que les Micro, en réponse à l'enrichissement du milieu en nutriments (Marty and Chiaverini, 2010).

De plus, plusieurs régions eutrophes sont caractérisées par la présence d'une grande abondance de Micro ; on peut citer le plateau continental du Golfe de Lion dans le Nord de la Méditerranée, le Golfe de Gabes, en Tunisie, le Golfe de Vénice au nord de l'Adriatique et le Delta du Nile dans le bassin Levantin. Ces régions sont caractérisées par un apport important de nutriment provenant de plusieurs sources, tel que la présence d'un plateau continental, siège de la reminéralisations de la matière organique (Premuzic *et al.*, 1982; Hedges and Keil, 1995), les activités anthropiques et les apports terrigènes des grands fleuves comme l'Ebro

(Côte Est de l'Espagne), le Rhône (Golfe du Lion), le Po (Golfe de Vénice) et le Nil (Nord de l'Égypte).

Dans les eaux riches en nutriment, ce sont les Diatomées (Micro) qui majoritairement prolifèrent grâce à la présence d'un élément essentiel à leur croissance : la silice. La présence de silice est particulière aux Diatomées, pour la construction de leurs thèques siliceuses. Dans les eaux du large, un bloom de Diatomées cesse une fois la silice est épuisée (Egge and Aksnes, 1992).

La dynamique antagoniste du Pico observé dans cette étude par rapport à celle du Micro et du Nano peut être expliquée par leur capacité autotrophe et hétérotrophe, qui permet au Pico de proliférer pendant la saison estivale et de recycler la matière organique dissoute d'une manière efficace et sous le stress de la limitation en nutriments (Sieburth, Smetacek and Lenz, 1978; Le Quéré *et al.*, 2005). Dans le bassin Levantin où les conditions oligotrophes règnent et particulièrement en été, les *Synechococcus* (Pico) ont été toujours observés comme groupe dominant, (Siokou-Frangou *et al.*, 2010; Uitz *et al.*, 2012).

Nos résultats sont en accord avec les études précédentes faites sur la Méditerranée (Vidussi, Marty and Chiave, 2000; Vidussi *et al.*, 2001; Uitz *et al.*, 2012; Navarro *et al.*, 2014, 2017; Sammartino *et al.*, 2015; Di Cicco *et al.*, 2017). Cela suggère que notre approche d'étude de PSC est bien adaptée à l'observation de la variabilité du phytoplancton à l'échelle du bassin méditerranéen et de ses sous-bassins.

4.2. Evaluation de la méthodologie:

La SOM-Pigments a montré une performance tout à fait satisfaisante lors de la cross-validation. Les pigments estimés nous ont permis d'obtenir les trois classes de taille de phytoplancton ainsi que leur variabilité. La comparaison avec des données réelles de comptage *in-situ* a été apport primordial quant au bien fondée de notre méthode.

Aux trois stations étudiées, la $Chla_{PSC}$ était en phase avec le comptage de phytoplancton, valorisé par les corrélations de Spearman. De plus, la pente positive que nous avons eue au niveau de la régression linéaire entre $Chla_{PSC}$ et comptage *in-situ* nous indique que la relation entre ces deux composantes est illustrée par l'augmentation de la $Chla_{PSC}$ avec l'augmentation du comptage. Simultanément, l'observation de la $Chla$ superposée sur les scatter-plots des

Chla_{PSC} vs. Comptage in-situ fait preuve que les Micro et Nano contribuent majoritairement au total du contenu en Chla. Mais la dispersion de la Chla superposée sur le scatter-plot de Pico nous informe de la contribution minoritaire de cette classe à la Chla total. Cela peut être mis en valeur par le fait que le Pico arrive à son maximum durant l'été, là où la Chla est à son minimum. Cet événement peut être expliqué par la réduction du contenu en pigment photosynthétique dans les cellules de Pico, sous des radiations lumineuses élevées, et une durée d'exposition à ces radiations plus longue pendant l'été (Calvo-Díaz, Morán and Suárez, 2008).

Par conséquent, nous pouvons dire que l'utilisation de la SOM, et l'estimation des PSC à travers les pigments diagnostiques reproduit efficacement le cycle saisonnier de chaque PSC. Néanmoins, l'observation des distributions et le coefficient de régression entre la Chla_{PSC} et les comptages *in-situ* de phytoplancton confirment que la relation n'est pas aussi simple que ça ; un maximum de Chla_{PSC} n'implique pas directement la présence d'un maximum de comptage de phytoplancton. Cette relation est sujette à plusieurs incertitudes de plusieurs origines :

Cependant, de nombreuses incertitudes cumulées peuvent affecter les performances du SOM:

- Incertitudes au niveau de Chla / Rrs

Les erreurs sur l'estimation de Chla et des Rrs peuvent également être prises en compte dans l'application d'algorithmes standard et de corrections atmosphériques standard (Kowalczyk *et al.*, 2010), qui sont principalement développés à l'aide d'une base de données globale in-situ, réduisant ainsi la spécificité de l'algorithme. De plus, dans certains cas d'eaux troubles, le CDOM élevé et la turbidité influencent l'estimation de la Chla et conduisent à une surestimation des valeurs de Chla.

- « Package Effect »

Les propriétés d'absorption et de diffusion du phytoplancton sont importantes affectant directement les calculs des Rrs (via le rapport rétrodiffusion sur absorption). Ces propriétés optiques du phytoplancton sont modulées par la morphologie cellulaire et la teneur en pigment. En d'autres termes, l'absorption par le phytoplancton n'est pas une simple somme

des coefficients d'absorption de pigments individuels. Les spectres d'absorption du phytoplancton varient en ampleur et en forme en raison de la composition différente des pigments intracellulaire et de son état physiologique (Bidigare *et al.*, 1990; Hoepffner and Sathyendranath, 1991; Bricaud, Ciotti and Gentili, 2012); En outre, l'augmentation de la concentration en pigment cellulaire et de la taille des cellules aplatira les spectres d'absorption spécifiques (Morel and Bricaud, 1981). Par ailleurs, les coefficients de diffusion et de rétrodiffusion du phytoplancton dépendent fortement de la taille, de la forme et de l'indice de réfraction de la cellule de phytoplancton (Volten *et al.*, 1998; Witkowski *et al.*, 1998; Jonasz and Fournier, 2007; Sullivan and Twardowski, 2009) et affectant l'intensité du signal Rrs. Pour cela, avec une telle influence sur le signal Rrs, on peut suspecter la fiabilité et la représentativité des informations de phytoplancton contenues dans ce signal et donc affecter la récupération des informations de Chla du capteur.

- Incertitudes dans l'estimation des pigments

La performance de la SOM dépend de la qualité des données utilisées pour la phase de calibration. Les données doivent être représentatives de la situation dans le temps et dans l'espace. Les incertitudes concernant la qualité des données satellitaires et les mesures HPLC impliquent des incertitudes cumulées supplémentaires dans l'estimation des pigments.

- Incertitudes liées à l'application DPA

L'application de l'analyse des pigments diagnostiques DPA a révélé de manière satisfaisante la variabilité des groupes de phytoplancton. Par ailleurs, l'utilisation des DP pour estimer l'abondance du phytoplancton n'est pas toujours précise lorsqu'un certain DP peut être trouvé dans plusieurs groupes de phytoplancton, ce qui peut donner lieu à une interprétation erronée. Ainsi, les incertitudes sur les pourcentages de PSC peuvent entraîner une surestimation ou une sous-estimation de la valeur $Chla_{PSC}$ relative et, par conséquent, altérer la relation avec les comptages in-situ.

5. Conclusion

L'utilisation des cartes auto-organisatrices s'est avérée être un outil fiable pour la reconstruction de la variabilité spatio-temporelle du phytoplancton avec une performance significative. La disponibilité des pigments à une haute fréquence temporelle et sur une vaste

échelle spatiale a permis l'estimation des pourcentages de PSC qui présentent des patterns géographiques cohérents.

L'utilisation d'une base de données globale de mesure HPLC, complémentaires aux données satellitaires de GlobColour était basée sur les propriétés optiques, et cela a permis une estimation efficace de la distribution régionale des phytoplanctons en termes de classe de taille dans la mer Méditerranée. Ces classes ont présentés une variabilité particulière vue la complexité de l'environnement quant aux facteurs biogéochimiques et hydro-physiques. De plus, cette étude a évalué la relation entre PSC et comptage de PSC in-situ en utilisant deux longues séries temporelles. Des coefficients de corrélations de Spearman ont mis en évidence une bonne performance vis-à-vis la reconstruction des patterns saisonniers par la SOM-Pigments. En effet, des études de régression log-linéaires ont pu décrire des tendances positives entre la fraction de Chla PSC pour chaque classe de Micro, Nano et Pico-phytoplancton étudiées et les comptages in-situ de phytoplancton. Le coefficient de régression qui s'est avéré moyen ($0.3 < R^2 < 0.5$), dénote plusieurs limitations afin de mieux restituer la relation entre la fraction de Chla pour chaque PSC et l'abondance de ce dernier ; la qualité des observations satellitaires, les incertitudes sur les estimations des pigments secondaires par la SOM-Pigments, la physiologie du phytoplancton et les caractéristiques optiques.

Chapitre V : Régionalisation de la Méditerranée par la méthode de 2S-SOM

1. Introduction:

La régionalisation se définit par la délimitation de provinces dans lesquelles les conditions physiques, les propriétés chimiques et les communautés biologiques sont homogènes. Depuis les années 90s, la télédétection spatiale a permis de décrire la dynamique de la production primaire via la télédétection de la Chl_a phytoplanctonique pour délimiter des provinces à l'échelle d'un bassin (Sathyendranath *et al.*, 1995) ou dans l'océan global (Longhurst, 1998).

Les travaux de Longhurst, (1998) décrivent la Méditerranée comme une région spécifique, sans proposer de sous régions au sein du bassin malgré son hétérogénéité. Depuis, le système de classification biogéographique proposé par Spalding *et al.*, (2007), la Méditerranée est décrite comme une province tempérée appartenant au domaine Nord atlantique, et est subdivisée en sept écorégions : la mer Adriatique, la mer Egée, le Bassin Levantin, le Plateau tunisien/Gulf of Sidra, la mer Ionienne, la Méditerranée Occidentale et la Mer Alboran.

La régionalisation de la Méditerranée a été le sujet de très nombreuses études. La raison de ces nombreuses études est attribuée au fait que ce bassin est vu comme un océan miniature (Lejeusne *et al.*, 2010), et peut servir de modèle réduit pour la compréhension de la dynamique de l'océan global. En effet, des traits importants de l'océan global sont présents en Méditerranée mais avec une échelle spatiale et temporelle plus petite ce qui en facilite l'étude (Bethoux *et al.*, 1999), telle que la circulation thermohaline influencée par la formation d'eau dense. De plus, plusieurs caractéristiques de la Méditerranée (océan quasi-tropical l'été avec une thermocline bien formée, océan quasi polaire l'hiver avec des convections profondes, multiples bassins de caractéristiques différentes) en font un hotspot de la biodiversité marine (Coll *et al.*, 2010) mais vulnérable au changement climatique, d'où l'importance d'une vision de conservation de ce mécosystème.

Les études effectuées sont souvent issues de méthodes de classification appliquées à des paramètres décrivant les eaux du large. Nous en rappelons quelques unes des plus récentes.

D'Ortenzio and Ribera d'Alcalà, (2008), Palmiéri, (2004), et Mayot *et al.*, (2016) se sont basés sur la phénologie du phytoplancton ; (Nieblas *et al.*, 2014) and (Reygondeau *et al.*, 2017) ont utilisé des variables géochimiques comme la température, la salinité et la concentration en nutriments; Berline *et al.*, (2014), Nieblas *et al.*, (2014), et Rossi *et al.*, (2014) se sont basés sur les propriétés hydrodynamiques des eaux de surface du bassin ; Reygondeau *et al.* (2014)

a exploité la composition biologique donnée par une modélisation des habitats d'espèces marines sur le bassin.

Les dernières études de régionalisation de Reyngondeau *et al.*, 2017 confirment que les variables classiques tels que la SST, Chla et la profondeur de la couche de mélange (Mixed Layer Depth : MLD) utilisées dans les classifications décrivent bien le gradient West-Est en surface de la Méditerranée et permettent de partitionner l'eau de surface. Nous proposons dans ce chapitre une régionalisation de la Méditerranée en prenant en compte d'une manière simultanée ces 3 paramètres (SST et Chla issue des données satellitaires de MODIS Aqua et MLD issue des observations in-situ Argo), en utilisant une méthode novatrice de classification par réseaux neuronaux dites 2S-SOM. 2S-SOM est un algorithme de classification neuronal basé sur la méthode des K-moyennes (Chen *et al.*, 2012). Il repose sur une modification de la SOM en introduisant un système de poids. L'utilisation de ce type d'algorithme dans ce problème offre l'avantage de prendre en compte les relations existantes entre les 3 paramètres et en particulier, permet de prendre en compte de manière régionale les corrélations existantes entre les cycles de climatologies mensuelles.

Ce chapitre est articulé en deux parties : la première est consacrée au développement de la méthode et présente les résultats de la régionalisation de la Méditerranée en fonction des cycles climatologiques mensuels de SST, Chla et MLD ; la deuxième présente une étude de la variabilité des communautés phytoplanctoniques en Méditerranée à partir de l'application sur chaque sous région des méthodes SOM-Pigments et SOM-PFT.

2. Etat de l'art ; Biorégionalisation de la Méditerranée

Dans cette section nous résumons les méthodes utilisées par différents auteurs pour effectuer une régionalisation de la Méditerranée en faisant apparaître les principales variables qui permettent de distinguer différentes régions. Pour toutes les études décrites ici, les algorithmes de classification sont le plus souvent l'algorithme des k-moyennes et des méthodes de classification hiérarchique et le calcul des exposants de Lyapunov. Les différentes variables proposées qui concernent les biorégions sont analysées.

- D'Ortenzio and Ribera d'Alcala, (2008) ont utilisé un algorithme de classification (k-moyennes) pour regrouper des pixels en se basant sur leur cycle de climatologie hebdomadaire de Chla calculé sur 10 ans de données de SeaWiFs. Sept régimes trophiques ont été identifiés caractérisés par des patterns spatiaux coïncidents avec les structures océanographiques typiques du bassin.
- Mayot *et al.*, (2016) a enrichi le travail précédent en ajoutant 6 années de données MODIS. L'analyse de la classification en termes de variabilité annuelle a permis de détecter de nouveaux régimes trophiques. Ces régimes correspondent à des scénarios extrêmes non observés avec des moyennes de climatologies associées à des événements variables de convection profonde en hiver, à des instabilités sur les fronts, à l'advection de l'eau de l'atlantique et de la Mer Noire et au taux d'influence des rejets riverains.
- Palmiéri, (2014) a introduit l'aspect tridimensionnel en utilisant les sorties d'un modèle hydrodynamique et biogéochimique pour ses classifications. Il a pris en compte les cycles de chlorophylle (surface, intégrée, maximum suivant la profondeur, MLD, biomasse) pour définir les biorégions. La régionalisation est relativement similaire à celles de D'Ortenzio and Ribera d'Alcala (2008) et Mayot *et al.*, (2016), avec des régions caractérisées par des blooms phytoplanctoniques induits par une couche de mélange profonde (D'Ortenzio *et al.*, 2005; Lavigne *et al.*, 2013). Des différences ont été observées au niveau des régionalisations pdu fait de la prise en compte du profil de la Chla qui est une caractéristique importante au sein de la Méditerranée et qui dépend des phénomènes saisonniers de radiation solaire et de la disponibilité des

nutriments dans la colonne d'eau. En effet la limitation principale de ce type d'approche est l'utilisation de données de modèles uniquement.

- D'une façon similaire, Reygondeau *et al.*, (2017) a proposé une approche tridimensionnelle mais en utilisant des observations in-situ tri-dimensionnelles à la place des sorties de modèles. Cette approche subdivise la Méditerranée en 60 régions biogéochimiques. Les classes sont réparties dans les trois dimensions : 12 pour les couches épipélagiques, 12 mesopélagiques, 13 bathypélagiques et 26 pour le fond marin. Ces régions ont été délimitées via un consensus entre plusieurs algorithmes de classification appliqués à des climatologies de 16 variables hydrologiques et biogéochimiques moyennées pour chaque couche en profondeur. Les résultats de cette approche confirment que les variables classiques tels que la SST, Chl_a et MLD utilisées dans les classifications décrivent bien le gradient West-Est en surface de la Méditerranée et sont suffisantes pour partitionner l'eau de surface.
- Nieblas *et al.*, (2014) ont utilisé des climatologies mensuelles de variables classiques (température de surface de la mer, chlorophylle de surface et bathymétrie), ainsi que des caractéristiques de mésoéchelle (*Eddy kinetic energy*, exposants de Lyapunov « finite-size » et gradients frontaux de surface) pour décrire la surface. Grâce au partitionnement k-means, les auteurs ont identifié quatre régions biogéochimiques en Méditerranée en utilisant des variables «classiques», quatre régions en utilisant des caractéristiques à mésoéchelle et cinq régions lors de la combinaison des caractéristiques classiques et à moyenne échelle. Globalement, la régionalisation classique est stable dans le temps et dans l'espace, alors celles prenant en compte la mésoéchelle seule ou une combinaison de la mésoéchelle avec les variables classiques sont très variables. Cela montre que la stabilité apparente trouvée à partir de variables classiques uniquement n'est pas représentative de la «vraie» variabilité in situ de l'environnement du bassin Méditerranéen. Il est intéressant de noter que l'ajout de caractéristiques dynamiques de moyenne échelle aux variables biogéochimiques est important pour délimiter davantage les régions des eaux du large.
- Enfin, Reygondeau *et al.*, (2014) ont proposé une écorégionalisation prenant en compte les composants biologiques des écosystèmes de la mer Méditerranée. Ils ont utilisé des modèles de niches écologiques pour projeter la répartition potentielle de

plus de 800 espèces marines, réparties à tous les niveaux trophiques (producteur primaire, consommateur primaire, consommateur secondaire, prédateur supérieur). Ces communautés modélisées ont été regroupées en régions avec la même approche multi-algorithmes que Reygondeau *et al.*, (2017), appliqué à chaque niveau trophique. Ces régions ont été fusionnées par niveau trophique en 25 écorégions, ou 15 écorégions en utilisant uniquement des couches pélagiques. Ce travail propose une approche basée sur des variables biotiques, mais est limité par le faible nombre d'occurrences de ces variables enregistrées dans les bases de données, sur lesquelles les modèles de niche écologique ont été construits, en particulier pour les espèces de niveau trophique inférieur (producteurs primaires et consommateurs primaires). De plus, seules les occurrences méditerranéennes ont été prises en compte, même pour des espèces présentes ailleurs, ce qui fait que cette calibration régionale peut aboutir à des niches écologiques peu représentatives.

En conclusion, toutes ces études montrent qu'il devrait être possible de définir des biorégions à l'aide d'observations satellitaires de variables océaniques (SST, Chla de surface) disponibles en très grand nombre et d'observation in situ de la profondeur de la couche de mélange.

Dans notre travail, nous proposons une régionalisation de la Méditerranée à l'aide de variables classiques ; qui sont la Chla et la SST représentées par les valeurs mensuelles de leur cycle annuel, issues d'observations spatiales (MODIS Aqua) entre 2002 et 2018, combinées aux climatologies mensuelles des MLD obtenues à partir des observations Argos entre 2000 et 2018. Notre approche utilise un algorithme neuronal de classification plus complexe (2S-SOM), que la SOM permettant d'évaluer les relations inter-variables et leurs niveaux de contributions à la détermination des régions. Dans ce qui suit, on décrit les bases de données utilisées ainsi que l'algorithme appliqué.

3. Base de Données

3.1. Couleur de l'Océan

La Chla et la SST sont issues des produits moyennés de MODIS-Aqua, sur une durée de 17 ans (2002-2017). Ces données de climatologie mensuelle ont été téléchargées depuis le portail d'oceancolor ; oceancolor.gsfc.nasa.gov, avec une résolution de 4 km. Par default, ces paramètres sont calculés par des algorithmes spécifiques décrits par la suite.

3.1.1. Chlorophylle-a (Chla):

La Chla est calculée par une relation empirique dérivée des mesures in-situ de Chla et de réflectance marine (Rrs) dans la région bleu-vert du spectre visible. L'implémentation courante de l'algorithme de Chla emploie par default l'algorithme OC3M pour le capteur MODIS, fusionné avec l'algorithme CI (color index) de (Hu, Lee and Franz, 2012). Comme décrit dans ce papier, l'utilisation de tel algorithme est restreinte aux eaux relativement claires, et l'apport de cette utilisation réside dans la réduction des artefacts et biais au niveau de l'estimation de la Chla dans les eaux claires.

L'algorithme CI est basé sur une différence de 3 bandes de Rrs :

$$CI = Rrs(\lambda_{blue}) - \left[Rrs(\lambda_{blue}) + \frac{\lambda_{green} - \lambda_{blue}}{\lambda_{red} - \lambda_{blue}} * (Rrs(\lambda_{red}) - Rrs(\lambda_{blue})) \right] \quad (5.1)$$

Où λ_{blue} , λ_{green} et λ_{red} sont les longueurs d'ondes spécifiques au capteur, les plus proches à 443, 555 et 670 nm respectivement.

Alors que l'OC3M est une relation polynomiale du 4^{ème} ordre entre Rrs et Chla :

$$\log_{10}(Chla) = a_0 + \sum_{i=1}^4 a_i (\log_{10}(Rrs(\lambda_{blue})/Rrs(\lambda_{green})))^i \quad (5.2)$$

Tableau V.1 : valeurs des coefficients a_i pour l'algorithme OC3M

		λ_{blue}	λ_{green}	a_0	a_1	a_2	a_3	a_4
OC3M	MODIS	443>488	547	0.2424	-2.7423	1.8017	0.0015	-1.2280

Pour une estimation de Chla inférieure à 0.15 mg m⁻³, l'algorithme CI est employé. Pour une estimation de Chla supérieur à 0.2 mg m⁻³, l'algorithme OC3M est employé. Entre ces deux valeurs, les algorithmes CI et OC3M sont fusionnés par des moyennes pondérées.

3.1.2. Température de Surface de la mer (SST)

L'algorithme utilisé pour le calcul de la SST est basé sur une version modifiée de l'algorithme de (Walton *et al.*, 1998);

$$\begin{aligned}
 SST_{sat} = & a_{0ij} + a_{1ij} T_{11\mu m} + a_{2ij} (T_{11\mu m} - T_{12\mu m}) T_{sfc} \\
 & + a_{3ij} \left(\sec(\theta - 1) (T_{11\mu m} - T_{12\mu m}) \right) + a_{4ij} (mirror) + a_{5ij} (\theta^*) \\
 & + a_{6ij} (\theta^2) \quad (5.3)
 \end{aligned}$$

Avec:

- $T_{11\mu m}$ = Brightness Temperature 11 μm channel
- $T_{12\mu m}$ = Brightness temperature 12 μm channel
- T_{sfc} = SST Reference
- θ = angle zénithal du capteur
- θ^* = angle zénithal du capteur rendu négatif pour les pixels dans le premier demi de la ligne scannée
- Mirror = numéro du miroir
- coefficients a_{ij} = coefficients de l'algorithme, défini pour un moi de l'année i , et une zone de latitude j .

Les coefficients a_{ij} sont dérivés et continuellement vérifiés, basés sur des match-ups entre les mesures satellitaires de luminance et mesures de SST in-situ. Ces coefficients dépendent de la zone de latitudes et de la période de l'année.

3.2. Climatologie de profondeur de la couche de Mélange, Mixed Layer Depth (MLD)

Cette climatologie mensuelle ((Holte *et al.*, 2017), <http://mixedlayer.ucsd.edu/>) consiste à l'incorporation de 1.5M de profile issue des mesures Argo entre janvier 2000 et avril 2018 , à l'aide d'une méthode hybride décrite par Holte *et al.*, (2009). L'algorithme cherche à modéliser l'allure générale de chaque profil, identifie des caractéristiques physiques et calcule des seuils et gradients en assemblant plusieurs valeurs de MLDs. Une MLD finale est estimée à la suite de cette analyse.

La climatologie de MLD est déterminée par regroupement des profils dans des bin de 1 degré. Ensuite, pour chaque mois, dans chaque regroupement la moyenne, la médiane, le maximum, l'écart-type, la température, la salinité et la densité moyenne de la couche de mélange sont calculés. Ces variables sont calculées séparément en utilisant un algorithme et seuil de densité.

3.3. Base de données d'apprentissage :

Cette base de données regroupe les 3 paramètres Chla, SST et MLD de manière à former une climatologie annuelle des valeurs mensuelles de ces paramètres dans bassin Méditerranéen.

La zone d'étude est délimitée par les coordonnées géographiques suivantes : 46.5 N, 30 N, et 7.5 W, 36 E. A chaque pixel de cette zone correspond 12 valeurs mensuelles de Chla, de SST et de MLD moyennées sur la durée de la période d'étude (2002-2017).

La taille du pixel (1°) des composites de MLD s'avère être une limitation ; une image mensuelle de MLD sur la Méditerranée est composée de $56 \times 17 = 952$ pixels, alors que celle de la Chla ou SST est de 1081×397 pixels = 429157 pixels. Par conséquent, à chaque pixel de MLD, on attribue la moyenne des valeurs de Chla et SST qui coïncident spatialement avec le pixel de MLD.

Ainsi la base de données finale (D) est constituée d'une matrice de dimension $952 \times (12 \times 3)$ où les 952 lignes correspondent chacune à un pixel de la région extraite, et les 36 colonnes sont les variables regroupant 12 climatologie mensuelle de Chla, 12 climatologie mensuelle de SST et 12 climatologie mensuelle de MLD.

La structure de D a rendu primordiale l'utilisation de 2S-SOM pour ce type de problème. Cette méthode permet d'évaluer la contribution de chaque paramètre en termes de blocs dans le processus de classification. Dans la partie qui suit, on présente l'algorithme 2S-SOM qui permet de pondérer l'influence de chaque paramètre dans la classification.

4. Méthodologie:

4.1. Définition de l'Algorithme 2S-SOM

2S-SOM est une extension de l'algorithme de type subspace clustering basé sur la méthode des K-moyennes (Chen *et al.*, 2012). Il repose sur une modification de la fonction de coût de l'algorithme de SOM en introduisant un double système de poids α_{ck} et β_{ckj} définis respectivement sur les blocs ($b = 1, \dots, B$) et sur les variables ($j = 1, \dots, p_b$) pour chaque neurone c de la grille C . La classification est donc obtenue par optimisation de la fonction objective J_{2S-SOM} définie en (5.4).

$$J_{2S-SOM}^T(\chi, W, \alpha, \beta) = \sum_c \left(\sum_{b=1}^B \left(\sum_{z_i \in D} \alpha_{cb} K^T(\delta(c, \chi(z_i))) d_{\beta_{cb}}(i) + J_{cb} \right) + I_c \right) \quad (5.4)$$

Avec

$$d_{\beta_{cb}}(i) = \sum_{j=1}^{P_b} \beta_{cbj} (z_{ib}^j - w_{ib}^j)^2 \quad (5.5)$$

Et c appartient à la grille de 2S-SOM. Tout ces poids sont soumis aux contraintes suivantes :

$$\sum_{b=1}^B \alpha_{cb} = 1; \alpha_{cb} \in [0,1] \forall c \in [2S - SOM] \quad (5.6)$$

Et

$$\sum_{j=1}^{P_b} \beta_{cbj} = 1; \beta_{cbj} \in [0,1] \forall c \in [2S - SOM]; \forall b \in [1, B] \quad (5.7)$$

I_c et J_{cb} représentent les entropies négatives pondérées et associées aux vecteurs poids relatifs aux blocs et aux variables pour neurone c . Ces termes permettent d'ajuster, selon les hyper-paramètres μ et η , les contributions relatives apportées par les variables et les blocs dans la classification :

$$I_c = \mu \sum_{b=1}^{P_b} \alpha_{cb} \log(\alpha_{cb}) \quad (5.8)$$

Et

$$J_{cb} = \eta \sum_{j=1}^B \beta_{cbj} \log(\beta_{cbj}) \quad (5.9)$$

Les propriétés de conservation topologique de 2S-SOM sont influencées par les poids α_{ck} et β_{ckj} dans la classification en fonction des hyper-paramètres μ , η et le paramètre de voisinage T ;

- Lorsque μ tend vers l'infini et η est fixé, les poids α_{ck} associés aux blocs étant tous égaux à $1/B$. Alors, seuls les poids β_{ckj} des variables des blocs définissent les neurones.

Pour une valeur de μ très grande, l_c desorme important. A ce niveau, l'algorithme de 2S-SOM cherche à minimiser pour chaque c l'entropie du vecteur α_{cb} , et force tout ses coefficient a etre egaux.

- Lorsque η tend vers l'infini et μ fixé, les poids β_{ckj} des variables d'un bloc sont identiques à $1/p_k$. Dans ce cas seuls les blocs sont pénalisés selon leur capacité à définir les neurones. Dans ce contexte, 2S-SOM permet alors de déterminer pour chaque neurone les blocs qui lui sont spécifiques.

Pour une valeur de η très grande, j_{cb} desorme important. A ce niveau, l'algorithme de la 2S-SOM cherche à minimiser pour chaque c dans chaque bloque b , l'entropie du vecteur β_{cbj} , et force tout ses coefficient a etre egaux.

- Lorsque μ et η tendent vers une valeur très grande, les blocs sont équiprobables de même que les variables. Ainsi, l'algorithme 2S-SOM est comparable à celui de SOM.

4.2. Fonctionnement de l'algorithme 2S-SOM :

Pour μ et η fixés, l'algorithme de 2S-SOM se présente de la manière suivante :

- Étape 0 : Initialisation avec iteration de l'algorithme SOM, en fixant α et β à des valeurs homogenes.

L'optimisation de la fonction de coût J_{2S-SOM} s'effectue en iterent les 3 etapes (1, 2, et 3) presenter ci-dessous. Les deux premières phases d'affectation des observations aux classes et d'actualisation des vecteurs référents sont similaires à celles de SOM.

- Étape 1 : Les référents W , les poids α et β sont connus et fixés, les observations sont affectées aux neurones en respectant la fonction d'affectation:

$$c(zi) = \chi(z_i) = \underset{c \in C}{\operatorname{argmin}} \left(\sum_{r \in C} K^T(\delta(r, c)) \left(\sum_{b=1}^B \alpha_{cb} d_{\beta_{cb}}(i) \right) \right) \quad (5.10)$$

- Étape 2 : Actualisation des centres de neurone (les référents W) suivant la formule de l'algorithme SOM (2.3).
- Etape 3 : la fonction d'affectation et les referens W étant fixés, α et β sont déterminés suivant les equations (5.11, 5.12, 5.13, 5.14); Il s'agit de minimiser la fonction de cout J_{2S-SOM} relativement à α et β sous les contraintes de 5.6 et 5.7.

$$\alpha_{cb} = \frac{\exp\left(\frac{-\psi_{cb}}{\mu}\right)}{\sum_{b=1}^B \exp\left(\frac{-\psi_{cb}}{\mu}\right)} \quad (5.11)$$

Avec

$$\psi_{cb} = \sum_{zi \in D} K^T(\chi(z_i), c) d_{\beta_{cb}}(i) \quad (5.12)$$

Et

$$\beta_{cbj} = \frac{\exp\left(\frac{-\Phi_{cbj}}{\eta}\right)}{\sum_{b=1}^{p_b} \exp\left(\frac{-\Phi_{cbj}}{\eta}\right)} \quad (5.13)$$

Avec

$$\Phi_{cbj} = \sum_{zi \in D} \alpha_{cb} K^T(\chi(z_i), c) (z_{ib}^j - w_{cb}^j)^2 \quad (5.14)$$

Cet algorithme est répété en échantillonnant les hyper-paramètres μ et η .

Les coefficients de pondération α_{ck} et β_{ckj} définis par 2S-SOM indiquent respectivement l'importance relative des blocs et des variables dans les classes. Ainsi, plus le poids d'un bloc k ou d'une variable v_j est important, plus le bloc ou la variable contribue à la définition de la classe au sens où elle permet de réduire la variabilité des observations dans la cellule et dans son voisinage proche.

Finalement, à la convergence, 2S-SOM fournit d'une part une carte topologique permettant de visualiser les données et d'autre part des systèmes de poids pour les neurones de la carte.

4.3. Application de l'algorithme :

Phase 1 : Traitement de D :

Premièrement, les 12 variables de climatologie mensuelle de Chla sont rendues en valeur \log_{10} , permettant de mieux identifier la variabilité des faibles Chla face à celles qui sont élevées. Par suite, les 36 variables de climatologie de Chla, SST et MLD sont normalisées par leurs variances afin de réduire l'influence des différents ordres de grandeurs des variables sur l'organisation de la 2S-SOM.

Nous avons utilisé l'algorithme 2S-SOM en considérant 3 blocs (12 mois de Chla, 12 mois de SST, 12 mois de MLD) pour bien faire jouer l'influence de chaque variable en fonction de sa position géographique. Nous avons considéré qu'à l'intérieur d'un bloc les variables étaient équiprobables. Nous faisons ainsi l'hypothèse que chaque mois doit avoir la même influence. Cette définition des blocs est une étape essentielle pour la classification par 2S-SOM puisque les poids déterminés pendant l'apprentissage influencent la construction des vecteurs référent w_c de chaque neurone c de la carte.

Phase 2 : initialisation de l'algorithme et apprentissage de 2S-SOM:

Pour introduire des poids équiprobables sur les variables nous avons fixé η à une valeur très grande (egale à 2000) et échantillonné uniquement la valeur de μ . Ceci rend les valeurs de β homogènes égales à $1/p_k$, et les valeurs de α variables entre 0 et 1.

Dans cet application de l'algorithme, nous procéderons comme suit :

La valeur de μ est échantillonnée entre 4 et 0.01 avec un pas de 0.05, définissant 80 iterations. Dans ce contexte, 2S-SOM permet alors de pondérer pour chaque neurone les blocs qui le caractérisent.

À iteration notée $N_{iter}=1, \dots, 80$:

- Etape 1 : Une carte neuronale de dimension $10*5=50$ neurones est initialisée avec la base de données (D).
- Etape 2 : nous fixons μ à $0.01+0.05*N_{iter}$ avec η fixe dès le départ
- Etape 3 : Apprentissage de 2S-SOM

Ceci provoque la création de 80 grilles de 2S-SOM, chacune différente selon la valeur de μ .

La meilleure valeur de μ est choisie en fonction d'une mesure de performance obtenue pour chaque valeur de μ . La plus petite valeur de cette mesure définit la valeur idéale de μ .

La performance est définie comme une mesure de distorsion, soit l'erreur de quantification estimée quand la fonction d'allocation χ associe un vecteur référent w_j d'un neurone à une observation x_i . L'apprentissage a une tendance à minimiser l'erreur, donc la valeur de distorsion.

La mesure de distorsion s'écrit :

$$Distortion = \sum_i \sum_c K^T(\delta(bmu(z_i), c))(w_c - z_i)^2 \quad (5.15)$$

Avec $bmu(z_i)$, la fonction d'allocation qui associe une observation i au neurone c le plus proche (best matching unit, bmu), w_c le vecteur référent du neurone j , et z_i le vecteur de données de l'observation i .

La figure V.1 représente la distorsion en fonction des 80 valeurs de μ passant de 0.01 à 4 avec un pas de 0.05. Cette courbe met en relief une distorsion quasi-inchangeable au-delà de $\mu = 1.5$.

Simultanément, le passage de μ à une valeur relativement grande rend la distribution des poids des blocs plus homogènes, d'où une contribution équivalente des blocs vis-à-vis de la construction des vecteurs référents des neurones de la carte 2S-SOM.

En effet nous cherchons à avoir une distribution la plus hétérogène possible des poids. Nous choisissons donc une carte avec faible distorsion et forte variabilité des poids : dans notre cas μ et η ont été fixés à 1.65 et 2000 respectivement.

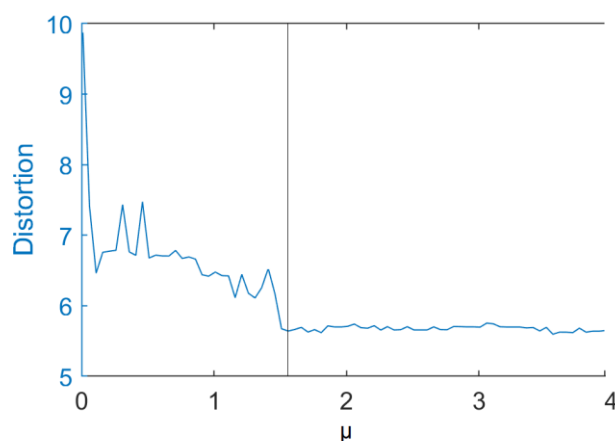


Figure V.1 : Diagramme représentant la distorsion et l'écart-type des poids (α) des blocs de variables pour chaque pas de η . La ligne noire verticale désigne la valeur de η choisi pour une valeur de distorsion minimale avec une hétérogénéité maximale de poids.

Phase 3 : Application d'une Classification hiérarchique Ascendante et regroupement des classes sur 2S-SOM.

La classification par 2S-SOM a permis une quantification vectorielle très fine entre les pixels. L'utilisation par la suite d'une CAH permet d'introduire un second niveau d'agrégation pour

obtenir de grandes régions cohérentes en termes de cycles climatologiques mensuels de Chla, SST et MLD. Dans notre cas, nous avons choisi la distance de Ward qui regroupe les classes de façon à ce que l'inertie interclasse soit maximale, ou en d'autres termes, à ce que l'inertie intra-classe soit minimale. Par la suite, l'arbre de classification hiérarchique a été coupé de manière à répartir les 50 neurones en 7 classes, le nombre de classe a été choisi après inspection des sauts observables dans l'arbre et une expertise liée au domaine.

Phase 4 : Représentation des Classes sur l'espace géographique de la Méditerranée.

La carte 2S-SOM (Chla, SST, MLD) est apprise avec une résolution géographique de 1° pour les paramètres. Pour obtenir des cartes plus fines (résolution de 4km), nous avons décidé d'utiliser la distance tronquée (section Article1) qui permet de décoder la carte en utilisant uniquement une partie des informations, c'est-à-dire les 12 valeurs de Chla et les 12 valeurs de SST. Au cours de la projection sur 2S-SOM, nous avons associé à chaque pixel son neurone le plus proche et nous lui avons associé ses valeurs de (Chla, SST, MLD). Le pixel prend alors la classe de son neurone.

Phase 5: Test de significativité des variables, one-way analysis of variance (ANOVA)

Afin de caractériser chacune des 7 classes, on a réalisé une analyse de significativité sur les variables, ceci de deux manières.

- Comparaison de la moyenne du cycle de la classe à la moyenne générale de la base entière **D**, vis-à-vis de la Chla, SST et MLD. Ceci met en relief la significativité de la différence de la Chla, SST et MLD entre une classe et **D**.
- La même analyse est réalisée mois par mois pour chacune des variables, permettant d'affiner la description des classes.

5. Résultat

5.1. Analyse de la 2S-SOM :

5.1.1. L'organisation des variables :

La figure V.2 présente pour chaque neurone de la grille 2S-SOM, les composantes mensuelles de son référent pour les mois de Janvier, Avril, Juillet et Octobre sur la grille de la 2S-SOM. Cette dernière s'est bien organisée sur chaque mois suivant des gradients de Chla croissant allant de haut en bas et de gauche à droite. Dans un deuxième temps, la SST en climatologie mensuelle présente une transition de gradient passant d'un mois à un autre ; Les mois de Janvier, Avril et Octobre présentent un gradient positif de SST ordonné du haut en bas et de droite à gauche, tandis que les mois de Juillet présentent un gradient positif de SST ordonné de haut en bas et de gauche à droite. La MLD représente un gradient positif de droite à gauche. En analysant les différentes organisations des variables sur la 2S-SOM, une tendance de corrélation peut être observée ; Pendant l'hiver la MLD atteint des maximums qui coïncident avec les fortes valeurs de Chla et faibles SST ; d'une manière opposée en été les fortes valeurs de SST coïncident avec des faibles valeurs de SST et de Chla. Chaque neurones spatialise la carte géographique et est représenté par une région fermée sur celle-ci.

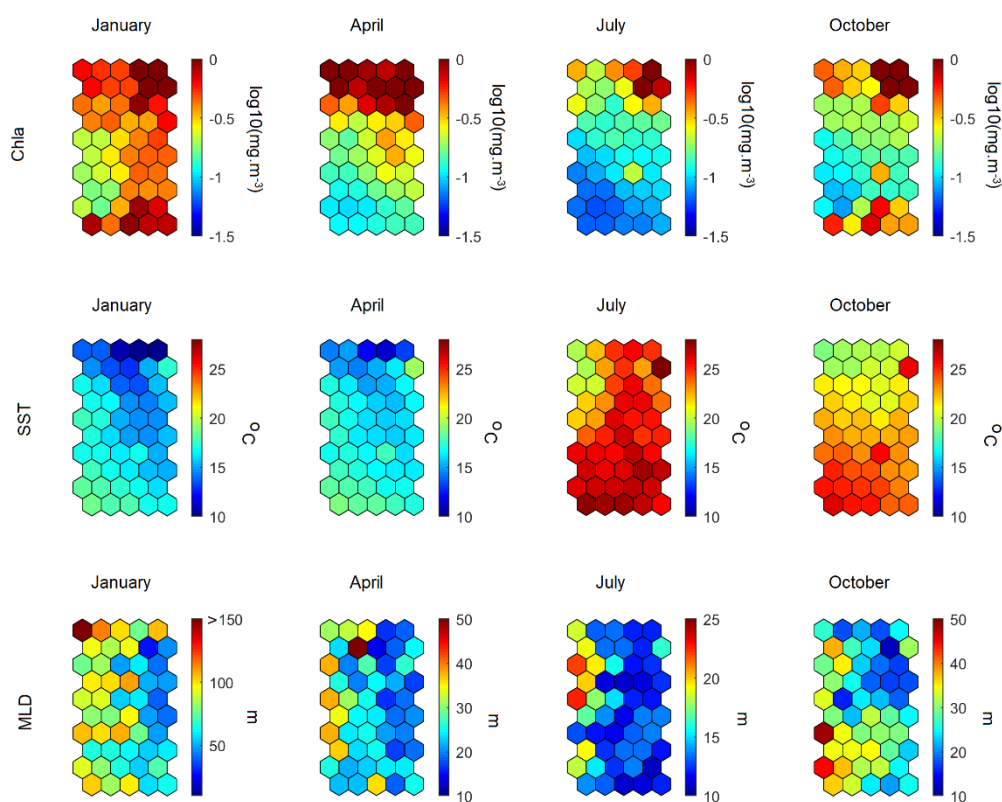


Figure V.2 : Représentation des différentes variables mensuelles (Janvier, Avril, Juillet, Octobre) de Chla, SST et MLD au niveau des neurones de la 2S-SOM après apprentissage.

5.1.2. Analyse des Poids des blocs

La figure V.3 représente les différents poids attribués aux blocs de variables pour la détermination des différents neurones de la 2S-SOM. Par normalisation dans l'algorithme d'apprentissage, pour un neurone donné la somme des valeurs des poids des 3 blocs est égale à 1.

On remarque qu'en fin d'apprentissage, les poids ont été distribués d'une façon hétérogène sur la 2S-SOM. Des neurones pour lesquels le poids d'une des trois variables a une très forte valeur et les deux autres étant quasi nuls, représentent des zones géographiques où cette variable est vraiment prépondérante dans la classification 2S-SOM.

Pour la Chla, les poids proches de 1 ont été attribués aux neurones qui représentent des faibles ou des fortes Chla. De même pour la SST les poids proches de 1 ont été attribués en général aux neurones de faible SST ; un seul neurone correspond à une forte SST (les eaux du cas 2 du Bassin Est). Pour la MLD les poids proches de 1 ont été attribués aux neurones caractérisés par de faibles valeurs de MLD.

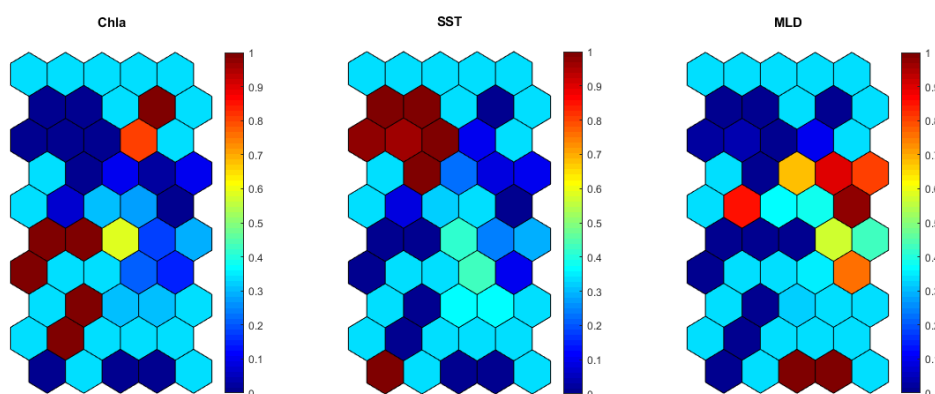


Figure V.3 : Représentation des différents poids des blocs de Chla, SST et MLD au niveau des neurones de la 2S-SOM après apprentissage.

5.1.3. Classification des neurones.

L'application de la CAH sur les neurones de la 2S-SOM a permis de regrouper 7 classes différentes. La représentation de la distribution des classes sur la carte topologique est observée dans la figure V.4. Les 7 classes sont bien délimitées mettant en relief la bonne

organisation de la 2S-SOM. Pour chaque classe, on peut voir pour les 3 variables la distribution des poids des neurones qui la constitue, cela donne une vision de l'homogénéité en termes de poids des classes. Par exemple, la classe 5 a des poids de 0.33 attachés à chaque variable et chaque neurone, indiquant une prise en compte égale des trois variables. De manière plus complète on peut dire que :

- La classe 1 est caractérisée par des poids variables sur la MLD, la Chla étant très peu prise en compte ;
- La classe 2 est plus caractérisée par des poids forts sur la MLD
- Les classes 3 et 4 sont caractérisées par des poids forts sur la SST
- Les classes 6 et 7 sont caractérisées par des poids plus fort en moyenne sur la Chla

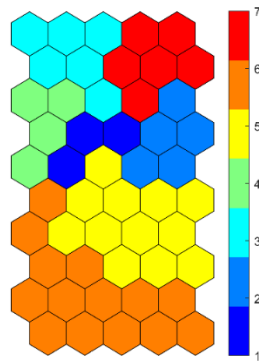


Figure V.4 : Résultat du regroupement des neurones de la 2S-SOM par la CAH. 7 classes ont été bien délimitées sur la carte neuronale.

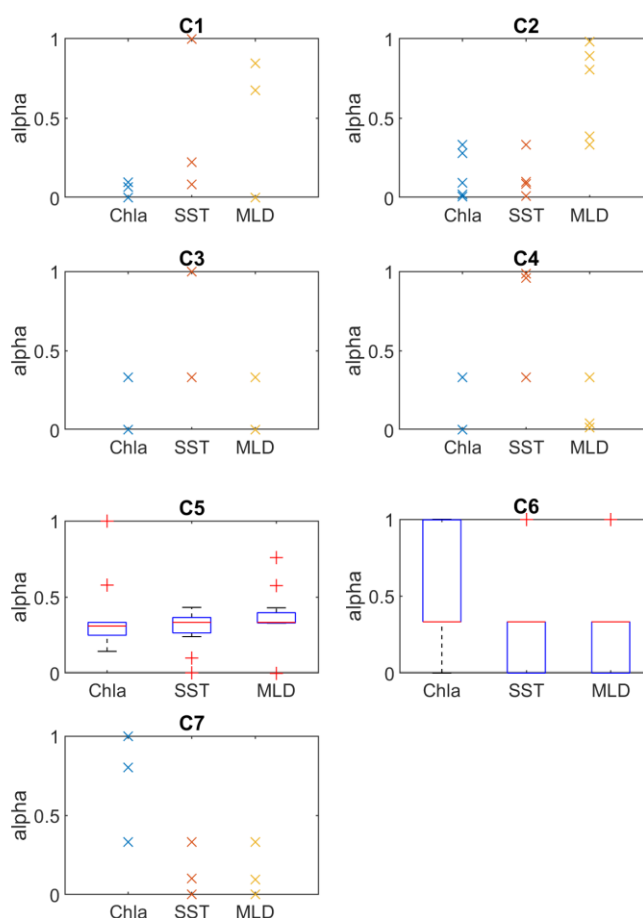


Figure V.5 : Representation des poids pour chaque bloc de Chla, SST et MLD au niveau des différentes classes de neurones de 1 à 7. C5 et C6 ont été représenté sous forme de boxplot vu le grand nombre de neurones regroupés dans ces classes

5.2. Caractérisation des Cycles de Climatologie mensuelle de Chla, SST et MLD de Chaque classe

Après apprentissage, chaque neurone de la 2S-SOM est caractérisé par un vecteur référent constitué de 3 cycles de climatologie annuelle représentés par leurs valeurs mensuelles (un de Chla, un de SST et un de MLD). Les classes qui regroupent ces neurones sont ainsi représentées par des cycles moyens issus de ce regroupement. La figure V.6 suivante représente des boxplots de la distribution des différents cycles de climatologie mensuelles de chaque classe de 1 à 7 comparées aux cycles moyens (derniers boxplots à droite). Les étoiles représentent une différence significative entre la moyenne de la classe et la moyenne de la distribution totale vis-à-vis d'un paramètre donnée (Chla, SST, MLD). Les flèches représentent les valeurs du test (ANOVA) (noires à pvalue < 5% et rouges à pvalue < 1%), en notant une

différence significative entre le mois au sein d'une classe, et le même mois de la distribution totale. L'analyse des différentes classes en termes de valeurs prises par les paramètres océaniques montrent que :

- C1 regroupe des cycles à faible Chla et SST relativement modérée tout au long de l'année, avec une différence non significative comparée au cycle moyen. Tandis que la MLD représente un cycle caractérisé par des valeurs significativement importantes (>100 m) pour les mois de Février et Mars, et significativement petites pour le mois de Septembre.
- C2 est caractérisée par des cycles de Chla et de SST comparables à ceux de C1, tandis que le cycle de MLD est significativement différent de celui de la distribution moyenne, et représente une MLD peu profonde tout au long du cycle. Les mois d'Aout à Janvier sont caractérisés par une MLD significativement superficielle.
- Les cycles de Chla, SST et MLD de C3 sont significativement différents de ceux de l'ensemble ; la Chla augmente brusquement aux mois de Février-Avril, coïncidant avec des faibles SST tout au long du cycle, et des MLD significativement profondes spécifiquement lors de la période qui s'étend entre Janvier et Mars.
- C4 présente des SST et MLD significativement différentes de l'ensemble, avec un cycle de SST de faible amplitude et des MLD moyennement profondes en hiver. La Chla, comme les classes C1 et C2, est caractérisée par des faibles concentrations avec une augmentation de celles-ci durant les mois d'hiver.
- C5 et C6 sont caractérisées par des Chla et SST significativement différentes de l'ensemble, soit une Chla minimale tout au long du Cycle et des SST relativement élevées. C6 représente des valeurs de Chla inférieures et des SST supérieures à celles de C5. La MLD est peu profonde dans les deux classes, et présente un mélange pendant les mois d'hiver.
- C7 représente une classe à forte variabilité de Chla et de SST, significativement différente de point de vue de ces 2 paramètres. La Chla arrive à 2mg en moyenne pendant les mois d'hiver-printemps et diminue pendant les mois d'été. Le cycle de SST présente une forte amplitude, le minimum atteint en moyenne les 10 degrés C en Février et le maximum 25 degrés C en Aout. La MLD est peu profonde, présentant une couche de mélange pendant les mois d'hiver.

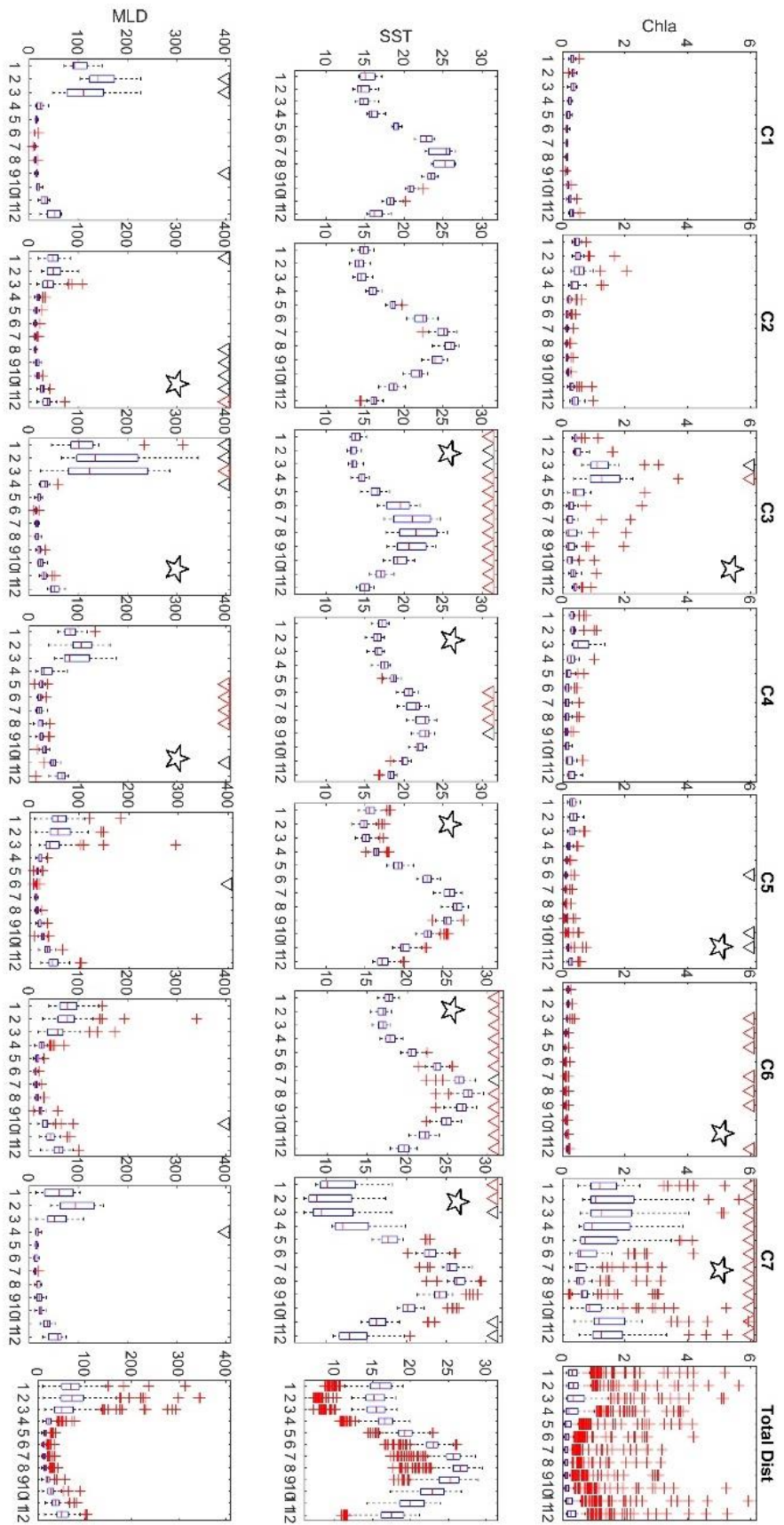


Figure V.6 : Boxplot représentant les différents cycles de climatologie mensuelle de Chla, SST et MLD pour chaque classe de 1 à 7 et de même que le cycle moyen de la Méditerranée. Les crois en rouge représentent les données «aberrantes». Les résultats du test ANOVA sont marqués par une étoile notant une différence significative pour les cycles de chaque classe (à pvalue<1%) et par des triangles notant une différence significative pour les mois de chaque cycle (noire pvalue<5%, rouge <1%).

5.3. Distribution géographiques des classes

Comme nous l'avons décrit dans la section 3.2 phase 4, les classes sont réparties géographiquement en Méditerranée selon les cycles annuels de Chla, SST et MLD. Les différentes classes ont une distribution géographique bien délimitée créant des espaces homogènes et non pixélisés (Figure V.7). Nous allons les interpréter en termes de biorégions

- C1 est présente dans le centre de l'Adriatique et de la Mer Egée. Ces zones sont le siège de convection hivernale et de formation d'eau Adriatique d'une part et de LIW d'autre part. Ceci est mis en relief par la MLD qui est significativement importante pendant les mois d'hivers et le niveau de pondération qui a été automatiquement établi par la 2S-SOM sur ce paramètre.
- C2 caractérise la Mer d'Alboran, est présente le long des côtes espagnoles sud, longe les côtes marocaines, algériennes et tunisiennes. Elle caractérise aussi les eaux du centre de l'Adriatique qui s'écoulent vers la mer Ioniène. On rencontre aussi des eaux de type C2 en mer Tyrrhénienne nord, au large des côtes Italiennes et Françaises. Ces eaux sont associées à des courants impliquant des advections importantes et permanentes. Ceci est mis en évidence par le poids attribué à la MLD qui se traduit par des valeurs de MLD peu profondes tout au long du cycle saisonnier. La présence d'eaux C2 au niveau du détroit de Gibraltar, puis le long de la côte africaine est lié à l'advection d'eau atlantique de densité légère entrant en Méditerranée et formant le courant algérien. Au niveau des côtes nord du bassin occidental, C2 est associée à l'advection par le courant Nord d'une eau atlantique modifiée. La présence de C2 au nord de la mer Egée lié à l'advection d'eaux provenant de la mer noire via les Dardanelles.
- C3 caractérise la Mer des Baléares. Les valeurs de la Chla, de la SST et de la MLD dénotent une classe où se produit une convection profonde en hiver. La Mer Baléare est le siège d'une convection hivernale très profonde qui engendre un mélange important et par conséquent la remonté d'eau froide vers la surface et riche en nutriment. Le poids attribué à la SST met en relief l'influence de celle-ci sur la caractérisation de la C3. En effet, la région Nord-Ouest de la Méditerranée Occidentale est marquée par les phénomènes de convection profonde en hiver dans le Golfe du Lion responsable de la formation d'eau dense (WMDW) engendrés par les coups de

vents forts caractérisant le Mistral et la Tramontane. Ce phénomène dont l'occurrence et l'amplitude varient considérablement d'une année à l'autre, influence la dynamique du phytoplancton. Cette convection explique le 'bloom' phytoplanctonique très remarquable observé à la fin de la convection aux mois de Mars-Avril. Ces zones de convection et de formation d'eaux denses, sont aussi les moteurs de toute la circulation thermohaline Méditerranéenne (Millot and Taupier-Letage, 2005).

- C4 caractérise l'eau Atlantique de surface dans le golfe de Cadix et est observée dans le détroit de Gibraltar sous la forme d'une petite langue voulant entrer en Méditerranée. Cette masse d'eau est bien identifiée par la 2S-SOM, étant donné sa SST et sa MLD. Cette eau atlantique de surface a une densité plus légère que l'eau méditerranéenne qui sort en profondeur, due à une température plus chaude et une salinité plus faible.
- C5 et C6 définissent les masses d'eau caractérisant les eaux centrales du bassin occidental (C5) et du bassin oriental (C6). C5 qui occupe en grande partie le centre du bassin Occidental a des SST plus faibles et des Chl_a plus importantes que celle de C6. Le bassin occidental est caractérisé par la réception d'apports fluviaux assez importants, par un niveau d'évaporation moins intense que celui de la Méditerranée Orientale, et par sa structure de circulation complexe. La C6 occupe le bassin orientale qui est connue par son caractère oligotrophe (Tsiola, Pitta and Fodelianakis, 2016) avec des fortes stratifications estivales. Le tourbillon cyclonique de Rhodes (Rhodes Gyre) situé entre Crète et Chypre, est le siège d'une convection hivernale intense avec formation de LIW qui remplit les couches de profondeur moyenne du bassin oriental. Cette convection engendre un refroidissement des eaux de surface et un enrichissement en sels nutritifs qui engendre une augmentation de la production phytoplanctonique ce qui explique que le Rhodes gyre appartient à C5 plutôt qu'à C6.
- C7 regroupe des zones côtières telles que celles qui longent la côte Est (Adriatique) et Ouest (Tyrrhénienne) de l'Italie, une partie des côtes du Golfe du Lion, les côtes de l'Égypte et du Liban. Les eaux C7 sont des eaux fortement influencées par les décharges des fleuves (Pô - (Syvitski and Saito, 2007); Rhône ; Ebre ; Nil - (Shata and El Fayoumy, 1970) et la pression anthropique. On rencontre aussi la classe C7 dans des

zones de faible profondeur où la marée peut être intense et remettre en suspension les sédiments (Golfe de Gabes, Nord de l'Adriatique, mer de Marmara)

En conclusion, la classification selon des critères biophysiques (SST, MLD, Chla) des eaux de la Méditerranée par 2S-SOM a mis en évidence des classes significativement différentes en regroupant des cycles de climatologie saisonniers de Chla, SST et MLD ayant des propriétés communes. En analysant les différents aspects de cette classification, ces classes peuvent être caractérisées comme des biorégions régies par des processus physique et biogéochimiques spécifiques.

Dans la partie qui suit, nous allons analyser la variabilité des groupes de phytoplancton estimés par la méthode des SOM décrites précédemment.

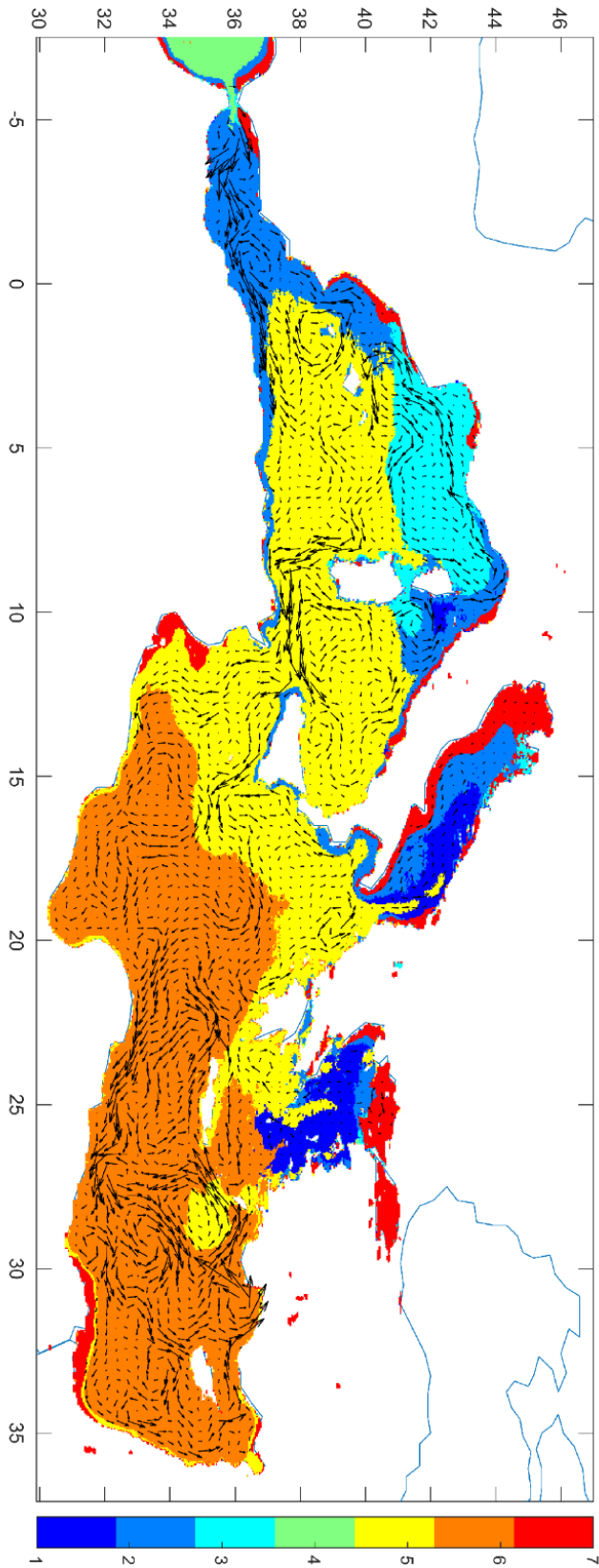


Figure V.7 : Distribution géographique des classes de 1 à 7 au niveau de la Méditerranée

5.4. Etude de la variabilité du phytoplancton au niveau des biorégions

Méditerranéennes

En utilisant les cycles de climatologie mensuelle de Chla, SST et MLD, nous avons pu partitionner la Méditerranée en 7 biorégions bien distinctes.

Nous avons ensuite essayé de caractériser ces biorégions en termes de diversité et de complexité des communautés phytoplanctoniques. Pour cela, nous nous sommes servi des groupes dominants de phytoplancton à 4km de résolution que nous avons déterminé au chapitre IV à l'aide de la SOM-PFT avec comme 'inputs' les images satellitaires quotidiennes de couleur de l'océan et de SST fournies par GlobColour entre 1998 et 2018. Pour chaque biorégion, nous avons calculé le pourcentage de pixels attribué à chacun des groupes de phytoplancton (Hapto, Chloro, Crypto, Proc, Syne et Diatoms). La Figure V.8 illustre pour les 7 biorégions, la variabilité du pourcentage des groupes de phytoplancton par rapport à la Chla OC5. Une climatologie annuelle a ensuite été calculée dont le cycle moyen est représenté sur la Figure V.9.

- Comme premier résultat, nous remarquons que, si l'on considère dans l'ordre les biorégions C7-C3-C2-C1-C5 et C6, on remarque un gradient négatif de la concentration en Chla, passant d'une fluctuation qui atteint les 4 mg m⁻³ au niveau de C7 à 0.2 mg m⁻³ au niveau de C6.
- Les groupes Hapto et Chloro possèdent en général un cyclique bien régulier. Ces groupes s'étendent depuis le début d'automne jusqu'à la fin du printemps atteignant des valeurs de 80% pour les Chloro et de 50% pour les Hapto respectivement ; leur prévalence diminue graduellement si l'on suit les classes dans l'ordre C3-C2-C1-C5, jusqu'à atteindre un cycle irrégulier au niveau de C6, qui présente un pic de Chloro en Hiver (50%), et un autre pic d'Hapto (25%) au printemps. Les Chloro sont les contributeurs majeurs à la Chla dans les biorégions C1, C2, C5 et C6, suivi par les Hapto. Les Crypto font leur apparition pendant les mois d'été arrivant à 30% de prévalence.
- Les Diatomées sont prédominantes dans C7 qui s'est avérée être une biorégion confinée aux zones côtières influencées par des apports terrigènes. De même C2 est caractérisée par des faibles fluctuations de Diatomées, marquant une influence des apports côtiers. C3 est caractérisée par des blooms brusques de Diatomées, coïncidant avec la fin de la convection profonde et contribue fortement aux valeurs élevées de

Chla qui atteignent 1.5mg m⁻³ lorsque les Diatomées prédominent. Cette prédominance de ce groupe de phytoplancton est expliquée par la prédisposition des sels nutritifs en grande quantité et notamment la silice qui provient de l'érosion des roches terrigènes par l'action du ruissèlement ; ce sont les Diatomées qui sont les premières à exploiter les sels nutritifs.

- En été, on observe une dominance de plusieurs types de cyanobactéries, tels les *Synechococcus* et *Prochlorococcus*, principalement dans les eaux du bassin oriental qui appartiennent à la biorégion C6. Cette biorégion est marquée par une importante stratification estivale limitant les échanges entre les eaux de surface et les couches plus profondes et par conséquent la remontée des nutriments. Dans ce type d'eau oligotrophe, ce sont les cyanobactéries qui prolifèrent le plus, au détriment des autres groupes phytoplanctoniques à cause de leurs adaptations physiologiques différentes (comme la photo-adaptation face au fort éclaircissement estival et aussi la mixotrophie). La prédominance des *Synechococcus* a été rapportée dans plusieurs travaux analysant des observations in-situ, caractérisant ainsi les eaux du bassin Est tout au long du cycle annuel.

C4 qui est classé loin des autres classes par la 2S-SOM, représente les eaux de l'Atlantique. Au niveau de la dynamique du phytoplancton, cette biorégion est caractérisée par un cycle d'Hapto et de Chloro d'amplitude plus faible en hiver comparé à celui de la Méditerranée orientale. Il présente de faibles blooms de diatomées en hiver en raison de l'homogénéisation de la colonne d'eau et de l'influence des côtes adjacentes qui rendent disponibles les matières nutritives essentielles pour ces blooms.

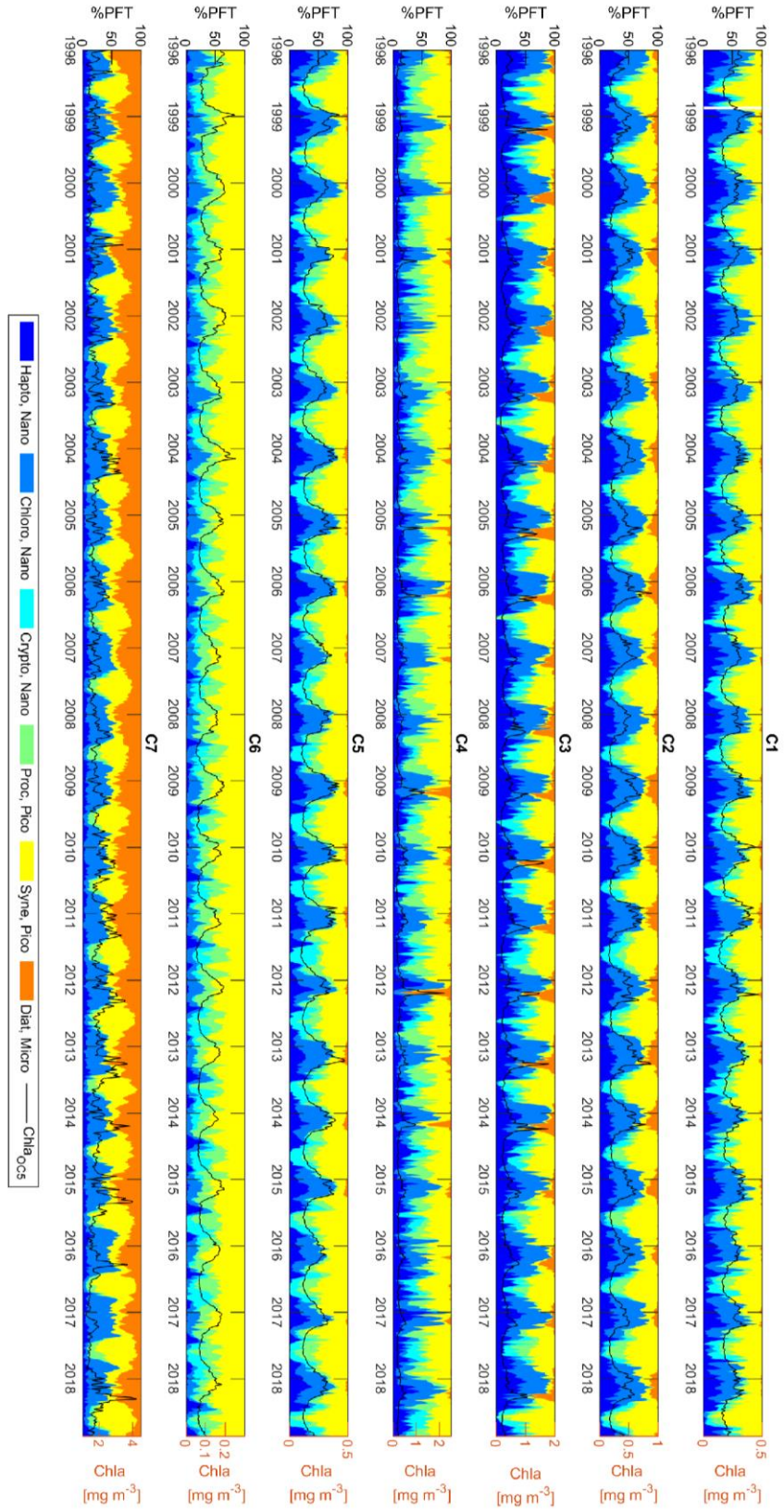


Figure V.8 : Reconstruction de la variabilité temporelle des différents groupes de phytoplancton estimé par la méthode des SOMs (SOM-Pigments/-PFT) au niveau de chaque bio régions obtenues.

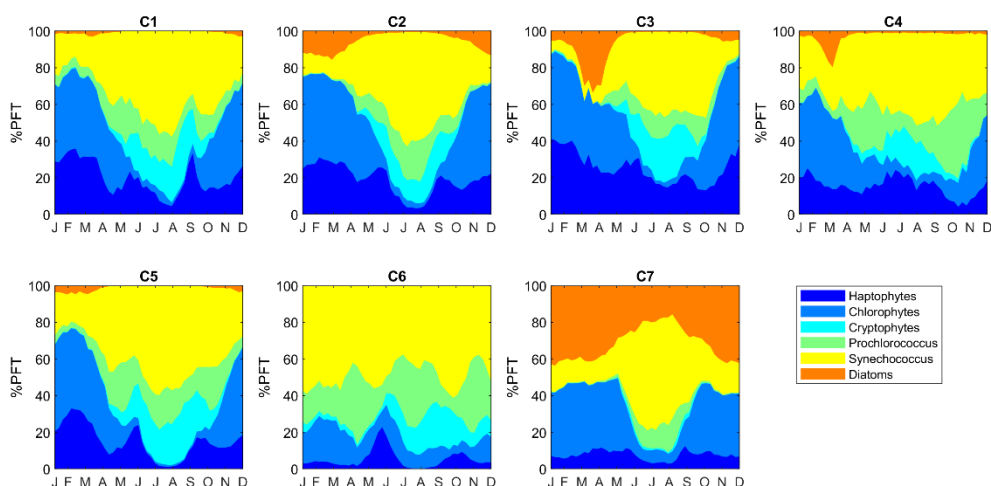


Figure V.9 : Cycles moyens (1998-2018) des différents groupes de phytoplancton estimé par la méthode des SOMs (SOM-Pigments/-PFT) au niveau de chaque biorégion obtenues.

6. Conclusion:

En conclusion, la classification effectuée par 2S-SOM en regroupant des cycles de climatologie annuelle représentés par les valeurs mensuelles de Chla, SST et MLD a permis de mettre en valeur des classes significativement différentes. En analysant les différents aspects de cette classification, ces classes peuvent être caractérisées comme des biorégions régies par des processus physique et biogéochimiques spécifiques. La classification a permis d'identifier 7 biorégions distinctes :

La première met en valeur les eaux de la mer Egée et de la mer Adriatique, toutes deux sièges de formation d'eau profonde. La deuxième est régie par des courants de surface importants, avec une influence côtière. La troisième est confinée au bassin des Baléares qui est caractérisé par des phénomènes de convection hivernal profonde, enrichissant cette région en matières nutritives provenant des couches profondes, provoquant ainsi une poussée phytoplanctonique importante à la fin de l'hiver. La 4ème biorégion caractérise les eaux atlantiques, correspondant à une combinaison des patterns de Chla, MLD et SST bien spécifique. C5 représente les eaux du large du bassin occidental et est influencée par les processus hydrodynamiques qui régissent le bassin occidental, comme l'advection de l'eau atlantique, ce qui fait que le phytoplancton prolifère étant donné les nutriments présents dans l'eau de surface. C6 caractérise les eaux les plus oligotrophes du bassin et s'étend du

centre du bassin méditerranéen jusqu'à l'Est. Ces eaux caractérisées par des fortes SST ce qui provoque des phénomènes de stratification en été empêchant ainsi la prolifération du phytoplancton. Finalement les eaux de C7 sont les plus eutrophes, dus à des décharges de rivières importantes (Nil, Pô, Rhones, Ebre), des zones côtières influencées par les ruissellements continentaux et des zones de marée telque le Golfe de Gabe ou de Triestes.

La variabilité des groupes de phytoplancton estimés par la méthode des SOM a permis de mieux caractériser ces biorégions vis-à-vis de la complexité de la communauté phytoplanctonique ; la dominance des groupes de Nanophytoplancton est observée en grande partie dans le bassin occidental pendant la période qui va de l'automne jusqu'au printemps, région représentée par C3 et C5. Alors que la dominance des différents types de cyanobactéries *Synechococcus* et *Prochlorococcus* est mise en relief en été et plus précisément dans les eaux de la biorégions C6 du bassin oriental. Les diatomées dominent tout au long de l'année dans les régions côtières et estuariennes marquées par C7, ce qui peut être expliqué par l'apport de silice terrigène indispensable pour la prolifération de ce type de phytoplancton. De même les diatomées bénéficient en grande partie de la forte convection profonde au niveau de la biorégions C3 marqué par un bloom important vers la fin de la convection hivernale en mars.

Conclusion Générale :

Nous pensons que nous avons apporté une contribution substantielle aux différents objectifs de cette thèse qui était axée sur la compréhension de la variabilité du phytoplancton à différentes échelles, allant du global au régional, pour arriver à une étude fine de la Méditerranée. Les apports que nous avons faits sur la compréhension de la dynamique du phytoplancton en utilisant la télédétection satellitaire s'avèrent très diversifiés et ont nécessité le développement de recherches novatrices aussi bien théoriques que techniques.

Dans notre première étude, nous avons développé une méthode d'estimation des pigments secondaires à partir des images satellitaires de différents capteurs de couleur de l'océan (base de données GlobColour) et de SST (AVHRR), à l'échelle globale. Les bonnes performances de notre méthode sont dues d'une part à l'efficacité remarquable des cartes topologiques auto-organisatrices, d'autre part à la grande base de données que nous avons pu constituer pour l'apprentissage des cartes topologiques qui comprend colocations entre mesures satellitaires (Chla, Rrs et SST) et pigments phytoplanctoniques mesurés in-situ. Les pigments secondaires que nous avons estimés à partir des mesures satellitaires permettent de remonter aux PFT et PSC.

Comparée aux approches empiriques qui ont été appliquées par de nombreux chercheurs pour la reconstruction de la variabilité des pigments phytoplanctoniques secondaires, l'utilisation de la SOM-Pigments a permis de mieux évaluer la relation qui existe entre la Chla qui prédomine au sein du phytoplancton et les pigments secondaires, en prenant en compte l'information de la variabilité du second ordre fournie par le signal de réflectance. La température de surface de la mer s'est avérée une composante importante de notre approche, en ajoutant une rythmicité saisonnière, et une bonne délimitation des structures de transports des pigments secondaires. La méthode a aussi permis de caractériser les incertitudes sur les données de réflectance via le développement d'un masque et aussi d'évaluer les incertitudes sur les estimations des pigments. La cross-validation montre que la méthode donne des résultats significatifs ; on note un coefficient de régression $R^2 > 0.75$ et une RMSE moyenne de 0.016 mg m^{-3} . Le protocole conçu pour l'application de la méthode fait que la qualité des données est contrôlée, ce qui permet une bonne estimation des différents pigments à l'échelle du globe. Une première estimation à l'échelle de l'océan global

de la climatologie saisonnière des pigments secondaires a pu être effectuée, en respectant la variabilité du phytoplancton ; les diatomées caractérisées par la fucoxanthine, sont prédominantes au niveau des pôles géographiques, des zones d'upwelling et des zones côtières. La 19HF délimite bien la répartition des coccolitophores dans les régions de moyennes et hautes latitudes, et au niveau du courant circumpolaire au nord de l'Antarctique. De même, la Zea donne des informations sur la répartition des cyanobactéries du type *Synechococcus* à l'équateur, alors que la DVChla représente les *Prochlorococcus*, omniprésents dans les régions oligotrophes de l'océan global. Des limitations majeures sont mises en relief lors de l'application de la méthode dans la zone de l'antarctique qui est caractérisée par une dynamique de communautés phytoplanctoniques bien spécifiques et dont les caractéristiques optiques sont différentes du reste du Globe. Ceci est évalué grâce au masque qui compare les données de réflectances observées par la SOM-Pigments et les données de réflectance mesurée au niveau de l'océan austral.

L'application de cette méthode sur une étude plus régionalisée au niveau de la Méditerranée met en relief la robustesse de cette approche. SOM-Pigments a été considérée comme générateur d'images quotidiennes de concentration en pigments secondaires sur la Méditerranée, et a permis de constituer une base de données pour l'apprentissage d'une seconde carte topologique régionalisée dédiée à l'identification des groupes de phytoplanctons dominants (SOM-PFT). Par cette méthode, six groupes de phytoplancton ont été identifiés: Haptophytes, Chlorophytes, Cryptophytes, *Synechococcus*, *Prochlorococcus* et Diatoms. Ces groupes ont montré une variabilité particulière due aux caractéristiques hydrophysiques et biogéochimiques complexes de la mer Méditerranée ; Les Haptophytes et les Chlorophytes dominent pendant l'hiver et principalement dans le bassin ouest de la Méditerranée, tandis que les *Synechococcus* et les *Prochlorococcus* sont prédominants pendant l'été. La dominance des diatomées a été principalement observée au printemps dans la mer des Baléares en réponse à des phénomènes de convection en eau profonde qui génèrent un apport de nutriments et près du littoral et des estuaires en raison d'importants apports continentaux de nutriments. Nous avons constaté que les Cryptophytes ont une faible présence en automne. Les tests de validation effectués sur des observations in-situ ont donné des résultats satisfaisants (<30% d'erreurs) et ont montré la capacité de la méthode à

reconstruire efficacement les patterns spatio-temporels des groupes de phytoplancton en mer Méditerranée.

Une deuxième exploitation de la SOM-Pigments au niveau de la Méditerranée a mis en évidence la variabilité du phytoplancton en termes de classe de tailles (PSC), étude qui est en accord avec l'étude de la variabilité des groupes de phytoplancton. De plus, cette étude a évalué la relation entre PSC et comptage de PSC in-situ en utilisant deux longues séries temporelles. Des coefficients de corrélation de Spearman ont mis en évidence une bonne performance vis-à-vis la reconstruction des patterns saisonniers par la SOM-pigments, et des études de régression de type log-linéaires ont pu décrire des tendances positives entre la fraction de Chla pour chaque classe de Micro, Nano et Pico-phytoplancton étudiée et les comptages in-situ de phytoplancton. Le coefficient de régression, qui s'est avéré être de qualité moyenne ($0.3 < R^2 < 0.5$), dénote plusieurs limitations de la relation qui lie la fraction de Chla de chaque PSC à l'abondance de ce dernier ; la qualité des observations satellitaires, les incertitudes sur les estimations des pigments secondaires par la SOM-Pigments, la physiologie du phytoplancton et les caractéristiques optiques, affectent cette relation.

Finalement une dernière étude, utilisant les caractéristiques hydrodynamiques de la méditerranée et la phénologie hétérogène du phytoplancton, a permis d'effectuer une bio-régionalisation à partir des données satellitaires de SST et de réflectances marines, couplées à des données in-situ d'Argo-floats permettant d'obtenir une climatologie mensuelle de la MLD. Cette bio-régionalisation a été effectuée en utilisant une méthode neuronale issue d'une modification de la SOM, appelée 2S-SOM, qui consiste à introduire des poids différents selon les variables. Cette dernière offre un outil pour évaluer la contribution des différentes variables à la régionalisation de la Méditerranée. Suivant cette approche, on a défini 7 biorégions, ayant des distinctions bien significatives entre elles qui sont ;

- une biorégion (C3) délimitant la zone de convection profonde hivernale en mer des Baléares
- deux biorégions caractérisant l'une les eaux centrales de la Méditerranée occidentale (C5), l'autre les eaux centrales de la Méditerranée orientale (C6). Ces biorégions diffèrent par leur niveau trophique, le bassin oriental étant nettement plus oligotrophique et plus chaud que le bassin occidental.

- une biorégion (C2) caractérisant des eaux soumises à une forte advection et assez riches en chlorophylle que l'on trouve en mer d'Alboran, le long de la côte Algérienne, au centre de l'Adriatique et en mer Tyrrhénienne.
- une biorégion (C1) caractérisant les eaux côtières de l'Est de l'Adriatique et les eaux centrales de la mer Egée, qui sont chacune le siège de formation d'eaux profondes typiques de la Méditerranée Est.
- une biorégion (C7) caractérisée par l'advection des courants de surface avec une influence côtière (le long de la cote égyptienne, libanaise et syrienne, et le long de la côte adriatique italienne et aussi dans le golfe de Gabes) caractérisée par des eaux du cas 2, dus soit à l'apport des fleuves (Nil, Pô, Rhones, Ebre) soit à des plateaux continentaux peu profonds permettant la remise en suspension de sédiments avec des marées importantes.
- Enfin une eau Atlantique présente dans le golfe de Cadiz qui pénètre un peu en mer d'Alboran en franchissant le détroit de Gibraltar (C4)

Cette classification robuste et bien délimitée spatialement montre la robustesse de la 2S-SOM, qui permet d'identifier et de délimiter des masses d'eaux différentes selon des critères bio-physiques (SST, MLD, Chla).

La variabilité des groupes de phytoplancton estimés par la méthode des SOM (SOM-Pigments et SOM-PFT) a permis de mieux caractériser ces biorégions en rapport avec la complexité de la communauté phytoplanctonique. La dominance des groupes de Nanophytoplancton est observée en grande partie dans le bassin occidental (biorégion C5) pendant la période qui s'étend de l'automne jusqu'au printemps. Alors que la dominance des différents types de cyanobactéries *Synechococcus* et *Prochlorococcus* est mise en relief en été, et plus précisément dans les eaux de la biorégion (C6) du bassin oriental, les diatomées dominent tout au long de l'année au niveau des régions côtières et estuariennes constituant la biorégion (C7). Ceci est expliqué par l'apport de la silice terrigène indispensable à la prolifération de ce type de phytoplancton. De même les diatomées bénéficient en grande partie des conséquences de la forte convection profonde au niveau de la biorégion (C3) de la mer Baléare, marquée par un bloom important à la fin de la convection hivernale en mars.

Ces différentes études ont permis de montrer les caractéristiques et la variabilité des communautés phytoplanctoniques à partir de leur composition pigmentaire, pour arriver à

l'étude des groupes ou des classes de taille de phytoplancton. Les méthodes mises au point au cours de cette thèse offrent une plateforme diversifiée pour l'étude du phytoplancton, à des échelles temporelles et spatiales variées, qui sont applicables au niveau de l'océan global aussi bien que sur des régions océaniques ciblées

A ce niveau, plusieurs perspectives peuvent être suggérées à la lumière de ces méthodes ;

- la valorisation de la base de données d'images de l'océan global des pigments secondaires estimées à travers la SOM-Pigments pour les études de la biodiversité marine. En effet, plus de pigments secondaires autres que les neufs utilisés peuvent y être incorporés et analysés. Ceci permettra de déterminer potentiellement plus de groupes de phytoplancton, ou bien de raffiner la représentation de la diversité de la communauté phytoplanctonique.
- l'application de telles méthodes sur les données de très haute résolution telles que les images de Sentinel 2, pour une étude fine de l'évolution des structures hydrodynamiques en utilisant le phytoplancton comme traceur biologique des courants.
- Utilisation des données HPLC dans l'océan australe pour développer un algorithme régionalisé et robuste pour l'étude du phytoplancton de cette vaste région, qui est mal connue et qui fait l'objet de nombreux questionnements tant du point de vue biologique que méthodologique.

Références:

- [1] Abboud-Abi Saab, M. (1985) 'Etude quantitative et qualitative du phytoplancton des eaux cotières libanaises', *Lebanese Science Bulletin*, 1(2), pp. 197–222.
- [2] Abdel-Moati, A. R. (1990) 'Particulate organic matter in the subsurface chlorophyll maximum layer of the Southeastern Mediterranean', *Oceanologica Acta*, 13(3), pp. 307–315. Available at: <https://archimer.ifremer.fr/doc/00131/24250/22243.pdf> (Accessed: 11 February 2019).
- [3] Aiken, J., Pradhan, Y., Barlow, R., Lavender, S., Poulton, A. and Hardman-Mountford, N. (2009) 'Phytoplankton pigments and functional types in the Atlantic Ocean: A decadal assessment, 1995–2005', *Deep Sea Research Part II: Topical Studies in Oceanography*. Pergamon, 56(15), pp. 899–917. doi: 10.1016/J.DSR2.2008.09.017.
- [4] Ainsworth, E. J. and Jones, I. S. F. (1999) 'Radiance spectra classification from the Ocean Color and Temperature Scanner on ADEOS', *IEEE Transactions on Geoscience and Remote Sensing*, 37(3), pp. 1645–1656. doi: 10.1109/36.763281.
- [5] Alvain, S., Moulin, C., Dandonneau, Y. and Bréon, F. M. (2005) 'Remote sensing of phytoplankton groups in case 1 waters from global SeaWiFS imagery', *Deep-Sea Research Part I: Oceanographic Research Papers*, 52(11), pp. 1989–2004. doi: 10.1016/j.dsr.2005.06.015.
- [6] Alvain, S., Moulin, C., Dandonneau, Y. and Loisel, H. (2008) 'Seasonal distribution and succession of dominant phytoplankton groups in the global ocean: A satellite view', *Global Biogeochemical Cycles*, 22(3), pp. 1–15. doi: 10.1029/2007GB003154.
- [7] Antoine, D. and Andr, J. (1995) 'Algal pigment distribution and primary production in the', 100.
- [8] Antoine, D., André, J.-M. and Morel, A. (1996) 'Oceanic primary production: 2. Estimation at global scale from satellite (Coastal Zone Color Scanner) chlorophyll', *Global Biogeochemical Cycles*. Wiley-Blackwell, 10(1), pp. 57–69. doi: 10.1029/95GB02832.
- [9] Arrigo, K. R., Worthen, D., Schnell, A. and Lizotte, M. P. (1998) 'Primary production in Southern Ocean waters', *Journal of Geophysical Research: Oceans*. Wiley-Blackwell, 103(C8), pp. 15587–15600. doi: 10.1029/98JC00930.
- [10] Artegiani, A., Paschini, E., Russo, A., Bregant, D., Raicich, F., Pinardi, N., Artegiani, A., Paschini, E., Russo, A., Bregant, D., Raicich, F. and Pinardi, N. (1997a) 'The Adriatic Sea General Circulation. Part I: Air–Sea Interactions and Water Mass Structure', *Journal of Physical Oceanography*, 27(8), pp. 1492–1514. doi: 10.1175/1520-0485(1997)027<1492:TASGCP>2.0.CO;2.
- [11] Artegiani, A., Paschini, E., Russo, A., Bregant, D., Raicich, F., Pinardi, N., Artegiani, A., Paschini, E., Russo, A., Bregant, D., Raicich, F. and Pinardi, N. (1997b) 'The Adriatic Sea General Circulation. Part II: Baroclinic Circulation Structure', *Journal of Physical Oceanography*, 27(8), pp. 1515–1532. doi: 10.1175/1520-0485(1997)027<1515:TASGCP>2.0.CO;2.
- [12] Barale, V., Jaquet, J. M. and Ndiaye, M. (2008) 'Algal blooming patterns and anomalies in the Mediterranean Sea as derived from the SeaWiFS data set (1998–2003)', *Remote Sensing of Environment*, 112(8), pp. 3300–3313. doi: 10.1016/j.rse.2007.10.014.
- [13] Barber, R. T. and Chavez, F. P. (1991) 'Regulation of primary productivity rate

- in the equatorial Pacific', *Limnology and Oceanography*. Wiley-Blackwell, 36(8), pp. 1803–1815. doi: 10.4319/lo.1991.36.8.1803.
- [14] Baumert, H. Z. (2008) 'The role of temperature, cellular quota and nutrient concentrations for photosynthesis, growth and light–dark acclimation in phytoplankton', *Limnologica - Ecology and Management of Inland Waters*. Urban & Fischer, 38(3–4), pp. 313–326. doi: 10.1016/J.LIMNO.2008.06.002.
- [15] Behrenfeld, M. J. and Boss, E. (2006) 'Beam attenuation and chlorophyll concentration as alternative optical indices of phytoplankton biomass', *Journal of Marine Research*, 64(3), pp. 431–451. doi: 10.1357/002224006778189563.
- [16] Behrenfeld, M. J., Boss, E., Siegel, D. A. and Shea, D. M. (2005) 'Carbon-based ocean productivity and phytoplankton physiology from space', *Global Biogeochemical Cycles*. Wiley-Blackwell, 19(1). doi: 10.1029/2004GB002299.
- [17] Behrenfeld, M. J. and Falkowski, P. G. (1997) 'Photosynthetic rates derived from satellite-based chlorophyll concentration', *Limnology and Oceanography*. Wiley-Blackwell, 42(1), pp. 1–20. doi: 10.4319/lo.1997.42.1.0001.
- [18] Belo Couto, A., Brotas, V., Mélin, F., Groom, S. and Sathyendranath, S. (2016) 'Inter-comparison of OC-CCI chlorophyll- *a* estimates with precursor data sets', *International Journal of Remote Sensing*. Taylor & Francis, 37(18), pp. 4337–4355. doi: 10.1080/01431161.2016.1209313.
- [19] Béranger, K., Mortier, L., Gasparini, G.-P., Gervasio, L., Astraldi, M. and Crépon, M. (2004) 'The dynamics of the Sicily Strait: a comprehensive study from observations and models', *Deep Sea Research Part II: Topical Studies in Oceanography*. Pergamon, 51(4–5), pp. 411–440. doi: 10.1016/J.DSR2.2003.08.004.
- [20] Bergamasco, A. and Malanotte-Rizzoli, P. (2010) 'The circulation of the Mediterranean Sea: a historical review of experimental investigations', *Advances in Oceanography and Limnology*. Taylor & Francis Group , 1(1), pp. 11–28. doi: 10.1080/19475721.2010.491656.
- [21] Berline, L., Rammou, A.-M., Doglioli, A., Molcard, A. and Petrenko, A. (2014) 'A Connectivity-Based Eco-Regionalization Method of the Mediterranean Sea', *PLoS ONE*. Edited by B. R. MacKenzie. Public Library of Science, 9(11), p. e111978. doi: 10.1371/journal.pone.0111978.
- [22] Bethoux, J. ., Gentili, B., Morin, P., Nicolas, E., Pierre, C. and Ruiz-Pino, D. (1999) 'The Mediterranean Sea: a miniature ocean for climatic and environmental studies and a key for the climatic functioning of the North Atlantic', *Progress in Oceanography*. Pergamon, 44(1–3), pp. 131–146. doi: 10.1016/S0079-6611(99)00023-3.
- [23] Bethoux, J. P. (1980) 'Mean water fluxes across sections in the Mediterranean Sea, evaluation on the basis of water and salt budgets and of observed salinities', *Oceanologica Acta*, 3(1), pp. 79–88. Available at: <https://archimer.ifremer.fr/doc/00122/23298/21125.pdf> (Accessed: 9 February 2019).
- [24] Bhadury, P. (2015) 'Effects of ocean acidification on marine invertebrates-a review', *Indian Journal of Geo-Marine Sciences*, 44(4), pp. 454–464. Available at: [http://nopr.niscair.res.in/bitstream/123456789/34717/1/IJMS_44\(4\)_454-464.pdf](http://nopr.niscair.res.in/bitstream/123456789/34717/1/IJMS_44(4)_454-464.pdf) (Accessed: 6 June 2018).
- [25] Bidigare, R. R., Ondrusek, M. E., Morrow, J. H. and Kiefer, D. A. (1990) 'In-vivo absorption properties of algal pigments', *Proc. SPIE 1302, Ocean Optics X*, . Edited by R. W. Spinrad, 1302, p. 290. doi: 10.1117/12.21451.

- [26] Blain, S., Quéguiner, B. and Trull, T. (2008) 'The natural iron fertilization experiment KEOPS (KErguelen Ocean and Plateau compared Study): An overview', *Deep Sea Research Part II: Topical Studies in Oceanography*, 55(5–7), pp. 559–565. doi: 10.1016/j.dsr2.2008.01.002.
- [27] Bosc, E., Bricaud, A. and Antoine, D. (2004) 'Seasonal and interannual variability in algal biomass and primary production in the Mediterranean Sea, as derived from 4 years of SeaWiFS observations', 18, pp. 1–17. doi: 10.1029/2003GB002034.
- [28] Bracher, Astrid; Taylor, Marc H; Taylor, Bettina B; Dinter, Tilman; Röttgers, Rüdiger; Steinmetz, F. (2015) 'Phytoplankton pigments, hyperspectral downwelling irradiance and remote sensing reflectance during POLARSTERN cruises ANT-XXIII/1, ANT-XXIV/1, ANT-XXIV/4, ANT-XXVI/4, and Maria S. Merian cruise MSM18/3.', *PANGAEA*. doi: <https://doi.org/10.1594/PANGAEA.847820>.
- [29] Bracher, A., Bouman, H. A., Brewin, R. J. W., Bricaud, A., Brotas, V., Ciotti, A. M., Clementson, L., Devred, E., Di Cicco, A., Dutkiewicz, S., Hardman-Mountford, N. J., Hickman, A. E., Hieronymi, M., Hirata, T., Losa, S. N., Mouw, C. B., Organelli, E., Raitsos, D. E., Uitz, J., Vogt, M. and Wolanin, A. (2017a) 'Obtaining Phytoplankton Diversity from Ocean Color: A Scientific Roadmap for Future Development', *Frontiers in Marine Science*, 4(March), pp. 1–15. doi: 10.3389/fmars.2017.00055.
- [30] Bracher, A., Bouman, H. A., Brewin, R. J. W., Bricaud, A., Brotas, V., Ciotti, A. M., Clementson, L., Devred, E., Di Cicco, A., Dutkiewicz, S., Hardman-Mountford, N. J., Hickman, A. E., Hieronymi, M., Hirata, T., Losa, S. N., Mouw, C. B., Organelli, E., Raitsos, D. E., Uitz, J., Vogt, M. and Wolanin, A. (2017b) 'Obtaining Phytoplankton Diversity from Ocean Color: A Scientific Roadmap for Future Development', *Frontiers in Marine Science*. Frontiers, 4, p. 55. doi: 10.3389/fmars.2017.00055.
- [31] Bracher, A., Taylor, M. H., Taylor, B., Dinter, T., Röttgers, R. and Steinmetz, F. (2015) 'Using empirical orthogonal functions derived from remote-sensing reflectance for the prediction of phytoplankton pigment concentrations', *Ocean Science*, 11(1), pp. 139–158. doi: 10.5194/os-11-139-2015.
- [32] Bracher, A., Vountas, M., Dinter, T., Burrows, J. P., Röttgers, R. and Peeken, I. (2009) *Quantitative observation of cyanobacteria and diatoms from space using PhytoDOAS on SCIAMACHY data*, *Biogeosciences*. Available at: www.biogeosciences.net/6/751/2009/ (Accessed: 11 February 2019).
- [33] Brewin, R. J. W., Devred, E., Sathyendranath, S., Lavender, S. J. and Hardman-Mountford, N. J. (2011) 'Model of phytoplankton absorption based on three size classes', *Applied Optics*, 50(22), p. 4535. doi: 10.1364/AO.50.004535.
- [34] Brewin, R. J. W., Sathyendranath, S., Hirata, T., Lavender, S. J., Barciela, R. M. and Hardman-Mountford, N. J. (2010) 'A three-component model of phytoplankton size class for the Atlantic Ocean', *Ecological Modelling*. Elsevier, 221(11), pp. 1472–1483. doi: 10.1016/J.ECOLMODEL.2010.02.014.
- [35] Bricaud, A., Ciotti, A. M. and Gentili, B. (2012) 'Spatial-temporal variations in phytoplankton size and colored detrital matter absorption at global and regional scales, as derived from twelve years of SeaWiFS data (1998–2009)', *Global Biogeochemical Cycles*. John Wiley & Sons, Ltd, 26(1), p. n/a-n/a. doi: 10.1029/2010GB003952.
- [36] Bricaud, A., Thiria, S., Chazottes, A. and Crépon, M. (2006) 'Statistical analysis of a database of absorption spectra of phytoplankton and pigment concentrations using self-organizing maps View project Statistical analysis of a database of absorption

- spectra of phytoplankton and pigment concentrations using self-organizing maps'. doi: 10.1364/AO.45.008102.
- [37] Bricaud, A., Thiria, S., Chazottes, A., Crépon, M. and Ras, J. (2007) 'Statistical analysis of absorption spectra of phytoplankton and of pigment concentrations observed during three POMME cruises using a neural network clustering method Astico View project ESA SEOM project SynSenPFT View project Statistical analysis of absorption spectra of phytoplankton and of pigment concentrations observed during three POMME cruises using a neural network clustering method'. doi: 10.1364/AO.46.003790.
- [38] Brunet, C., Casotti, R., Vantrepotte, V., Corato, F. and Conversano, F. (2006) 'Picophytoplankton diversity and photoacclimation in the Strait of Sicily (Mediterranean Sea) in summer. I. Mesoscale variations', *Aquatic Microbial Ecology*, 44(2), pp. 127–141. doi: 10.3354/ame044127.
- [39] Calvo-Díaz, A., Morán, X. A. G. and Suárez, L. Á. (2008) 'Seasonality of picophytoplankton chlorophyll a and biomass in the central Cantabrian Sea, southern Bay of Biscay', *Journal of Marine Systems*, 72(1–4), pp. 271–281. doi: 10.1016/j.jmarsys.2007.03.008.
- [40] Casey, K. S., Brandon, T. B., Cornillon, P. and Evans, R. (2010) 'The Past, Present, and Future of the AVHRR Pathfinder SST Program', in *Oceanography from Space*. Dordrecht: Springer Netherlands, pp. 273–287. doi: 10.1007/978-90-481-8681-5_16.
- [41] Cavazos, T. and Cavazos, T. (1999) 'Large-Scale Circulation Anomalies Conducive to Extreme Precipitation Events and Derivation of Daily Rainfall in Northeastern Mexico and Southeastern Texas', *Journal of Climate*, 12(5), pp. 1506–1523. doi: 10.1175/1520-0442(1999)012<1506:LSCACT>2.0.CO;2.
- [42] Cavender-Bares, K. K., Karl, D. M. and Chisholm, S. W. (2001) 'Nutrient gradients in the western North Atlantic Ocean: Relationship to microbial community structure and comparison to patterns in the Pacific Ocean', *Deep Sea Research Part I: Oceanographic Research Papers*. Pergamon, 48(11), pp. 2373–2395. doi: 10.1016/S0967-0637(01)00027-9.
- [43] Chase, A., Boss, E., Zaneveld, R., Bricaud, A., Claustre, H., Ras, J., Dall'Olmo, G. and Westberry, T. K. (2013) 'Decomposition of in situ particulate absorption spectra', *Methods in Oceanography*. Elsevier Ltd, 7(2013), pp. 110–124. doi: 10.1016/j.mio.2014.02.002.
- [44] Chase, A. P., Boss, E., Cetinić, I. and Slade, W. (2017) 'Estimation of Phytoplankton Accessory Pigments From Hyperspectral Reflectance Spectra: Toward a Global Algorithm', *Journal of Geophysical Research: Oceans*, 122(12), pp. 9725–9743. doi: 10.1002/2017JC012859.
- [45] Chavez, F. P. (1996) 'Forcing and biological impact of onset of the 1992 El Niño in central California', *Geophysical Research Letters*. Wiley-Blackwell, 23(3), pp. 265–268. doi: 10.1029/96GL00017.
- [46] Chavez, F. P. and Barber, R. T. (1987) 'An estimate of new production in the equatorial Pacific', *Deep Sea Research Part A. Oceanographic Research Papers*. Elsevier, 34(7), pp. 1229–1243. doi: 10.1016/0198-0149(87)90073-2.
- [47] Chazottes, A., Bricaud, A., Crépon, M. and Thiria, S. (2006) 'Statistical analysis of a database of absorption spectra of phytoplankton and pigment concentrations using self-organizing maps'.
- [48] Chisholm, S. W. (1995) 'The iron hypothesis: Basic research meets

- environmental policy', *Reviews of Geophysics*. Wiley-Blackwell, 33(S2), pp. 1277–1286. doi: 10.1029/95RG00743.
- [49] Di Cicco, A., Sammartino, M., Marullo, S. and Santoleri, R. (2017) 'Regional Empirical Algorithms for an Improved Identification of Phytoplankton Functional Types and Size Classes in the Mediterranean Sea Using Satellite Data', *Frontiers in Marine Science*. Frontiers, 4, p. 126. doi: 10.3389/fmars.2017.00126.
- [50] Ciotti, A. M. and Bricaud, A. (2006) 'Retrievals of a size parameter for phytoplankton and spectral light absorption by colored detrital matter from water-leaving radiances at SeaWiFS channels in a continental shelf region off Brazil', *Limnology and Oceanography: Methods*, 4, pp. 237–253. doi: 10.4319/lom.2006.4.237.
- [51] Coale, K. H., Johnson, K. S., Fitzwater, S. E., Gordon, R. M., Tanner, S., Chavez, F. P., Ferioli, L., Sakamoto, C., Rogers, P., Millero, F., Steinberg, P., Nightingale, P., Cooper, D., Cochlan, W. P., Landry, M. R., Constantinou, J., Rollwagen, G., Trasvina, A. and Kudela, R. (1996) 'A massive phytoplankton bloom induced by an ecosystem-scale iron fertilization experiment in the equatorial Pacific Ocean', *Nature*. Nature Publishing Group, 383(6600), pp. 495–501. doi: 10.1038/383495a0.
- [52] Coll, M., Piroddi, C., Steenbeek, J., Kaschner, K., Ben Rais Lasram, F., Aguzzi, J., Ballesteros, E., Bianchi, C. N., Corbera, J., Dailianis, T., Danovaro, R., Estrada, M., Frogli, C., Galil, B. S., Gasol, J. M., Gertwagen, R., Gil, J., Guilhaumon, F., Kesner-Reyes, K., Kitsos, M.-S., Koukouras, A., Lampadariou, N., Laxamana, E., López-Fé de la Cuadra, C. M., Lotze, H. K., Martin, D., Mouillot, D., Oro, D., Raicevich, S., Rius-Barile, J., Saiz-Salinas, J. I., San Vicente, C., Somot, S., Templado, J., Turon, X., Vafidis, D., Villanueva, R. and Voultsiadou, E. (2010) 'The Biodiversity of the Mediterranean Sea: Estimates, Patterns, and Threats', *PLoS ONE*. Edited by S. J. Bograd. Public Library of Science, 5(8), p. e11842. doi: 10.1371/journal.pone.0011842.
- [53] Collier, J. L. (2004) 'FLOW CYTOMETRY AND THE SINGLE COMPOUND IN PLANKTON ECOLOGY', *Journal of Phycology*. John Wiley & Sons, Ltd (10.1111), 40(5), pp. 805–807. doi: 10.1111/j.1529-8817.2004.40501.x.
- [54] D'Ortenzio, F., Iudicone, D., de Boyer Montegut, C., Testor, P., Antoine, D., Marullo, S., Santoleri, R. and Madec, G. (2005) 'Seasonal variability of the mixed layer depth in the Mediterranean Sea as derived from in situ profiles', *Geophysical Research Letters*. John Wiley & Sons, Ltd, 32(12), p. n/a-n/a. doi: 10.1029/2005GL022463.
- [55] D'Ortenzio, F. and Ribera d'Alcalà, M. (2008) 'On the trophic regimes of the Mediterranean Sea: a satellite analysis', *Biogeosciences Discussions*, 5(4), pp. 2959–2983. doi: 10.5194/bgd-5-2959-2008.
- [56] Dandonneau, Y., Deschamps, P.-Y., Nicolas, J.-M., Loisel, H., Blanchot, J., Montel, Y., Thieuleux, F. and Bécu, G. (2004) 'Seasonal and interannual variability of ocean color and composition of phytoplankton communities in the North Atlantic, equatorial Pacific and South Pacific', *Deep Sea Research Part II: Topical Studies in Oceanography*, 51(1–3), pp. 303–318. doi: 10.1016/j.dsr2.2003.07.018.
- [57] Dandonneau, Y., Montel, Y., Blanchot, J., Giraudeau, J. and Neveux, J. (2006) 'Temporal variability in phytoplankton pigments, picoplankton and coccolithophores along a transect through the North Atlantic and tropical southwestern Pacific', *Deep-Sea Research Part I: Oceanographic Research Papers*, 53(4), pp. 689–712. doi: 10.1016/j.dsr.2006.01.002.
- [58] Débora Iglesias-Rodríguez, M., Schofield, O. M., Batley, J., Medlin, L. K. and

- Hayes, P. K. (2006) 'Intraspecific genetic diversity in the marine coccolithophore *Emiliana Huxleyi* (Prymnesiophyceae): The use of microsatellite analysis in marine phytoplankton population studies', *J. Phycol.*, (42), pp. 526–536. doi: 10.1111/j.1529-8817.2006.00231.x.
- [59] Devred, E., Sathyendranath, S., Stuart, V., Maass, H., Ulloa, O. and Platt, T. (2006) 'A two-component model of phytoplankton absorption in the open ocean: Theory and applications', *Journal of Geophysical Research*. Wiley-Blackwell, 111(C3), p. C03011. doi: 10.1029/2005JC002880.
- [60] Devred, E., Sathyendranath, S., Stuart, V. and Platt, T. (2011) 'A three component classification of phytoplankton absorption spectra: Application to ocean-color data', *Remote Sensing of Environment*. Elsevier, 115(9), pp. 2255–2266. doi: 10.1016/J.RSE.2011.04.025.
- [61] Diouf, D., Niang, A., Brajard, J., Crepon, M. and Thiria, S. (2013) 'Retrieving aerosol characteristics and sea-surface chlorophyll from satellite ocean color multi-spectral sensors using a neural-variational method', *Remote Sensing of Environment*. Elsevier Inc., 130, pp. 74–86. doi: 10.1016/j.rse.2012.11.002.
- [62] Dowidar, N. M. (1984) 'Phytoplankton biomass and primary productivity of the south-eastern Mediterranean', *Deep Sea Research Part A. Oceanographic Research Papers*. Elsevier, 31(6–8), pp. 983–1000. doi: 10.1016/0198-0149(84)90052-9.
- [63] Dubelaar, G. B. J., Gerritzen, P. L. and Gerritzen, P. L. (2000) 'CytoBuoy: a step forward towards using flow cytometry in operational oceanography', *Scientia Marina*, 64(2), pp. 255–265. doi: 10.3989/scimar.2000.64n2255.
- [64] Dugdale, R. C. and Wilkerson, F. P. (1998) 'Silicate regulation of new production in the equatorial Pacific upwelling', *Nature*, 391(6664), pp. 270–273. doi: 10.1038/34630.
- [65] Egge, J. K. and Aksnes, D. L. (1992) 'Silicate as regulating nutrient in phytoplankton competition', *Marine Ecology Progress Series*, 83, pp. 281–289. Available at: <https://www.int-res.com/articles/meps/83/m083p281.pdf> (Accessed: 11 February 2019).
- [66] Ehsani, A. H. and Quiel, F. (2008) 'Geomorphometric feature analysis using morphometric parameterization and artificial neural networks', *Geomorphology*, 99(1–4), pp. 1–12. doi: 10.1016/j.geomorph.2007.10.002.
- [67] Emery, W. J., Yu, Y., Wick, G. A., Schluessel, P. and Reynolds, R. W. (1994) 'Correcting infrared satellite estimates of sea surface temperature for atmospheric water vapor attenuation', *Journal of Geophysical Research*. Wiley-Blackwell, 99(C3), p. 5219. doi: 10.1029/93JC03215.
- [68] Falkowski, P. G., Laws, E. A., Barber, R. T. and Murray, J. W. (2003) 'Phytoplankton and Their Role in Primary, New, and Export Production', in *Ocean Biogeochemistry*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 99–121. doi: 10.1007/978-3-642-55844-3_5.
- [69] Fenton, N., Priddle, J. and Tett, P. (1994) 'Regional variations in bio-optical properties of the surface waters in the Southern Ocean', *Antarctic Science*. Cambridge University Press, 6(04), pp. 443–448. doi: 10.1017/S0954102094000684.
- [70] Flombaum, P., Gallegos, J. L., Gordillo, R. A., Rincon, J., Zabala, L. L., Jiao, N., Karl, D. M., Li, W. K. W., Lomas, M. W., Veneziano, D., Vera, C. S., Vrugt, J. A. and Martiny, A. C. (2013) 'Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*', *Proceedings of the National*

- Academy of Sciences*, 110(24), pp. 9824–9829. doi: 10.1073/pnas.1307701110.
- [71] FOGG, G. E. (1991) 'The phytoplanktonic ways of life', *New Phytologist*. Wiley/Blackwell (10.1111), 118(2), pp. 191–232. doi: 10.1111/j.1469-8137.1991.tb00974.x.
- [72] Fragoso, G. M., Poulton, A. J., Yashayaev, I. M., Head, E. J. H. and Purdie, D. A. (2016) 'Spring phytoplankton communities of the Labrador Sea (2005-2014): pigment signatures, photophysiology and elemental ratios', *Biogeosciences Discussions*, pp. 1–43. doi: 10.5194/bg-2016-295.
- [73] Fuda, J.-L., Etiope, G., Millot, C., Favali, P., Calcara, M., Smriglio, G. and Boschi, E. (2002) 'Warming, salting and origin of the Tyrrhenian Deep Water', *Geophysical Research Letters*. John Wiley & Sons, Ltd, 29(19), pp. 4-1-4–4. doi: 10.1029/2001GL014072.
- [74] Fujiwara, A., Hirawake, T., Suzuki, K. and Saitoh, S.-I. (2011) 'Remote sensing of size structure of phytoplankton communities using optical properties of the Chukchi and Bering Sea shelf region', *Biogeosciences*, 8(12), pp. 3567–3580. doi: 10.5194/bg-8-3567-2011.
- [75] Gerin, R., Poulain, P., Millot, C., Ismail, S. Ben, Sammari, C., Grotta, B., Trieste, S., National, I. and Mer, D. (2009) 'Surface Circulation in the Eastern Mediterranean Using Lagrangian Drifters', *Circulation*, (June 1986), pp. 83507–83507. doi: 10.5194/osd-6-525-2009.
- [76] Gieskes, W. W. C. and Kraay, G. W. (1983) 'Dominance of Cryptophyceae during the phytoplankton spring bloom in the central North Sea detected by HPLC analysis of pigments', *Marine Biology*. Springer-Verlag, 75(2–3), pp. 179–185. doi: 10.1007/BF00406000.
- [77] Gohin, F. (2011) 'Annual cycles of chlorophyll-a, non-algal suspended particulate matter, and turbidity observed from space and in-situ in coastal waters', *Ocean Science*, 7(5), pp. 705–732. doi: 10.5194/os-7-705-2011.
- [78] Gordon, H. R. (1997) 'Atmospheric correction of ocean color imagery in the Earth Observation System era', *Journal of Geophysical Research*, 102(D14), pp. 17081–17106.
- [79] Gorricha, J. and Lobo, V. (2012) 'Improvements on the visualization of clusters in geo-referenced data using Self-Organizing Maps', *Computers and Geosciences*, 43, pp. 177–186. doi: 10.1016/j.cageo.2011.10.008.
- [80] Gower, J., Young, E. and King, S. (2013) 'Satellite images suggest a new Sargassum source region in 2011', *Remote Sensing Letters*. Taylor & Francis, 4(8), pp. 764–773. doi: 10.1080/2150704X.2013.796433.
- [81] Greene, R. M., Kolber, Z. S., Swift, D. G., Tindale, N. W. and Falkowski, P. G. (1994) 'Physiological limitation of phytoplankton photosynthesis in the eastern equatorial Pacific determined from variability in the quantum yield of fluorescence', *Limnology and Oceanography*. John Wiley & Sons, Ltd, 39(5), pp. 1061–1074. doi: 10.4319/lo.1994.39.5.1061.
- [82] Guenard, V., Drobinski, P., Caccia, J.-L., Campistron, B. and Bench, B. (2005) 'An Observational Study of the Mesoscale Mistral Dynamics', *Boundary-Layer Meteorology*. Kluwer Academic Publishers, 115(2), pp. 263–288. doi: 10.1007/s10546-004-3406-z.
- [83] Guillard, R. R. L., Murphy, L. S., Foss, P. and Liaaen-Jensen, S. (1985) 'Synechococcus spp. as likely zeaxanthin-dominant ultraphytoplankton in the North

- Atlantic1', *Limnology and Oceanography*. Wiley-Blackwell, 30(2), pp. 412–414. doi: 10.4319/lo.1985.30.2.0412.
- [84] Häder, D., Häder, D.-P., Villafañe, V. E. and Helbling, E. W. (2014) 'Productivity of aquatic primary producers under global climate change Photochemical & Photobiological Sciences PERSPECTIVE Productivity of aquatic primary producers under global climate change', *Photochemical & Photobiological Sciences*, 13. doi: 10.1039/c3pp50418b.
- [85] Hamad, N., Millot, C. and Taupier-Letage, I. (2005) 'A new hypothesis about the surface circulation in the eastern basin of the mediterranean sea', *Progress in Oceanography*. Pergamon, 66(2–4), pp. 287–298. doi: 10.1016/J.POCEAN.2005.04.002.
- [86] Hedges, J. I. and Keil, R. G. (1995) 'Sedimentary organic matter preservation: an assessment and speculative synthesis', *Marine Chemistry*. Elsevier, 49(2–3), pp. 81–115. doi: 10.1016/0304-4203(95)00008-F.
- [87] Hewitson, B. C. and Crane, R. G. (2002) 'Self-organizing maps: applications to synoptic climatology', *Climate Research*. Inter-Research Science Center, pp. 13–26. doi: 10.2307/24868302.
- [88] Hirata, T., Aiken, J., Hardman-Mountford, N., Smyth, T. J. and Barlow, R. G. (2008) 'An absorption model to determine phytoplankton size classes from satellite ocean colour', *Remote Sensing of Environment*. Elsevier, 112(6), pp. 3153–3159. doi: 10.1016/J.RSE.2008.03.011.
- [89] Hirata, T., Hardman-Mountford, N. J., Brewin, R. J. W., Aiken, J., Barlow, R., Suzuki, K., Isada, T., Howell, E., Hashioka, T., Noguchi-Aita, M. and Yamanaka, Y. (2011) 'Synoptic relationships between surface Chlorophyll-a and diagnostic pigments specific to phytoplankton functional types', *Biogeosciences*, 8(2), pp. 311–327. doi: 10.5194/bg-8-311-2011.
- [90] Hoepffner, N. and Sathyendranath, S. (1991) 'Effect of pigment composition on absorption properties of phytoplankton', *Marine Ecology Progress Series*, 73, pp. 11–23. Available at: <https://www.int-res.com/articles/meps/73/m073p011.pdf> (Accessed: 11 February 2019).
- [91] Holligan, P. M., Groom, S. B. and Harbour, D. S. (1993) 'What controls the distribution of the coccolithophore, *Emiliania huxleyi*, in the North Sea?', *Fisheries Oceanography*. Wiley/Blackwell (10.1111), 2(3–4), pp. 175–183. doi: 10.1111/j.1365-2419.1993.tb00133.x.
- [92] Holte, J., Talley, L. D., Gilson, J. and Roemmich, D. (2017) 'An Argo mixed layer climatology and database', *Geophysical Research Letters*. John Wiley & Sons, Ltd, 44(11), pp. 5618–5626. doi: 10.1002/2017GL073426.
- [93] Holte, J., Talley, L., Holte, J. and Talley, L. (2009) 'A New Algorithm for Finding Mixed Layer Depths with Applications to Argo Data and Subantarctic Mode Water Formation*', *Journal of Atmospheric and Oceanic Technology*, 26(9), pp. 1920–1939. doi: 10.1175/2009JTECHO543.1.
- [94] El Hourany, R., Abboud-Abi Saab, M., Faour, G., Aumont, O., Crépon, M. and Thiria, S. (2019) 'Estimation of secondary phytoplankton pigments from satellite observations using self-organizing maps (SOM)', *Journal of Geophysical Research: Oceans*. John Wiley & Sons, Ltd. doi: 10.1029/2018JC014450.
- [95] Hu, C., Lee, Z. and Franz, B. (2012) 'Chlorophyll a algorithms for oligotrophic oceans: A novel approach based on three-band reflectance difference', *J. Geophys.*

- Res*, 117. doi: 10.1029/2011JC007395.
- [96] Hu, X. and Weng, Q. (2009) 'Estimating impervious surfaces from medium spatial resolution imagery using the self-organizing map and multi-layer perceptron neural networks', *Remote Sensing of Environment*. Elsevier, 113(10), pp. 2089–2102. doi: 10.1016/J.RSE.2009.05.014.
- [97] Hülse, D., Arndt, S., Wilson, J. D., Munhoven, G. and Ridgwell, A. (2017) 'Understanding the causes and consequences of past marine carbon cycling variability through models', *Earth-Science Reviews*. Elsevier, 171, pp. 349–382. doi: 10.1016/J.EARSCIREV.2017.06.004.
- [98] Iglesias-Rodríguez, M. D., Brown, C. W., Doney, S. C., Kleypas, J., Kolber, D., Kolber, Z., Hayes, P. K. and Falkowski, P. G. (2002) 'Representing key phytoplankton functional groups in ocean carbon cycle models: Coccolithophorids', *Global Biogeochemical Cycles*. Wiley-Blackwell, 16(4), pp. 47-1-47–20. doi: 10.1029/2001GB001454.
- [99] Ignatiades, L., Gotsis-Skretas, O., Pagou, K. and Krasakopoulou, E. (2009) 'Diversification of phytoplankton community structure and related parameters along a large-scale longitudinal east-west transect of the Mediterranean Sea', *Journal of Plankton Research*. Oxford University Press, 31(4), pp. 411–428. doi: 10.1093/plankt/fbn124.
- [100] Iskandar, I. and Iskhaq (2010) 'VARIABILITY OF SATELLITE-OBSERVED SEA SURFACE HEIGHT IN THE TROPICAL INDIAN OCEAN: COMPARISON OF EOF AND SOM ANALYSIS', *MAKARA of Science Series*, 13(2). doi: 10.7454/mss.v13i2.421.
- [101] Jansá, A. (1987) 'Distribution of the Mistral: A satellite observation', *Meteorology and Atmospheric Physics*. Springer-Verlag, 36(1–4), pp. 201–214. doi: 10.1007/BF01045149.
- [102] Jeffrey, S. and Mantoura, R. (1997) 'Development of pigment methods for oceanography: SCOR-supported working groups and objectives'.
- [103] Jeffrey, S. W. (1980) 'Algal Pigment Systems', in *Primary Productivity in the Sea*. Boston, MA: Springer US, pp. 33–58. doi: 10.1007/978-1-4684-3890-1_3.
- [104] Jeffrey, S. W. and Hallegraeff, G. M. (1987) 'Chlorophyllase distribution in ten classes of phytoplankton: a problem for chlorophyll analysis', *Marine Ecology Progress Series*. Inter-Research Science Center, pp. 293–304. doi: 10.2307/24825001.
- [105] Jiang, Q., Smith, R. B. and Doyle, J. (2003) 'The nature of the mistral: Observations and modelling of two MAP events', *Quarterly Journal of the Royal Meteorological Society*. John Wiley & Sons, Ltd, 129(588), pp. 857–875. doi: 10.1256/qj.02.21.
- [106] Jonasz, M. and Fournier, G. R. (2007) *Light Scattering by Particles in Water, Light Scattering by Particles in Water*. Elsevier/Academic Press. doi: 10.1016/B978-012388751-1/50006-5.
- [107] Jouini, M., Lévy, M., Crépon, M. and Thiria, S. (2013) 'Reconstruction of satellite chlorophyll images under heavy cloud coverage using a neural classification method', *Remote Sensing of Environment*. Elsevier, 131, pp. 232–246. doi: 10.1016/J.RSE.2012.11.025.
- [108] Jurčec, V. (1981) 'On Mesoscale Characteristics of Bora Conditions in Yugoslavia', in *Weather and Weather Maps*. Basel: Birkhäuser Basel, pp. 640–657. doi: 10.1007/978-3-0348-5148-0_15.
- [109] Kobayashi, F. and Takahashi, K. (2002) 'Distribution of diatoms along the

- equatorial transect in the western and central Pacific during the 1999 La Niña conditions', *Deep Sea Research Part II: Topical Studies in Oceanography*. Pergamon, 49(13–14), pp. 2801–2821. doi: 10.1016/S0967-0645(02)00059-0.
- [110] Kohonen, T. (2013) 'Essentials of the self-organizing map', *Neural Networks*. Pergamon, 37, pp. 52–65. doi: 10.1016/J.NEUNET.2012.09.018.
- [111] Korb, R., Whitehouse, M., Atkinson, A. and Thorpe, S. (2008) 'Magnitude and maintenance of the phytoplankton bloom at South Georgia: a naturally iron-replete environment', *Marine Ecology Progress Series*, 368, pp. 75–91. doi: 10.3354/meps07525.
- [112] Kostadinov, T. S., Cabré, A., Vedantham, H., Marinov, I., Bracher, A., Brewin, R. J. W., Bricaud, A., Hirata, T., Hirawake, T., Hardman-Mountford, N. J., Mouw, C., Roy, S. and Uitz, J. (2017) 'Inter-comparison of phytoplankton functional type phenology metrics derived from ocean color algorithms and Earth System Models', *Remote Sensing of Environment*. Elsevier, 190, pp. 162–177. doi: 10.1016/J.RSE.2016.11.014.
- [113] Kostadinov, T. S., Siegel, D. A. and Maritorena, S. (2009) 'Retrieval of the particle size distribution from satellite ocean color observations', *Journal of Geophysical Research*. John Wiley & Sons, Ltd, 114(C9), p. C09015. doi: 10.1029/2009JC005303.
- [114] Kowalczyk, P., Cooper, W. J., Durako, M. J., Kahn, A. E., Gonsior, M. and Young, H. (2010) 'Characterization of dissolved organic matter fluorescence in the South Atlantic Bight with use of PARAFAC model: Relationships between fluorescence and its components, absorption coefficients and organic carbon concentrations', *Marine Chemistry*. Elsevier, 118(1–2), pp. 22–36. doi: 10.1016/J.MARCHEM.2009.10.002.
- [115] Krom, M. D., Emeis, K. C. and Van Cappellen, P. (2010) 'Why is the Eastern Mediterranean phosphorus limited?', *Progress in Oceanography*. Elsevier Ltd, 85(3–4), pp. 236–244. doi: 10.1016/j.pocean.2010.03.003.
- [116] Krom, M. D., Herut, B. and Mantoura, R. F. C. (2004) 'Nutrient budget for the Eastern Mediterranean: Implications for phosphorus limitation', *Limnology and Oceanography*, 49(5), pp. 1582–1592. doi: 10.4319/lo.2004.49.5.1582.
- [117] Krom, M. D., Woodward, E. M. S., Herut, B., Kress, N., Carbo, P., Mantoura, R. F. C., Spyres, G., Thingsted, T. F., Wassmann, P., Wexels-Riser, C., Kitidis, V., Law, C. and Zodiatis, G. (2005) 'Nutrient cycling in the south east Levantine basin of the eastern Mediterranean: Results from a phosphorus starved system', *Deep-Sea Research Part II: Topical Studies in Oceanography*, 52(22–23), pp. 2879–2896. doi: 10.1016/j.dsr2.2005.08.009.
- [118] Lacombe, H., Gascard, J. C., Gonella, J. and Béthoux, J. P. (1981) 'Response of the Mediterranean to the water and energy fluxes across its surface, on seasonal and interannual scales', *Oceanologica Acta*, 4(2), pp. 247–255. Available at: <https://archimer.ifremer.fr/doc/00121/23209/21054.pdf>.
- [119] Lascaratos, A. (1993) 'Estimation of deep and intermediate water mass formation rates in the Mediterranean Sea', *Deep Sea Research Part II: Topical Studies in Oceanography*. Pergamon, 40(6), pp. 1327–1332. doi: 10.1016/0967-0645(93)90072-U.
- [120] Lavigne, H., D'Ortenzio, F., Migon, C., Claustre, H., Testor, P., d'Alcalà, M. R., Lavezza, R., Houpert, L. and Prieur, L. (2013) 'Enhancing the comprehension of mixed layer depth control on the Mediterranean phytoplankton phenology', *Journal of Geophysical Research: Oceans*. John Wiley & Sons, Ltd, 118(7), pp. 3416–3430. doi:

- 10.1002/jgrc.20251.
- [121] Lejeusne, C., Chevaldonné, P., Pergent-Martini, C., Boudouresque, C. F. and Pérez, T. (2010) 'Climate change effects on a miniature ocean: the highly diverse, highly impacted Mediterranean Sea', *Trends in Ecology and Evolution*. Elsevier Current Trends, pp. 250–260. doi: 10.1016/j.tree.2009.10.009.
- [122] Letelier, R. M., Bidigare, R. R., Hebel, D. V., Ondrusek, M., Winn, C. D. and Karl, D. M. (1993) 'Temporal variability of phytoplankton community structure based on pigment analysis', *Limnology and Oceanography*. Wiley-Blackwell, 38(7), pp. 1420–1437. doi: 10.4319/lo.1993.38.7.1420.
- [123] Li, W. K. W., Dickie, P. M., Harrison, W. G. and Irwin, B. D. (1993) 'Biomass and production of bacteria and phytoplankton during the spring bloom in the western North Atlantic Ocean', *Deep Sea Research Part II: Topical Studies in Oceanography*. Pergamon, 40(1–2), pp. 307–327. doi: 10.1016/0967-0645(93)90019-J.
- [124] Litchman, E., Klausmeier, C. A., Schofield, O. M. and Falkowski, P. G. (2007) 'The role of functional traits and trade-offs in structuring phytoplankton communities: scaling from cellular to ecosystem level', *Ecology Letters*. John Wiley & Sons, Ltd (10.1111), 10(12), pp. 1170–1181. doi: 10.1111/j.1461-0248.2007.01117.x.
- [125] Liu, Y. and Weisberg, R. H. (2005) 'Patterns of ocean current variability on the West Florida Shelf using the self-organizing map', *Journal of Geophysical Research*. Wiley-Blackwell, 110(C6), p. C06003. doi: 10.1029/2004JC002786.
- [126] Liu, Y., Weisberg, R. H. and Mooers, C. N. K. (2006) 'Performance evaluation of the self-organizing map for feature extraction', *Journal of Geophysical Research*. Wiley-Blackwell, 111(C5), p. C05018. doi: 10.1029/2005JC003117.
- [127] Llasat, M. C., Llasat-Botija, M., Petrucci, O., Pasqua, A. A., Rosselló, J., Vinet, F. and Boissier, L. (2013) 'Towards a database on societal impact of Mediterranean floods within the framework of the HYMEX project', *Natural Hazards and Earth System Sciences*, 13(5), pp. 1337–1350. doi: 10.5194/nhess-13-1337-2013.
- [128] Lochte, K., Ducklow, H. ., Fasham, M. J. . and Stienen, C. (1993) 'Plankton succession and carbon cycling at 47°N 20°W during the JGOFS North Atlantic Bloom Experiment', *Deep Sea Research Part II: Topical Studies in Oceanography*. Pergamon, 40(1–2), pp. 91–114. doi: 10.1016/0967-0645(93)90008-B.
- [129] Loisel, H., Nicolas, J.-M., Sciandra, A., Stramski, D. and Poteau, A. (2006) 'Spectral dependency of optical backscattering by marine particles from satellite remote sensing of the global ocean', *Journal of Geophysical Research*. John Wiley & Sons, Ltd, 111(C9), p. C09024. doi: 10.1029/2005JC003367.
- [130] Longhurst, A. R. (1998) *Ecological geography of the sea*. San Diego: Academic Press.
- [131] Longhurst, A., Sathyendranath, S., Platt, T. and Caverhill, C. (1995) 'An estimate of global primary production in the ocean from satellite radiometer data', *Journal of Plankton Research*. Oxford University Press, 17(6), pp. 1245–1271. doi: 10.1093/plankt/17.6.1245.
- [132] Luo, Y.-W., Doney, S. C., Anderson, L. A., Benavides, M., Berman-Frank, I., Bode, A., Bonnet, S., Boström, K. H., Böttjer, D., Capone, D. G., Carpenter, E. J., Chen, Y. L., Church, M. J., Dore, J. E., Falcón, L. I., Fernández, A., Foster, R. A., Furuya, K., Gómez, F., Gundersen, K., Hynes, A. M., Karl, D. M., Kitajima, S., Langlois, R. J., LaRoche, J., Letelier, R. M., Marañón, E., McGillicuddy, D. J., Moisander, P. H., Moore, C. M., Mouriño-Carballido, B., Mulholland, M. R., Needoba, J. A., Orcutt, K. M., Poulton, A. J.,

- Rahav, E., Raimbault, P., Rees, A. P., Riemann, L., Shiozaki, T., Subramaniam, A., Tyrrell, T., Turk-Kubo, K. A., Varela, M., Villareal, T. A., Webb, E. A., White, A. E., Wu, J. and Zehr, J. P. (2012) 'Database of diazotrophs in global ocean: abundance, biomass and nitrogen fixation rates', *Earth System Science Data*, 4(1), pp. 47–73. doi: 10.5194/essd-4-47-2012.
- [133] Mackey, M., Mackey, D., Higgins, H. and Wright, S. (1996) 'CHEMTAX - a program for estimating class abundances from chemical markers: application to HPLC measurements of phytoplankton', *Marine Ecology Progress Series*, 144, pp. 265–283. doi: 10.3354/meps144265.
- [134] Manca, B., Ibello, V., Pacciaroni, M., Scarazzato, P. and Giorgetti, A. (2006) 'Ventilation of deep waters in the Adriatic and Ionian Seas following changes in thermohaline circulation of the Eastern Mediterranean', *Climate Research*, 31(2–3), pp. 239–256. doi: 10.3354/cr031239.
- [135] Mann, K. H. (Kenneth H. and Lazier, J. R. N. (2006) *Dynamics of marine ecosystems : biological-physical interactions in the oceans*. Blackwell Pub. Available at: <https://www.wiley.com/en-us/Dynamics+of+Marine+Ecosystems%3A+Biological+Physical+Interactions+in+the+Oceans%2C+3rd+Edition-p-9781405111188> (Accessed: 3 May 2018).
- [136] Maritorena, S., Siegel, D. A. and Peterson, A. R. (2002) 'Optimization of a semianalytical ocean color model for global-scale applications', *Applied Optics*. Optical Society of America, 41(15), p. 2705. doi: 10.1364/AO.41.002705.
- [137] Marty, J.-C., Chiavérini, J., Pizay, M.-D. and Avril, B. (2002) 'Seasonal and interannual dynamics of nutrients and phytoplankton pigments in the western Mediterranean Sea at the DYFAMED time-series station (1991–1999)', *Deep Sea Research Part II: Topical Studies in Oceanography*. Pergamon, 49(11), pp. 1965–1985. doi: 10.1016/S0967-0645(02)00022-X.
- [138] Mayot, N., D'Ortenzio, F., Uitz, J., Gentili, B., Ras, J., Vellucci, V., Golbol, M., Antoine, D. and Claustre, H. (2017) 'Influence of the Phytoplankton Community Structure on the Spring and Annual Primary Production in the Northwestern Mediterranean Sea', *Journal of Geophysical Research: Oceans*. Wiley-Blackwell, 122(12), pp. 9918–9936. doi: 10.1002/2016JC012668.
- [139] Mayot, N., Ortenzio, F. D., Ribera, M., Lavigne, H., Claustre, H., Biologica, O., Dohrn, S. Z. A. and Comunale, V. (2016) 'Interannual variability of the Mediterranean trophic regimes from ocean color satellites', pp. 1901–1917. doi: 10.5194/bg-13-1901-2016.
- [140] McClain, E. P., Pichel, W. G. and Walton, C. C. (1985) 'Comparative performance of AVHRR-based multichannel sea surface temperatures', *Journal of Geophysical Research*. Wiley-Blackwell, 90(C6), p. 11587. doi: 10.1029/JC090iC06p11587.
- [141] Millot, C. (1999) *Circulation in the Western Mediterranean Sea*, *Journal of Marine Systems*. Available at: http://hermes.dt.insu.cnrs.fr/moose/DOC_BIBLIO/millot99.pdf (Accessed: 10 February 2019).
- [142] Millot, C. and Taupier-Letage, I. (2005) 'The Mediterranean Sea', *Hdb Env Chem*. Springer, 5, pp. 29–66. doi: 10.1007/b107143.
- [143] Mitchell, B. G., Brody, E. A., Holm-Hansen, O., McClain, C. and Bishop, J. (1991) 'Light limitation of phytoplankton biomass and macronutrient utilization in the

- Southern Ocean', *Limnology and Oceanography*. Wiley-Blackwell, 36(8), pp. 1662–1677. doi: 10.4319/lo.1991.36.8.1662.
- [144] Moon-van der Staay, S. Y., van der Staay, G. W. M., Guillou, L., Vaulot, D., Claustre, H. and Medlin, L. K. (2000) 'Abundance and diversity of prymnesiophytes in the picoplankton community from the equatorial Pacific Ocean inferred from 18S rDNA sequences', *Limnology and Oceanography*. Wiley-Blackwell, 45(1), pp. 98–109. doi: 10.4319/lo.2000.45.1.0098.
- [145] Morel, A. and Bricaud, A. (1981) 'Theoretical results concerning light absorption in a discrete medium, and application to specific absorption of phytoplankton', *Deep-Sea Research*, 28A(11), pp. 1375–1393. Available at: http://omtab.obs-vlfr.fr/fichiers_PDF/Morel_and_Bricaud_DSR_81.pdf (Accessed: 11 February 2019).
- [146] Morel, A. and Gentili, B. (1996) 'Diffuse reflectance of oceanic waters III Implication of bidirectionality for the remote-sensing problem', *Applied Optics*. Optical Society of America, 35(24), p. 4850. doi: 10.1364/AO.35.004850.
- [147] Morel, A. and Prieur, L. (1977) 'Analysis of variations in ocean color1', *Limnology and Oceanography*. John Wiley & Sons, Ltd, 22(4), pp. 709–722. doi: 10.4319/lo.1977.22.4.0709.
- [148] Mouw, C. B. and Yoder, J. A. (2010) 'Optical determination of phytoplankton size composition from global SeaWiFS imagery', *Journal of Geophysical Research: Oceans*, 115(12), pp. 1–20. doi: 10.1029/2010JC006337.
- [149] Ben Mustapha, Z., Alvain, S., Jamet, C., Loisel, H. and Dessailly, D. (2013) 'Automatic classification of water-leaving radiance anomalies from global SeaWiFS imagery: Application to the detection of phytoplankton groups in open ocean waters', *Remote Sensing of Environment*, (February 2016). doi: 10.1016/j.rse.2013.08.046.
- [150] Navarro, G., Almaraz, P., Caballero, I., Vázquez, Á. and Huertas, I. E. (2017) 'Reproduction of Spatio-Temporal Patterns of Major Mediterranean Phytoplankton Groups from Remote Sensing OC-CCI Data', *Frontiers in Marine Science*, 4(August), pp. 1–16. doi: 10.3389/fmars.2017.00246.
- [151] Navarro, G., Alvain, S., Vantrepotte, V. and Huertas, I. E. (2014) 'Identification of dominant phytoplankton functional types in the Mediterranean Sea based on a regionalized remote sensing approach', *Remote Sensing of Environment*. Elsevier Inc., 152(February 2016), pp. 557–575. doi: 10.1016/j.rse.2014.06.029.
- [152] Niang, A., Badran, F., Moulin, C., Crépon, M. and Thiria, S. (2006) 'Retrieval of aerosol type and optical thickness over the Mediterranean from SeaWiFS images using an automatic neural classification method', *Remote Sensing of Environment*. Elsevier, 100(1), pp. 82–94. doi: 10.1016/J.RSE.2005.10.005.
- [153] Nieblas, A.-E., Drushka, K., Reygondeau, G., Rossi, V., Demarcq, H., Dubroca, L. and Bonhommeau, S. (2014) 'Defining Mediterranean and Black Sea Biogeochemical Subprovinces and Synthetic Ocean Indicators Using Mesoscale Oceanographic Features', *PLoS ONE*. Edited by S. Mazucca. Public Library of Science, 9(10), p. e111251. doi: 10.1371/journal.pone.0111251.
- [154] O'Reilly, J. E., Maritorena, S., Mitchell, B. G., Siegel, D. a., Carder, K. L., Garver, S. a., Kahru, M. and McClain, C. (1998) 'Ocean color chlorophyll algorithms for SeaWiFS', *Journal of Geophysical Research*, 103(C11), p. 24937. doi: 10.1029/98JC02160.
- [155] O'Reilly, J. E., Maritorena, S., O'Brien, M. C., Siegel, D. a, Toole, D., Menzies, D.,

- Smith, R. C., Mueller, J. L., Mitchell, B. G., Kahru, M., Chavez, F. P., Strutton, P., Cota, G. F., Hooker, S. B., McClain, C. R., Carder, K. L., Müller-Karger, F. E., Harding, L., Magnuson, A., Phinney, D., Moore, G. F., Aiken, J., Arrigo, K. R., Letelier, R. and Culver, M. (2000) 'SeaWiFS Postlaunch Calibration and Validation Analyses, Part 3', *SeaWiFS Postlaunch Technical Report Series*, 11, p. 51pp.
- [156] Organelli, E., Bricaud, A., Antoine, D. and Uitz, J. (2013) 'Multivariate approach for the retrieval of phytoplankton size structure from measured light absorption spectra in the Mediterranean Sea (BOUSSOLE site)', *Applied Optics*. Optical Society of America, 52(11), p. 2257. doi: 10.1364/AO.52.002257.
- [157] Orr, J. C., Fabry, V. J., Aumont, O., Bopp, L., Doney, S. C., Feely, R. A., Gnanadesikan, A., Gruber, N., Ishida, A., Joos, F., Key, R. M., Lindsay, K., Maier-Reimer, E., Matear, R., Monfray, P., Mouchet, A., Najjar, R. G., Plattner, G.-K., Rodgers, K. B., Sabine, C. L., Sarmiento, J. L., Schlitzer, R., Slater, R. D., Totterdell, I. J., Weirig, M.-F., Yamanaka, Y. and Yool, A. (2005) 'Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms', *Nature*. Nature Publishing Group, 437(7059), pp. 681–686. doi: 10.1038/nature04095.
- [158] Palmiéri, J. (2004) *Modélisation biogéochimique de la mer méditerranée avec le modèle régional couplé NEMO-MED12/PISCES*. Versailles-Saint-Quentin.
- [159] Pan, X., Mannino, A., Russ, M. E., Hooker, S. B. and Harding, L. W. (2010) 'Remote sensing of phytoplankton pigment distribution in the United States northeast coast', *Remote Sensing of Environment*. Elsevier, 114(11), pp. 2403–2416. doi: 10.1016/J.RSE.2010.05.015.
- [160] Pan, X., Wong, G. T. F., Ho, T. Y., Shiah, F. K. and Liu, H. (2013) 'Remote sensing of picophytoplankton distribution in the northern South China Sea', *Remote Sensing of Environment*. Elsevier Inc., 128, pp. 162–175. doi: 10.1016/j.rse.2012.10.014.
- [161] Pandžić, K. and Likso, T. (2005) 'Eastern Adriatic typical wind field patterns and large-scale atmospheric conditions', *International Journal of Climatology*. John Wiley & Sons, Ltd, 25(1), pp. 81–98. doi: 10.1002/joc.1085.
- [162] Passow, U. and Carlson, C. A. (2012) 'The biological pump in a high CO₂ world', *Marine Ecology Progress Series*. Inter-Research Science Center, pp. 249–272. doi: 10.2307/24876215.
- [163] Pesant, S., Not, F., Picheral, M., Kandels-Lewis, S., Le Bescot, N., Gorsky, G., Iudicone, D., Karsenti, E., Speich, S., Troublé, R., Dimier, C., Searson, S. and Tara, & (2015) 'Open science resources for the discovery and analysis of Tara Oceans data Background & Summary'. doi: 10.1038/sdata.2015.23.
- [164] Premuzic, E. T., Benkovitz, C. M., Gaffney, J. S. and Walsh, J. J. (1982) 'The nature and distribution of organic matter in the surface sediments of world oceans and seas', *Organic Geochemistry*. Pergamon, 4(2), pp. 63–77. doi: 10.1016/0146-6380(82)90009-2.
- [165] Psarra, S., Zohary, T., Krom, M. D., Mantoura, R. F. C., Polychronaki, T., Stambler, N., Tanaka, T., Tselepides, A. and Frede Thingstad, T. (2005) 'Phytoplankton response to a Lagrangian phosphate addition in the Levantine Sea (Eastern Mediterranean)', *Deep-Sea Research Part II: Topical Studies in Oceanography*, 52(22–23), pp. 2944–2960. doi: 10.1016/j.dsr2.2005.08.015.
- [166] Le Quéré, C., Harrison, S. P., Colin Prentice, I., Buitenhuis, E. T., Aumont, O., Bopp, L., Claustre, H., Cotrim Da Cunha, L., Geider, R., Giraud, X., Klaas, C., Kohfeld, K. E., Legendre, L., Manizza, M., Platt, T., Rivkin, R. B., Sathyendranath, S., Uitz, J.,

- Watson, A. J. and Wolf-Gladrow, D. (2005) 'Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models', *Global Change Biology*. John Wiley & Sons, Ltd (10.1111), 0(0), p. 051013014052005-??? doi: 10.1111/j.1365-2486.2005.1004.x.
- [167] Reusch, D. B., Alley, R. B. and Hewitson, B. C. (2007) 'North Atlantic climate variability from a self-organizing map perspective', *Journal of Geophysical Research*. Wiley-Blackwell, 112(D2), p. D02104. doi: 10.1029/2006JD007460.
- [168] Reygondeau, G., Guieu, C., Benedetti, F., Irisson, J.-O., Ayata, S.-D., Gasparini, S. and Koubbi, P. (2017) 'Biogeochemical regions of the Mediterranean Sea: An objective multidimensional and multivariate environmental approach', *Progress in Oceanography*. Pergamon, 151, pp. 138–148. doi: 10.1016/J.POCEAN.2016.11.001.
- [169] Reygondeau, G., Irisson, J. O., Ayata, S., Gasparini, S., Benedetti, F., Albouy, C., Hattab, T., Guieu, C. and Koubbi., P. (2014) *Definition of the Mediterranean Eco-regions and Maps of Potential Pressures in These Eco-regions*.
- [170] Reynolds, C. (2006) *The ecology of phytoplankton*. Cambridge University Press. Available at: https://books.google.fr/books?hl=en&lr=&id=gDz5jGsPWZYC&oi=fnd&pg=PA1&dq=reynolds+2006+ecology+of+phytoplankton&ots=tTxV1anEiK&sig=DUpDSx3xCu_K00PpbIH0epEN9ec (Accessed: 9 February 2019).
- [171] Reynolds, C. S., Huszar, V., Kruk, C., Naselli-Flores, L. and Melo, S. (2002) 'Towards a functional classification of the freshwater phytoplankton', *Journal of Plankton Research*. Oxford University Press, 24(5), pp. 417–428. doi: 10.1093/plankt/24.5.417.
- [172] Richardson, A. ., Risien, C. and Shillington, F. . (2003) 'Using self-organizing maps to identify patterns in satellite imagery', *Progress in Oceanography*. Pergamon, 59(2–3), pp. 223–239. doi: 10.1016/J.POCEAN.2003.07.006.
- [173] Robinson, R. A., Leslie, G. W., Theocharis, A. and Lascaratos, A. (2001) 'Mediterranean Sea Circulation', *Ocean Currents*, pp. 1–19. doi: 10.1006/rwos.2001.0376.
- [174] Rossi, V., Ser-Giacomi, E., López, C. and Hernández-García, E. (2014) 'Hydrodynamic provinces and oceanic connectivity from a transport network help designing marine reserves', *Geophysical Research Letters*, 41(8), pp. 2883–2891. doi: 10.1002/2014GL059540.
- [175] Roy, S., Llewellyn, C., Egeland, E. and Johnsen, G. (2011) *Phytoplankton pigments: characterization, chemotaxonomy and applications in oceanography*. Available at: https://books.google.fr/books?hl=en&lr=&id=K_2AJHmdvxgC&oi=fnd&pg=PR9&dq=Roy+et+al.,+2011+phytoplankton&ots=YIXynpy6E_&sig=gm4mll1incd0vUH2ehHQ6nZywd4 (Accessed: 9 February 2019).
- [176] Roy, S., Sathyendranath, S., Bouman, H. and Platt, T. (2013) 'The global distribution of phytoplankton size spectrum and size classes from their light-absorption spectra derived from satellite data', *Remote Sensing of Environment*. Elsevier, 139, pp. 185–197. doi: 10.1016/J.RSE.2013.08.004.
- [177] Saha, K., Zhao, X., Zhang, H., Casey, K. S., Zhang, D., Baker-Yeboah, S., Kilpatrick, K. A., Evans, R. H., Ryan, T. and Relph, J. M. (2018) *AVHRR Pathfinder version 5.3 level 3 collated (L3C) global 4km sea surface temperature for 1981-Present*. NOAA National Centers for Environmental Information., Dataset. doi: doi:10.7289/V52J68XX.

- [178] Sammartino, M., Cicco, A. Di, Marullo, S. and Santoleri, R. (2015) 'phytoplankton in the Mediterranean Sea from satellite ocean colour data of SeaWiFS', pp. 759–778. doi: 10.5194/os-11-759-2015.
- [179] Sammartino, M., Marullo, S., Santoleri, R., Scardi, M., Marine, S. and Nazionale, C. (2018) 'Modelling the Vertical Distribution of Phytoplankton Biomass in the Mediterranean Sea from Satellite Data : A Neural Network Approach', pp. 1–24. doi: 10.3390/rs10101666.
- [180] Sarmiento, J. L., Slater, R., Barber, R., Bopp, L., Doney, S. C., Hirst, A. C., Kleypas, J., Matear, R., Mikolajewicz, U., Monfray, P., Soldatov, V., Spall, S. A. and Stouffer, R. (2004) 'Response of ocean ecosystems to climate warming', *Global Biogeochemical Cycles*. Wiley-Blackwell, 18(3), p. n/a-n/a. doi: 10.1029/2003GB002134.
- [181] Sarno, D. and Zingone, A. (2008) 'MARECHIARA-phytoplankton long-term time-series (1984-2006) at the fixed coastal station in the Gulf of Naples, Southern Tyrrhenian Sea.', *Pangea*. doi: <https://doi.org/10.1594/PANGAEA.701471>.
- [182] Sarzeaud, O. and Stephan, Y. (2000) 'Data interpolation using Kohonen networks', *Proceedings of the International Joint Conference on Neural Networks*, 6(C), pp. 197–202.
- [183] Sathyendranath, S., Aiken, J., Alvain, S., Barlow, R., Bouman, H., Bracher, A., Brewin, R., Bricaud, A., Brown, C. W., Ciotti, A. M., Clementson, L. A., Craig, S. E., Devred, E., Hardman-Mountford, N., Hirata, T., Hu, C., Kostadinov, T. S., Lavender, S., Loisel, H., Moore, T. S., Morales, J., Mouw, C. B., Nair, A., Raitsos, D., Roesler, C., Shutler, J. D., Sosik, H. M., Soto, I., Stuart, V., Subramaniam, A. and Uitz, J. (2014) *Phytoplankton functional types from Space*. IOCCG; 15. Edited by S. Sathyendranath and V. Stuart. Dartmouth, Nova Scotia, B2Y 4A2, Canada.: International Ocean-Colour Coordinating Group. Available at: <https://epic.awi.de/id/eprint/36000/> (Accessed: 9 February 2019).
- [184] Sathyendranath, S., Watts, L., Devred, E., Platt, T., Caverhill, C. and Maass, H. (2004) 'Discrimination of diatoms from other phytoplankton using ocean-colour data', *Marine Ecology Progress Series*, 272, pp. 59–68. doi: 10.3354/meps272059.
- [185] Schulz, K., Neill, C., Riebesell, U., Schulz, K. G., Bellerby, R. G. J., Botros, M., Fritsche, P., Meyerhöfer, M., Neill, C., Nondal, G., Oschlies, A., Wohlers, J. and Zöllner, E. (2007) 'Enhanced biological carbon consumption in a high CO₂ ocean FP7 EURO-BASIN View project Ocean Abyssal Carbon Experiment View project Enhanced biological carbon consumption in a high CO₂ ocean'. doi: 10.1038/nature06267.
- [186] Shata, A., & El Fayoumy, I. (1970) 'Remarks on the hydrogeology of the Nile delta, UAR. Hydrology of Deltas, 385-396.', *Hydrology of Deltas*, pp. 385–396.
- [187] Sieburth, J. M., Smetacek, V. and Lenz, J. (1978) 'Pelagic ecosystem structure: Heterotrophic compartments of the plankton and their relationship to plankton size fractions 1', *Limnology and Oceanography*. John Wiley & Sons, Ltd, 23(6), pp. 1256–1263. doi: 10.4319/lo.1978.23.6.1256.
- [188] Siokou-Frangou, I., Christaki, U., Mazzocchi, M. G., Montresor, M., Ribera D'Alcala, M., Vaque, D. and Zingone, A. (2010) 'Plankton in the open mediterranean Sea: A review', *Biogeosciences*, 7(5), pp. 1543–1586. doi: 10.5194/bg-7-1543-2010.
- [189] Smetacek, V. (1999) 'Diatoms and the Ocean Carbon Cycle', *Protist*. Urban & Fischer, 150(1), pp. 25–32. doi: 10.1016/S1434-4610(99)70006-4.
- [190] Smith, R. B. and Smith, R. B. (1987) 'Aerial Observations of the Yugoslavian Bora', *Journal of the Atmospheric Sciences*, 44(2), pp. 269–297. doi: 10.1175/1520-

- 0469(1987)044<0269:AOOTYB>2.0.CO;2.
- [191] Smyth, T. J., Moore, G. F., Groom, S. B., Land, P. E. and Tyrrell, T. (2002) 'Optical modeling and measurements of a coccolithophore bloom', *Applied Optics*. Optical Society of America, 41(36), p. 7679. doi: 10.1364/AO.41.007679.
- [192] Smyth, T. J., Moore, G. F., Hirata, T. and Aiken, J. (2006) 'Semianalytical model for the derivation of ocean color inherent optical properties: description, implementation, and performance assessment', *Applied Optics*. Optical Society of America, 45(31), p. 8116. doi: 10.1364/AO.45.008116.
- [193] Sorgente, R., Olita, A., Oddo, P., Fazioli, L. and Ribotti, A. (2011) 'Numerical simulation and decomposition of kinetic energy in the Central Mediterranean: insight on mesoscale circulation and energy conversion', *Ocean Science*, 7(4), pp. 503–519. doi: 10.5194/os-7-503-2011.
- [194] Sosik, H. M., Olson, R. J. and Armbrust, E. V. (2010) 'Flow Cytometry in Phytoplankton Research', in *Chlorophyll a Fluorescence in Aquatic Sciences: Methods and Applications*. Dordrecht: Springer Netherlands, pp. 171–185. doi: 10.1007/978-90-481-9268-7_8.
- [195] Subramaniam, A. and Carpenter, E. J. (1994) 'An empirically derived protocol for the detection of blooms of the marine cyanobacterium *Trichodesmium* using CZCS imagery', *International Journal of Remote Sensing*. Taylor & Francis Group, 15(8), pp. 1559–1569. doi: 10.1080/01431169408954191.
- [196] Sullivan, J. M. and Twardowski, M. S. (2009) 'Angular shape of the oceanic particulate volume scattering function in the backward direction', *Applied Optics*. Optical Society of America, 48(35), p. 6811. doi: 10.1364/AO.48.006811.
- [197] Syvitski, J. P. M. and Saito, Y. (2007) 'Morphodynamics of deltas under the influence of humans', *Global and Planetary Change*. Elsevier, 57(3–4), pp. 261–282. doi: 10.1016/J.GLOPLACHA.2006.12.001.
- [198] Tarran, G. A. (2011) *Abundance of phytoplankton heterotrophic nanoflagellates and bacteria through the water column at time series station L4 in the Western English Channel in 2007*, *Pangaea*. doi: 10.1594/PANGAEA.756537.
- [199] Testor, P., Béranger, K. and Mortier, L. (2005) 'Modeling the deep eddy field in the southwestern Mediterranean: The life cycle of Sardinian eddies', *Geophysical Research Letters*. John Wiley & Sons, Ltd, 32(13), p. L13602. doi: 10.1029/2004GL022283.
- [200] Theocharis, A., Balopoulos, E., Kioroglou, S., Kontoyiannis, H. and Iona, A. (1999) 'A synthesis of the circulation and hydrography of the South Aegean Sea and the Straits of the Cretan Arc (March 1994–January 1995)', *Progress in Oceanography*. Pergamon, 44(4), pp. 469–509. doi: 10.1016/S0079-6611(99)00041-5.
- [201] Tomas, C. R. and Hasle, G. R. (Grethe R. (1997) *Identifying marine phytoplankton*. Academic Press. Available at: <https://books.google.fr/books?hl=en&lr=&id=8WLABHmo-K8C&oi=fnd&pg=PP1&dq=Tomas,+1997+phytoplankton+microscope&ots=MGKeSCDzJK&sig=Z2wgEIIcXalWKJyUdFW58gG0hMM#v=onepage&q=Tomas%2C+1997+phytoplankton+microscope&f=false> (Accessed: 9 February 2019).
- [202] Tréguer, P., Nelson, D. M., Van Bennekom, A. J., Demaster, D. J., Leynaert, A. and Quéguiner, B. (1995) 'The silica balance in the world ocean: A reestimate', *Science*. American Association for the Advancement of Science, 268(5209), pp. 375–379. doi: 10.1126/science.268.5209.375.

- [203] Tsiola, A., Pitta, P. and Fodelianakis, S. et al (2016) 'Nutrient Limitation in Surface Waters of the Oligotrophic', *Microbial Ecology*, 71(3), pp. 575–588.
- [204] Turley, C., Bianchi, M., Christaki, U., Conan, P., Harris, J., Psarra, S., Ruddy, G., Stutt, E., Tselepides, A. and Van Wambeke, F. (2000) 'Relationship between primary producers and bacteria in an oligotrophic sea-the Mediterranean and biogeochemical implications', *Marine Ecology Progress Series*, 193, pp. 11–18. doi: 10.3354/meps193011.
- [205] Uitz, J., Claustre, H., Morel, A. and Hooker, S. B. (2006) 'Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll', *J. Geophys. Res.*, 111(C8), p. C08005. doi: 10.1029/2005jc003207.
- [206] Uitz, J., Claustre, H., Morel, A. and Hooker, S. B. (2006) 'Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll', *Journal of Geophysical Research*. John Wiley & Sons, Ltd, 111(C8), p. C08005. doi: 10.1029/2005JC003207.
- [207] Uitz, J., Stramski, D., Gentili, B., D'Ortenzio, F. and Claustre, H. (2012) 'Estimates of phytoplankton class-specific and total primary production in the Mediterranean Sea from satellite ocean color observations', *Global Biogeochemical Cycles*, 26(2), pp. 1–10. doi: 10.1029/2011GB004055.
- [208] Vargas-Yáñez, M., Plaza, F., García-Lafuente, J., Sarhan, T., Vargas, J. . and Vélez-Belchi, P. (2002) 'About the seasonal variability of the Alboran Sea circulation', *Journal of Marine Systems*. Elsevier, 35(3–4), pp. 229–248. doi: 10.1016/S0924-7963(02)00128-8.
- [209] Vidussi, F., Claustre, H., Manca, B. B., Luchetta, A. and Marty, J.-C. (2001) 'Phytoplankton pigment distribution in relation to upper thermocline circulation in the eastern Mediterranean Sea during winter', *Journal of Geophysical Research: Oceans*. Wiley-Blackwell, 106(C9), pp. 19939–19956. doi: 10.1029/1999JC000308.
- [210] Vidussi, F., Marty, J. and Chiave, J. (2000) 'Phytoplankton pigment variations during the transition from spring bloom to oligotrophy in the northwestern Mediterranean sea', 47.
- [211] Volten, H., de Haan, J. F., Hovenier, J. W., Schreurs, R., Vassen, W., Dekker, A. G., Hoogenboom, H. J., Charlton, F. and Wouts, R. (1998) 'Laboratory measurements of angular distributions of light scattered by phytoplankton and silt', *Limnology and Oceanography*. John Wiley & Sons, Ltd, 43(6), pp. 1180–1197. doi: 10.4319/lo.1998.43.6.1180.
- [212] Walton, C. C., Pichel, W. G., Sapper, J. F. and May, D. A. (1998) 'The development and operational application of nonlinear algorithms for the measurement of sea surface temperatures with the NOAA polar-orbiting environmental satellites', *Journal of Geophysical Research: Oceans*. John Wiley & Sons, Ltd, 103(C12), pp. 27999–28012. doi: 10.1029/98JC02370.
- [213] Werdell, P. J. and Bailey, S. W. (2005) 'An improved in-situ bio-optical data set for ocean color algorithm development and satellite data product validation', *Remote Sensing of Environment*, 98(1), pp. 122–140. doi: 10.1016/j.rse.2005.07.001.
- [214] Westberry, T., Behrenfeld, M. J., Siegel, D. A. and Boss, E. (2008) 'Carbon-based primary productivity modeling with vertically resolved photoacclimation', *Global Biogeochemical Cycles*. Wiley-Blackwell, 22(2), p. n/a-n/a. doi: 10.1029/2007GB003078.
- [215] Westberry, T. K., Siegel, D. A. and Subramaniam, A. (2005) 'An improved bio-

- optical model for the remote sensing of *Trichodesmium* spp. blooms', *Journal of Geophysical Research*. John Wiley & Sons, Ltd, 110(C6), p. C06012. doi: 10.1029/2004JC002517.
- [216] Witkowski, K., Król, T., Zieliriski, A. and Kuteń, E. (1998) 'A light-scattering matrix for unicellular marine phytoplankton', *Limnology and Oceanography*. John Wiley & Sons, Ltd, 43(5), pp. 859–869. doi: 10.4319/lo.1998.43.5.0859.
- [217] Wright, S. W. and Jeffrey, S. W. (1987) 'Fucoxanthin pigment markers of marine phytoplankton analysed by HPLC and HPTLC', *Marine Ecology Progress Series*. Inter-Research Science Center, pp. 259–266. doi: 10.2307/24825629.
- [218] Wu, C., Wu, J., Qi, J., Zhang, L., Huang, H., Lou, L. and Chen, Y. (2010) 'Empirical estimation of total phosphorus concentration in the mainstream of the Qiantang River in China using Landsat TM data', *International Journal of Remote Sensing*, 31(933814462), pp. 2309–2324. doi: 10.1080/01431160902973873.
- [219] Yacobi, Y. Z., Zohary, T., Kress, N., Hecht, A., Robarts, R. D., Waiser, M., Wood, A. M. and Li, W. K. W. (1995) 'Chlorophyll distribution throughout the southeastern Mediterranean in relation to the physical structure of the water mass', *Journal of Marine Systems*, 6(3), pp. 179–190. doi: 10.1016/0924-7963(94)00028-A.
- [220] Yentsch, C. M. and Horan, P. K. (1989) 'Cytometry in the Aquatic Sciences', *Cytometry*. John Wiley & Sons, Ltd, 10(5), pp. 497–499. doi: 10.1002/cyto.990100503.
- [221] Yogev, T., Rahav, E., Bar-Zeev, E., Man-Aharonovich, D., Stambler, N., Kress, N., Béjà, O., Mulholland, M. R., Herut, B. and Berman-Frank, I. (2011) 'Is dinitrogen fixation significant in the Levantine Basin, East Mediterranean Sea?', *Environmental Microbiology*, 13(4), pp. 854–871. doi: 10.1111/j.1462-2920.2010.02402.x.