

Etude et paramétrisation de la distribution verticale de la biomasse phytoplanctonique dans l'ocean global

Raphaëlle Sauzède

▶ **To cite this version:**

Raphaëlle Sauzède. Etude et paramétrisation de la distribution verticale de la biomasse phytoplanctonique dans l'ocean global. Océan, Atmosphère. Université Pierre et Marie Curie - Paris VI, 2015. Français. ⟨NNT : 2015PA066625⟩. ⟨tel-01342441⟩

HAL Id: tel-01342441

https://theses.hal.science/tel-01342441v1

Submitted on 6 Jul 2016

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.



HAL Authorization

Université Pierre et Marie Curie

Ecole doctorale des sciences de l'environnement d'Ile de France

Observatoire Océanologique de Villefranche / Equipe Optique Marine et Télédetection, Application Biogéochimiques

Etude et paramétrisation de la distribution verticale de la biomasse phytoplanctonique dans l'océan global

Par Raphaëlle SAUZÈDE

Thèse de doctorat spécialité : Océan Atmosphère, Climat et Observations Spatiales

Dirigée par Hervé CLAUSTRE & Julia UITZ

Présentée et soutenue publiquement le 11 Décembre 2015

Devant un jury composé de :

Mme Shubha SATHYENDRANATH	Professeur	Rapporteur
M. Emmanuel BOSS	Professeur	Rapporteur
M. Hubert LOISEL	Professeur	Examinateur
M. Stéphane BLAIN	Professeur	Examinateur
M. Cédric JAMET	Maître de conférences	Examinateur
Mme Julia UITZ	Chargée de recherche	Examinateur
M. Hervé CLAUSTRE	Directeur de recherche	Examinateur

RÉSUMÉ

Les travaux présentés dans cette thèse concernent la paramétrisation de la distribution verticale de la biomasse et de la structure des communautés phytoplanctoniques dans l'océan global. Nous avons tout d'abord développé une méthode neuronale de calibration de la fluorescence en concentration en chlorophylle-*a* ($[Chl]$) associée à la biomasse phytoplanctonique totale et à trois classes de taille de phytoplancton (i.e. micro-, nano- et pico-phytoplancton). Cette méthode, FLAVOR (Fluorescence to Algal communities Vertical distribution in the Oceanic Realm), a été entrainée et validée à l'aide une base de données de ~900 profils concomitants de fluorescence et de pigments mesurés pat HPLC. Une base de données globale de ~49000 profils de fluorescence a ensuite été assemblée et calibrée en termes de biomasse chlorophyllienne et composition du phytoplancton. Ce travail représente une première étape vers une vision globale tridimensionnelle de la biomasse phytoplanctonique. Nous avons ensuite développé deux réseaux de neurones (SOCA : Satellite Ocean Color and Argo data to infer vertical distribution of bio-optical properties) pour estimer la distribution verticale de deux indicateurs bio-optiques de la biomasse phytoplanctonique, $[Chl]$ et le coefficient de rétrodiffusion particulaire. Ces réseaux de neurones requièrent comme données d'entrée des données satellites de couleur de l'eau co-localisées avec un profil hydrologique collecté par un flotteur Argo. Ils ont été entrainés et validés avec une base de données globale composée de ~5 000 profils de propriétés bio-optiques et hydrologiques acquises par des flotteurs Bio-Argo de façon concomitante aux données satellitaires. Les bases de données utilisées pour développer les méthodes FLAVOR et SOCA proviennent de régions océaniques très variées, représentatives de l'océan global, permettant ainsi l'application de ces deux méthodes à la majorité des eaux océaniques. Finalement, nous avons mené une étude focalisée sur l'Atlantique Nord qui exploite les outils développés. Les champs tridimensionnels de biomasse obtenus, couplés à un modèle bio-optique de production primaire, permettent d'étudier les cycles saisonniers de la distribution verticale de la biomasse phytoplanctonique et de la production primaire dans différentes bio-régions de l'Atlantique Nord.

ABSTRACT

This PhD work focuses on the parameterization of the vertical distribution of phytoplankton biomass and community structure in the global open ocean. First we have developed a neural network-based method for the calibration of the fluorescence in chlorophyll-*a* concentration $[Chl]$ associated with the total phytoplankton biomass and with three phytoplankton size classes (i.e. micro, nano- and pico-phytoplankton). This method, (FLAVOR for Fluorescence to Algal communities Vertical distribution in the Oceanic Realm), was trained and validated using a database of ~900 concomitant fluorescence and HPLC-determined pigment profiles. A global database comprising ~49 000 fluorescence profiles was assembled and calibrated with FLAVOR. The resulting database represents a first step towards a global three-dimensional view of phytoplankton biomass and community composition. Second, two neural networks (SOCA for Satellite Ocean Color and Argo data to infer vertical distribution of bio-optical properties) were developed to infer the vertical distribution of two bio-optical proxies of the phytoplankton biomass, $[Chl]$ and the particulate backscattering coefficient, using as input satellite-derived products matched up with a hydrological Argo profile. The SOCA methods were trained and validated using a global database of ~5 000 profiles of bio-optical and hydrological properties collected from Bio-Argo floats with concomitant satellite products. The database used to develop FLAVOR and SOCA originates from various oceanic regions largely representative of the global ocean, making the methods applicable to most oceanic waters. Finally, we proposed a study dedicated to the North Atlantic where the tools developed in this thesis are used in conjunction with a bio-optical primary production model. This allows us to characterize the seasonal cycle of the vertical distribution of the phytoplankton biomass and primary production in various bio-regions of the North Atlantic.

Remerciements

Cette thèse a été réalisée au Laboratoire d'Océanographie de Villefranche, de l'Université Pierre et Marie Curie au sein de l'équipe "Optique Marine et Télédetection et Applications Biogéochimiques". Je remercie Antoine Sciandra, directeur du laboratoire, et Gaby Gorsky, directeur de l'Observatoire Océanologique de Villefranche, de m'avoir accueillie au sein de cette structure unique.

Je tiens à remercier tout particulièrement Hervé Claustre, mon directeur de thèse, pour avoir initié et dirigé ce travail. Sans lui, cette expérience unique n'aurait jamais eu lieu. Je luis suis extrêmement reconnaissante de m'avoir fait confiance en me permettant de faire cette thèse. Je le remercie d'avoir su me guider tout au long de ces trois ans en me laissant une grande autonomie. Merci pour sa disponibilité, son soutien et tous ses conseils qui m'ont permis de m'améliorer, de progresser et de développer ma critique scientifique.

Un grand merci à Julia Uitz, ma co-directrice de thèse, qui a su me conseiller et m'épauler tout au long de ma thèse. Merci pour sa grande disponibilité, son soutien dans les moments difficiles et pour sa motivation communicative qui m'a vraiment aidée à avancer.

Merci à Cédric Jamet pour notre collaboration tout au long de mon travail et pour ses conseils avisés sur les réseaux de neurones sans lesquels tout ce travail n'aurait jamais pu aboutir. Merci en plus de nos nombreux appels, de m'avoir accueillie à Wimereux et d'être venu à Villefranche pour échanger sur mon travail et m'aider à progresser.

Je remercie Hubert Loisel et Stéphane Blain pour avoir fait partie de mon comité de thèse. Ils m'ont amené un œil extérieur sur mon travail et m'ont permis chaque année d'avancer grâce à leurs conseils judicieux.

Je remercie Shubha Sathyendranath et Emmanuel Boss d'avoir accepté d'être rapporteurs de ce manuscrit. J'espère que la lecture de ce travail saura les intéresser et éveiller leur curiosité.

Merci également à Hubert Loisel, Stéphane Blain et Cédric Jamet d'avoir accepté d'évaluer mon

travail en faisant partie de mon jury de thèse.

Je tiens aussi à remercier le projet RemOcean et le Centre National d'Etudes Spatiales pour m'avoir accordé leurs financements.

La campagne GEOVIDE a été pour moi une expérience extraordinaire. Merci à Hervé de m'avoir permis de participer à cette mission et merci à toute l'équipe à bord qui a fait de cette expérience un moment inoubliable.

De manière générale, un grand merci à toute l'équipe OMTAB pour leur accueil chaleureux, pour leurs conseils en tout genre et pour toute l'aide que les personnes de l'équipe ont pu m'apporter au cours de ces trois ans.

Plus particulièrement, un immense merci au temps que Bernard Gentili m'a accordé pour me sortir des situations les plus critiques lorsque mon ordinateur avait décidé de ne plus me faire de cadeaux. Merci à lui également pour tous ses conseils en R qui vous changent la vie en terme de programmation !

Un grand merci également à Catherine Schmechtig et Antoine Poteau pour leur travail sur les données Bio-Argo sans lequel cette thèse n'aurait jamais pu avoir lieu.

Merci à Anick Bricaud et Fabrizio D'Ortenzio pour les discussions scientifiques et leurs nombreux conseils.

Je remercie Joséphine Ras de m'avoir appris les techniques de filtration HPLC et d'avoir consciencieusement relu mes papiers.

Merci à Nathan Briggs d'avoir relu mes documents en Anglais.

Merci à Linda Féré, Corinne Poutier et Isabelle Courtois d'avoir été les gestionnaires les plus efficaces que l'on puisse rêver d'avoir !

Je remercie également Carolyn Scheurle et Martina Ferraris pour m'avoir fait découvrir une autre facette de la science.

Je remercie les collègues de la mezzanine pour leur accueil chaleureux tous les matins et pour les petites discussions sur l'Atlantique Nord, le bloom, le b_{bp} ainsi que les problèmes Linux en tout genre : Marie, Nicolas, Léo et Christophe.

Merci aux personnes qui égayent le coin café : Catherine, Julia, Francis, Tristan, Joséphine, Edouard, Fabrizio, Emanuele, Emilie, Nathan, Céline, Thomas J., Thomas L., Mathieu, David ...

et tous les autres.

Je remercie les collègues, de tous les jours, devenus amis, qui m'ont aidé à décompresser en cas de besoin : Amanda, Orens, Laulau, Enzo, Sabine, Anouck, Mélek, Stéphanie, Lorenzo, Sophie, Justine, Marie, Simon, Nico, Maïa, Léo, Robin ...

Je n'oublie pas ceux qui sont passés par là puis partis vers de nouveaux horizons mais qui n'en ont pas moins été importants dans ma vie niçoise : Héloïse, Grigor, Florent, Malika ...

Merci aux amis de plus longue date pour votre soutien et tous les moments partagés qui se font d'ailleurs trop rares : Elsa, Noëlle, Florent, Delphine, Sam, Audrey, Grégoire, Alice, Joul's, Léa, Romain, Sacha, Hugo, Thomas...

Merci à Clément de m'avoir épaulée et d'avoir toujours cru en moi. Je te remercie pour les escapades au grand air dans l'arrière-pays niçois qui vident l'esprit. Et surtout merci pour tous les moments qu'on vit tous les deux qui égayent le quotidien. Et bien sûr je n'oublie pas de remercier le petit rayon de soleil Lisa.

Enfin, merci à toute ma famille d'avoir été là pour moi. D'abord je remercie mes grand-parents qui ont toujours été présents. Merci à mes petites sœurs, Alix et Emma, pour leur soutien et les rigolades qui détendent. Enfin, merci mille fois à mon papa et ma maman de m'avoir supportée jusqu'à la fin et de m'avoir offert tout ce qu'il fallait pour en arriver jusque là.

Table des matières

Liste des figures **I**

Liste des tableaux **V**

Liste des principaux acronymes **VII**

Liste des principales abréviations **IX**

1 Contexte scientifique **1**

1.1 Le phytoplancton dans l'océan mondial 2

1.1.1 Distribution de la biomasse phytoplanctonique 2

1.1.2 La structure des communautés phytoplanctoniques 5

1.2 Étude de la biomasse phytoplanctonique à l'échelle globale 7

1.2.1 Mesures satellitaires « couleur de l'eau » 7

1.2.2 Mesures *in situ* discrètes 10

1.2.3 Mesures *in situ* à haute fréquence 13

1.3 Objectifs de l'étude et description des axes de travail 20

1.4 Méthodologie : Approche neuronale 24

1.4.1 Un peu d'histoire 25

1.4.2 Modèle général d'un MLP 25

1.4.3 Les différentes phases du développement d'un MLP 28

2 Vers le développement d'une climatologie 3D des communautés phytoplanctoniques pour l'océan mondial **33**

2.1 Introduction 35

2.2 Calibration de la fluorescence en concentration de chlorophylle *a* totale et détermination de la composition des communautés phytoplanctoniques associées 40

2.2.1 Résumé de l'étude . . . 40
2.2.2 Article publié dans *Journal of Geophysical Research* : Retrieving the vertical distribution of chlorophyll *a* concentration and phytoplankton community composition from *in situ* fluorescence profiles : A method based on a neural network with potential for global-scale applications . . . 41
2.2.3 Conclusion et perspectives de l'étude . . . 62
2.3 Création d'une base de données de la distribution verticale de la composition des communautés phytoplanctoniques pour l'océan global . . . 62
2.3.1 Résumé de l'étude . . . 62
2.3.2 Article publié dans *Earth System Science Data* : Vertical distribution of chlorophyll *a* concentration and phytoplankton community composition from *in situ* fluorescence profiles : a first database for the global ocean . . . 64
2.3.3 Résultats complémentaires . . . 78
2.3.4 Exemple d'utilisation de la base de données : une étude spécifique à la Méditerranée . 80
2.3.5 Conclusion et perspectives de l'étude . . . 82
2.4 Conclusion générale . . . 83

3 Coupler des données satellites "couleur de l'eau" et des données *in situ* hydrologiques Argo dans le but de paramétriser la distribution verticale de plusieurs paramètres bio-optiques dans l'océan global 85

3.1 Introduction . . . 86
3.2 Paramétrisation de la distribution verticale du coefficient de rétrodiffusion particulaire . . . 89
3.2.1 Résumé de l'étude . . . 89
3.2.2 Article soumis à *Journal of Geophysical Research* : A neural network-based method for merging ocean color and Argo data to extend surface bio-optical properties to depth : particulate backscattering coefficient retrieval . . . 90
3.2.3 Conclusions et perspectives de l'étude . . . 108
3.3 Paramétrisation de la distribution verticale de la concentration en chlorophylle *a* . . . 109
3.3.1 Introduction . . . 109
3.3.2 Databases and data processing . . . 111
3.3.3 SOCA-CHL algorithm development . . . 113
3.3.4 Results and Discussion . . . 114

3.3.5 Conclusions and perspectives . . . 117
3.4 Conclusion générale . . . 118

4 Étude régionale en Atlantique Nord : variabilité saisonnière de la biomasse phytoplanctonique et de la production primaire 121

4.1 Introduction . . . 123
4.2 Données et méthodes . . . 126
4.2.1 Données utilisées . . . 126
4.2.2 Bio-régionalisation de l'Atlantique Nord . . . 126
4.2.3 Distribution verticale de plusieurs paramètres biogéochimiques à partir des méthodes SOCA . . . 127
4.2.4 Modèle de production primaire . . . 128
4.2.5 Réalisation de séries temporelles climatologiques pour plusieurs paramètres biogéochimiques . . . 130
4.3 Résultats et discussion . . . 132
4.3.1 Bio-régionalisation de l'Atlantique Nord en 5 zones écologiques . . . 132
4.3.2 Validation des modèles utilisés . . . 133
4.3.3 Variabilité saisonnière du coefficient de rétrodiffusion . . . 136
4.3.4 Variabilité saisonnière de la concentration en chlorophylle *a* . . . 138
4.3.5 Comparaison des cycles saisonniers du coefficient de rétrodiffusion et de la concentration en chlorophylle *a* . . . 141
4.3.6 Variabilité saisonnière de la composition des communautés phytoplanctoniques . . . 142
4.3.7 Variabilité saisonnière de la production primaire jusqu'au taux de croissance phytoplanctonique . . . 143
4.4 Conclusion générale et perspectives . . . 147

5 Conclusions générales 149

5.1 Rappels des objectifs principaux . . . 150
5.2 Résultats majeurs . . . 150
5.3 Intérêts et limites des approches présentées . . . 153
5.4 Travaux futurs et perspectives . . . 154

Bibliographie 159

A On the vertical distribution of the chlorophyll-a concentration in the Mediterranean Sea : a basin scale and seasonal approach 187

Liste des figures

1.1 Concentration moyenne de la concentration en chlorophylle *a* de surface en fonction des saisons 3
1.2 Schéma conceptuel de l'évolution saisonnière de la distribution verticale de la biomasse chlorophyllienne pour trois grandes zones océaniques 4
1.3 Adapté à partir de Vidussi *et al.* (2001). Classes de taille et pigments diagnostiques de 8 taxons phytoplanctoniques communs 6
1.4 Principales missions globales « couleur de l'océan » de 1975 à de nos jours 8
1.5 Extraite de Peloquin *et al.* (2013). Distribution géographique en surface des données pigments analysés par HPLC de la base de données MAREDAT 12
1.6 Extraite du CSIRO. Schéma conceptuel des cycles de mesures d'un flotteur profileur 17
1.7 Position des flotteurs Argo et Bio-Argo en Juin 2015 pour l'océan global 19
1.8 Nombre de profils de fluorescence mesurés en fonction des années pour différents modes d'acquisition 20
1.9 Schéma conceptuel des différents axes de travail présentés dans cette thèse 21
1.10 Schéma général d'un perceptron multicouche (MLP) 26
1.11 Exemple de sur-apprentissage d'un MLP 29
1.12 Évolution des erreurs d'apprentissage et de validation au cours de l'apprentissage d'un MLP . 30

2.1 Distribution des données « bouteilles » sur lesquelles reposent les climatologies du World Ocean Atlas pour 4 des 7 profondeurs standards accessibles (Conkright *et al.*, 2002) 36
2.2 Comparaison des données climatologiques saisonnières de surface de notre base de données calibrées (BD-FLAVOR) et du WOA pour l'hiver et le printemps 78
2.3 Comparaison des données climatologiques saisonnières de BD-FLAVOR et du WOA pour l'été et l'automne 79
2.4 Comparaison des données climatologiques de BD-FLAVOR et du WOA en surface et à 30m de profondeur 79

2.5 Comparaison des données climatologiques de BD-FLAVOR et du WOA à 75m et à 100m de profondeur 80
2.6 Distribution des profils de fluorescence calibrés en concentration en chlorophylle *a* dans BD-FLAVOR 80
2.7 Comparaison des relations entre la profondeur du DCM et la concentration en chlorophylle *a* de surface établie avec BD-FLAVOR et issue de Mignot *et al.* (2011) 80
2.8 Extraite de Lavigne *et al.* (2015). Climatologie saisonnière des profils verticaux de concentration en chlorophylle *a* pour quatre zones clés de la Méditerranée 81
2.9 Extraite de Lavigne *et al.* (2015). Evolution saisonnière de la profondeur du DCM pour quatre zones clés de la Méditerranée 81

3.1 Distribution géographique des 4725 profils Bio-Argo utilisés pour entrainer et valider SOCA-CHL et des 262 profils de campagnes océanographiques utilisés pour une validation indépendante de SOCA-CHL 112
3.2 Caractéristiques de la base de données Bio-Argo et de la base de données issue des campagnes océanographiques : distribution temporelle des données et histogrammes de la concentration en chlorophylle *a* satellitaire et de la couche de mélange 113
3.3 Evaluation de la méthode SOCA-CHL en comparaison aux données *in situ* HPLC et comparaison des résultats avec le modèle de Uitz *et al.* (2006) 114
3.4 Diagramme de dispersion des données climatologiques de concentration en chlorophylle *a* de surface déterminées par SOCA-CHL et dérivées des données satellites MODIS Aqua pour le mois de juin 116
3.5 Climatologies de la concentration en chlorophylle *a* estimées à partir de SOCA-CHL pour la surface, 75 m et 150 m de profondeur pour le mois de juin à l'échelle globale 117

4.1 Extraite de l'Ifremer (Pascale Lherminier). Schéma de circulation des courants en Atlantique Nord 123
4.2 Extraite de Longhurst (2010). Distribution géographique des 7 provinces écologiques de l'Atlantique Nord définies par Longhurst (2010) 125
4.3 Bio-régionalisation de l'Atlantique Nord en 5 grandes zones écologiques 131
4.4 Cycles saisonniers moyens de la profondeur de la couche de mélange et de la profondeur de la couche euphotique en fonction des 5 clusters de l'Atlantique Nord 133

4.5 Boîtes à moustaches (boxplots) mensuelles des rapports transformés en log entre le coefficient de rétrodiffusion estimé par SOCA-BBP à la surface et celui estimé par satellite 133

4.6 Distribution géographique des données HPLC de validation pour l'Atlantique Nord 135

4.7 Boîtes à moustaches (boxplots) mensuelles des rapports transformés en log entre la concentration en chlorophylle *a* estimée par SOCA-BBP à la surface et celle estimée par satellite 136

4.8 Cycles saisonniers climatologiques moyens du coefficient de rétrodiffusion estimé par SOCA-BBP en surface et intégré sur la couche productive pour les 5 clusters de l'Atlantique Nord . . 137

4.9 Cycles saisonniers climatologiques moyens de la concentration en chlorophylle *a* et du coefficient de rétrodiffusion de surface estimés par MODIS Aqua pour les 5 clusters de l'Atlantique Nord 137

4.10 Cycles saisonniers climatologiques moyens de la distribution verticale du coefficient de rétrodiffusion estimé par SOCA-BBP pour les 5 clusters de l'Atlantique Nord 138

4.11 Cycles saisonniers climatologiques moyens de la concentration en chlorophylle *a* estimée par SOCA-CHL-NA en surface et intégrée sur la couche productive pour les 5 clusters de l'Atlantique Nord 139

4.12 Cycles saisonniers climatologiques moyens de la distribution verticale de la concentration en chlorophylle *a* estimée par SOCA-CHL-NA pour les 5 clusters de l'Atlantique Nord 140

4.13 Cycles saisonniers climatologiques moyens superposés de la concentration en chlorophylle *a* et du coefficient de rétrodiffusion intégrés dans la couche productive pour chaque cluster de l'Atlantique Nord 141

4.14 Cycles saisonniers climatologiques moyens du PAR moyenné dans la couche de mélange pour les 5 clusters de l'Atlantique Nord 142

4.15 Cycles saisonniers climatologiques moyens des pourcentages de contribution de chaque classe de taille phytoplanctonique à la concentration en chlorophylle *a* intégrée dans la couche productive pour chaque cluster de l'Atlantique Nord 143

4.16 Cycles saisonniers climatologiques moyens de la production primaire intégrée sur la couche productive en fonction de la biomasse phytoplanctonique totale et des 3 classes de taille phytoplanctoniques pour les 5 clusters de l'Atlantique Nord 145

4.17 Cycles saisonniers climatologiques moyens du taux de croissance phytoplanctonique pour chaque cluster de l'Atlantique Nord 146

4.18 Cycles saisonniers climatologiques moyens du rapport carbone/chlorophylle phytoplanctonique (intégré sur la couche productive) pour chaque cluster de l'Atlantique Nord 147

Liste des tableaux

1.1 Pigments diagnostiques et leurs abréviations 12

4.1 Abréviations, localisation, période, nombre de stations et PI pour deux campagnes océanographiques en Atlantique Nord 134

4.2 Paramètres statistiques de validation du modèle SOCA-CHL-NA et comparaison avec le modèle établi par Uitz *et al.* (2006) 136

Liste des principaux acronymes

Acronymes	Définition
CDOM	Colored Dissolved Organic Matter
CTD	Conductivity Temperature Depth
CZCS	Coastal Zone Color Scanner
DCM	Deep Chlorophyll Maximum
ESA	European Spatial Agency
FLAVOR	Fluorescence to Algal communities
	Vertical distribution in the Oceanic Realm
IOP	Inherent Optical Properties
LEO	Low Earth Orbit
MERIS	Medium Resolution Imaging Spectrometer
MAPD	Median Absolute Percent Difference
MLP	Multi-Layered Perceptron
MODIS	Moderate Resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration
NN	Neural Network
NPQ	Non-Photochemical Quenching
OCR	Ocean Color Radiometry
PAR	Photosynthetically Available Radiation
POC	Particulate Organic Carbon
PP	Primary Production
RMSE	Root Mean Squared Error
SeaWiFS	Sea-viewing Wide Field-of-vien Sensor

Acronymes	Définition
SOCA	Satellite Ocean Color merged with Argo data to infer the vertical distribution of bio-optical properties
SOCA-BBP	Satellite Ocean Color merged with Argo data to infer the vertical distribution of particulate backscattering
SOCA-CHL	Satellite Ocean Color merged with Argo data to infer the vertical distribution of chlorophyll *a* concentration
SOCA-CHL-NA	Satellite Ocean Color merged with Argo data to infer the vertical distribution of chlorophyll *a* concentration and phytoplankton communities in the North Atlantic
VIIRS	Visible Infrared Imaging Radiometer Suite
VSF	Volume Scattering Function

Liste des principales abréviations

Abréviations	Définitions	Unités
b_{bp}	coefficient de rétrodiffusion particulaire	m^{-1}
$[Chl]$ ou Chl	concentration en chlorophylle *a*	$mg.m^{-3}$
C_{phyto}	carbone phytoplanctonique	$mg.C.m^{-3}$
f_{micro}	fraction de concentration en chlorophylle *a* associée au microphytoplancton	–
f_{nano}	fraction de concentration en chlorophylle *a* associée au nanophytoplancton	–
f_{pico}	fraction de concentration en chlorophylle *a* associée au picophytoplancton	–
K_{d490}	coefficient d'atténuation diffuse à 490 nm	m^{-1}
$[microChl]$	concentration en chlorophylle *a* associée au microphytoplancton	$mg.m^{-3}$
$[nanoChl]$	concentration en chlorophylle *a* associée au nanophytoplancton	$mg.m^{-3}$
$[picoChl]$	concentration en chlorophylle *a* associée au picophytoplancton	$mg.m^{-3}$
PP_{tot}	production primaire intégrée dans la zone productive et associée à la biomasse phytoplanctonique totale	$g.C.m^{-2}.d^{-1}$
PP_{micro}	production primaire intégrée dans la zone productive et associée au microphytoplancton	$g.C.m^{-2}.d^{-1}$
PP_{nano}	production primaire intégrée dans la zone productive et associée au nanophytoplancton	$g.C.m^{-2}.d^{-1}$

Abréviations	Définitions	Unités
PP_{pico}	production primaire intégrée dans la zone productive et associée au picophytoplancton	$g.C.m^{-2}.d^{-1}$
$[TChl]$	concentration en chlorophylle *a* associée à la biomasse phytoplanctonique totale	$mg.m^{-3}$
z	profondeur géométrique	m
Z_e	profondeur de la couche euphotique	m
Z_m	profondeur de la couche de mélange	m
Z_{pd}	profondeur de pénétration définie par $Z_{pd} = Z_e/4.6$ (Morel et Berthon, 1989)	m
Z_0 ou Z_{norm}	profondeur à laquelle la fluorescence commence à être constante en profondeur	m
ζ	profondeur normalisée : $\zeta = z/Z_0$	–
μ	taux de croissance phytoplanctonique	d^{-1}

CHAPITRE 1

Contexte scientifique

Sommaire

1.1 Le phytoplancton dans l'océan mondial 2

1.1.1 Distribution de la biomasse phytoplanctonique 2

1.1.2 La structure des communautés phytoplanctoniques 5

1.2 Étude de la biomasse phytoplanctonique à l'échelle globale 7

1.2.1 Mesures satellitaires « couleur de l'eau » 7

1.2.2 Mesures *in situ* discrètes 10

1.2.3 Mesures *in situ* à haute fréquence 13

1.3 Objectifs de l'étude et description des axes de travail 20

1.4 Méthodologie : Approche neuronale 24

1.4.1 Un peu d'histoire 25

1.4.2 Modèle général d'un MLP 25

1.4.3 Les différentes phases du développement d'un MLP 28

1.1 Le phytoplancton dans l'océan mondial

L'océan joue un rôle primordial dans la **régulation des flux de carbone** de la planète puisqu'environ la moitié de la production primaire annuelle globale est réalisée par les **organismes photosynthétiques océaniques** (Field, 1998). Par le processus de **photosynthèse**, le phytoplancton fixe le CO_2 atmosphérique et le transforme en matière organique. Ce carbone organique peut soit être fixé et recyclé dans les couches superficielles de l'océan, soit être exporté vers les niveaux trophiques supérieurs par le biais de la chaîne alimentaire ou bien vers les couches profondes de l'océan (par sédimentation de cellules, d'agrégats ou de pelotes fécales). C'est en particulier ce dernier **processus d'export**, appelé « pompe biologique à carbone », qui contribue à réguler la concentration en CO_2 atmosphérique de la planète (Takahashi *et al.*, 2009).

C'est l'équilibre entre les termes sources et puits de carbone dans la colonne d'eau qui détermine le flux net de carbone exporté vers les couches profondes et par conséquent, l'intensité locale de la pompe biologique (Volk et Hoffert, 1985). L'évolution future de l'efficacité des puits de CO_2 océaniques est encore très incertaine. Il est donc important de mieux comprendre les processus liés aux **flux de carbone**, de mieux **quantifier** ces flux, et de mieux évaluer leurs incertitudes dans l'océan à l'échelle globale (Le Quéré *et al.*, 2009). Pour cela, il est nécessaire d'étudier en amont la distribution spatiale et temporelle de la **biomasse phytoplanctonique** dans toute la **colonne d'eau**.

1.1.1 Distribution de la biomasse phytoplanctonique

La **distribution spatio-temporelle du phytoplancton** à la fois à la surface des océans mais aussi en profondeur dépend de nombreux facteurs qui peuvent être de nature **biotique** (e.g. sénescence des cellules, broutage par le zooplancton) ou **abiotique** (e.g. taux d'éclairement, température, concentration en nutriments, advection, diffusion). La distribution spatiale du phytoplancton est déjà bien documentée grâce aux observations satellitaires qui nous renseignent sur la concentration en chlorophylle *a* en surface avec une haute définition spatio-temporelle (voir Figure 1.1 et Sect. 1.2.1, McClain, 2009, Siegel *et al.*, 2013). La chlorophylle *a* étant le pigment majeur à la base de la photosynthèse, celui-ci est présent dans toutes les cellules phytoplanctoniques. La concentration en chlorophylle *a* ($[Chl]$) est le meilleur **estimateur** à ce jour de la quantité en phytoplancton présente dans l'océan.

De manière générale, les plus faibles teneurs en phytoplancton en surface sont observées dans les grands gyres subtropicaux. Les biomasses phytoplanctoniques les plus élevées se trouvent généralement dans les mers fermées et les systèmes d'upwelling. Enfin, des niveaux relativement élevés (en moyenne sur l'année) sont observés pour les zones de latitude $> 40°$- $50°$ Nord et Sud (Gregg, 2001, Gregg et Conkright, 2002), souvent en conséquence des floraisons printanières.

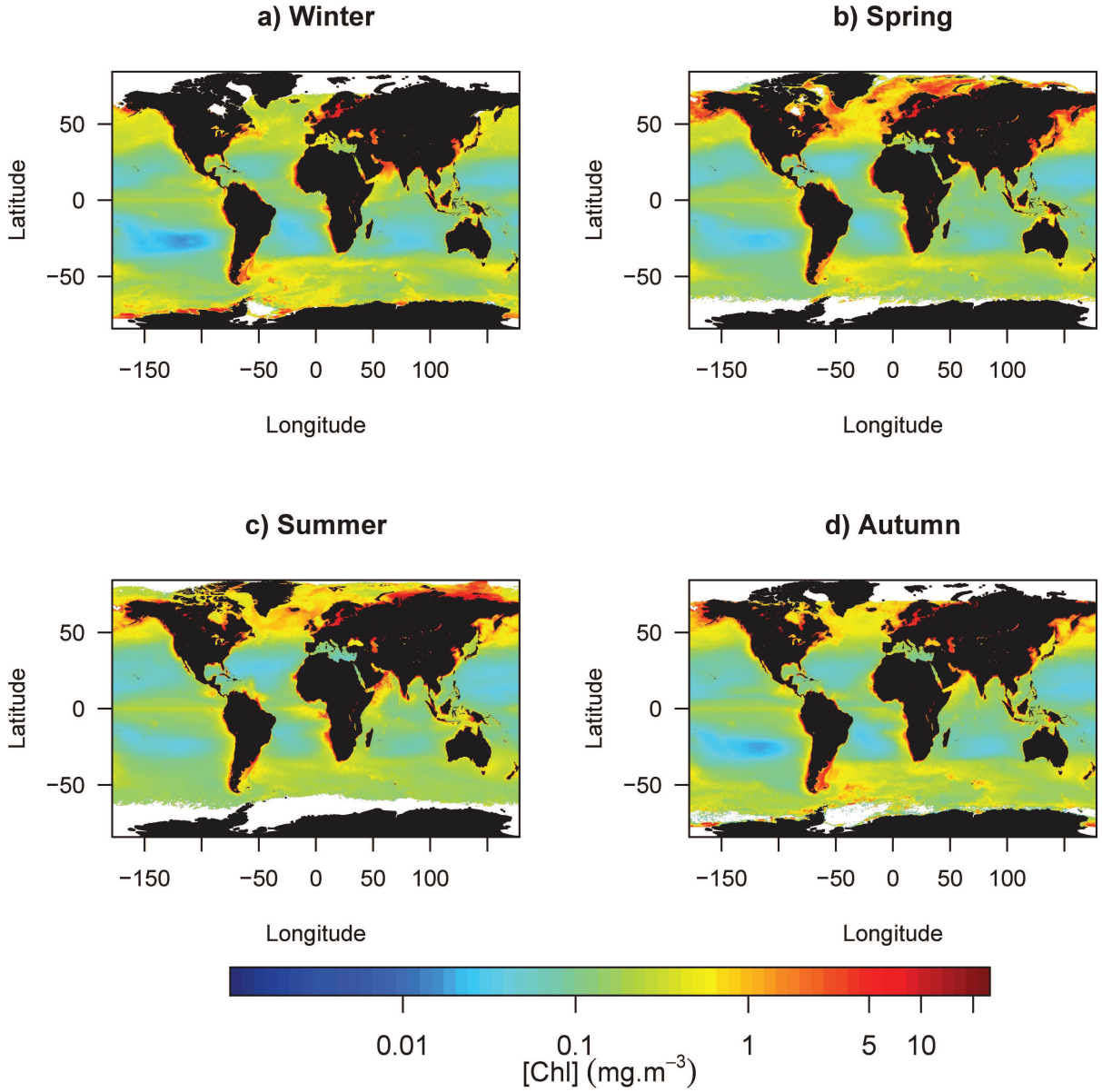


FIGURE 1.1 – *Concentration en chlorophylle a, (i.e. estimateur de la biomasse phytoplanctonique, $[Chl]$) en surface moyennée pour tout le composite MODIS Aqua (2002-2015) en fonction des saisons : a) hiver, b) printemps, c) été et d) automne.*

Il a été montré que les **propriétés physiques** océaniques de la colonne d'eau jouent un rôle fondamental dans la distribution du phytoplancton à la fois spatialement et temporellement (e.g. Wilson,

2005, Martinez *et al.*, 2009). En effet, l'intensité du mélange dans la colonne d'eau joue un rôle clé dans le contrôle notamment de la disponibilité en **lumière** et en **nutriments** (nitrates, phosphates, silicates, etc.), deux facteurs nécessaires à la **photosynthèse** et à la croissance des cellules phytoplanctoniques. La couche de l'océan où l'intensité lumineuse est favorable au processus de photosynthèse est appelée **zone euphotique**.

Il a été montré qu'au premier ordre la **variabilité saisonnière** de la biomasse phytoplanctonique est corrélée aux processus de mélange, stratification et déstratification de la colonne d'eau (influençant la disponibilité en nutriments) ainsi qu'à la disponibilité en lumière (Longhurst, 2010, Gregg, 2001, Cushing, 1959, Dutkiewicz *et al.*, 2001). Alors que dans les régions subtropicales le maximum de biomasse phytoplanctonique est observé en hiver, celui des régions subpolaires est observé en été (voir Figure 1.1). Les hautes latitudes ($> 40°$ Nord et Sud) sont caractérisées par une efflorescence printanière de phytoplancton, appelée plus communément « bloom » (e.g. Cullen *et al.*, 2002). Bien que ces dynamiques spatiales et temporelles soient déjà bien documentées en surface grâce à la haute résolution spatio-temporelle des mesures satellitaires, la distribution verticale associée est encore assez mal décrite, principalement à cause du **manque d'observations *in situ*** (physiques et biogéochimiques associées). Ainsi, des processus très étudiés comme l'initiation du bloom en Atlantique Nord, sont encore l'objet de controverses quant aux processus déterminants (e.g. Chiswell, 2011, Taylor et Ferrari, 2011, Behrenfeld, 2010, Mignot *et al.*, 2015).

Lorsque la colonne d'eau est **stratifiée**, la répartition verticale de la biomasse phytoplanctonique résulte de l'optimisation des ressources disponibles entre éléments nutritifs et lumière qui varient de manière opposée dans la colonne d'eau. Les **zones sub-tropicales** (de 10° à environ 40° Nord ou Sud) oligotrophes sont caractérisées par la quasi-omniprésence annuelle d'un maximum profond de biomasse chlorophyllienne (DCM pour « Deep Chlorophyll Maximum », Cullen, 1982, Fennel et Boss, 2003), principalement due au processus de photoacclimatation (Cullen, 1982, 2015, Mignot *et al.*, 2011). Bien que le DCM soit omniprésent au cours de l'année, une variation saisonnière est tout de même observée. Il a été montré que la profondeur du DCM était fortement corrélé à l'irradiance puisqu'il co-varie avec une certaine isolume (niveau de lumière constante, Letelier *et al.*, 2004, Mignot *et al.*, 2014). Ainsi, les isolumes et donc la profondeur du DCM s'approfondissent pendant l'été. Ceci permet au phytoplancton d'accéder à des couches profondes, plus riches en nutriments (voir Figure

1.2-a).

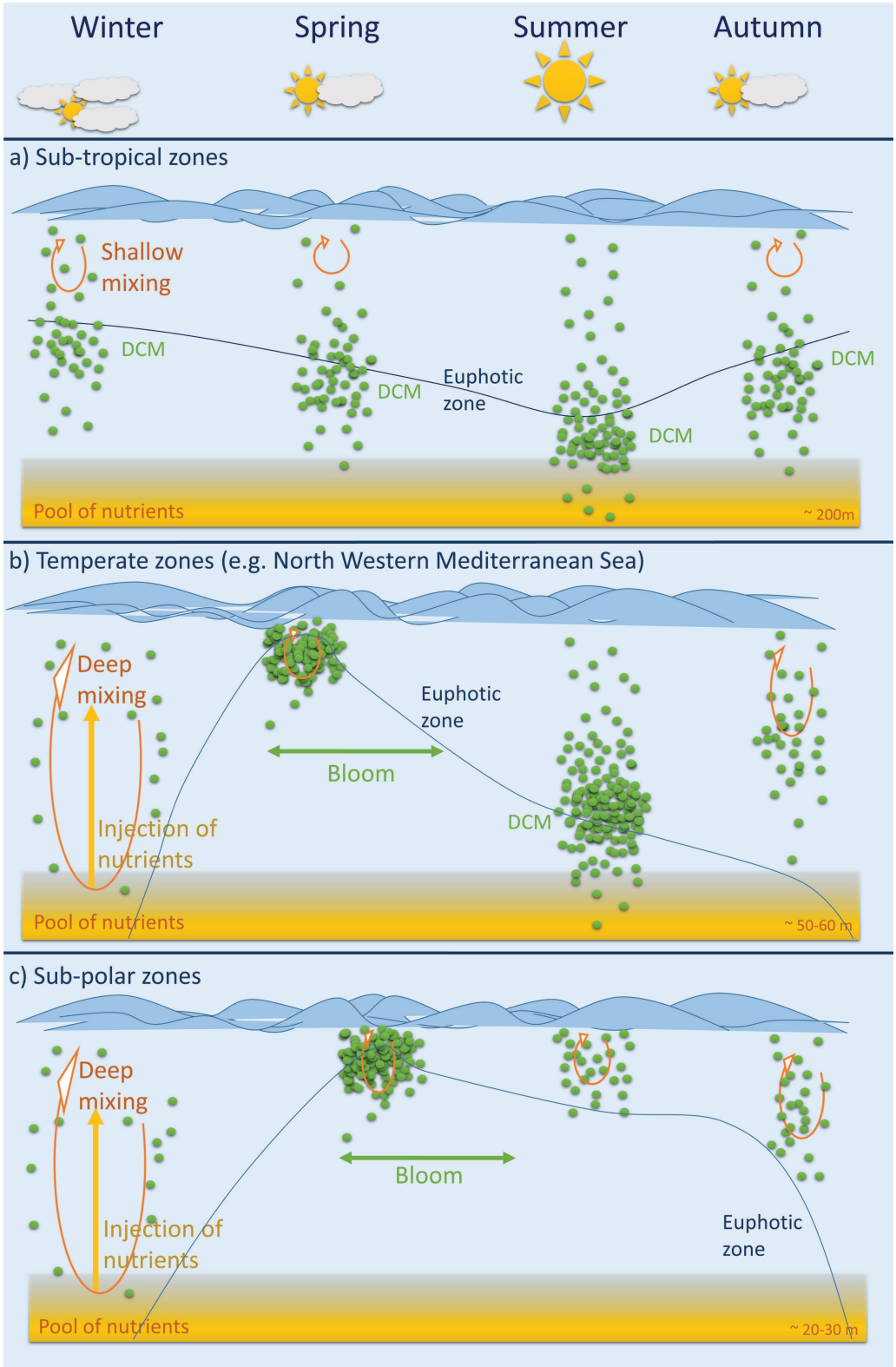


FIGURE 1.2 – *Schéma conceptuel de l'évolution saisonnière de la distribution verticale de la biomasse chlorophyllienne pour trois grandes zones océaniques : a) les zones sub-tropicales, b) les zones tempérées et c) les zones sub-polaires.*

Dans les **zones tempérées et sub-polaires** ($>$ 40° Nord et Sud), les **variations de mélange** dans la colonne d'eau ont un fort impact sur la distribution verticale et la croissance du phytoplancton à l'échelle annuelle. Le mélange peut déstratifier la colonne d'eau, apportant des eaux plus riches en nutriments dans les eaux de surface où la lumière est suffisante pour la photosynthèse, et ainsi permettre au phytoplancton de se développer (voir Figure 1.2-b et -c). Cependant, de forts mélanges dans les **hautes latitudes** ($>$ 60° Nord et Sud) peuvent réduire l'éclairement dans la couche de mélange à un niveau trop faible pour permettre la croissance phytoplanctonique (Sverdrup, 1953, Siegel, 2002a). Le réchauffement saisonnier conduit en général à la stratification de la colonne d'eau, ce qui permet de maintenir le phytoplancton dans une couche de surface suffisamment éclairée et riche en nutriments et de favoriser ainsi l'apparition d'un pic saisonnier de biomasse phytoplanctonique (i.e. **bloom**, Sverdrup, 1953). Pendant l'été, dans les **régions tempérées**, la stratification relativement profonde ne permet pas l'injection de nutriments dans les couches superficielles, ce qui entraîne le développement d'un DCM proche de la nutricline (voir Figure 1.2-b). La nutricline est définie par la zone de fort gradient des éléments nutritifs. Par contre dans les **régions sub-polaires**, la nutricline est plus proche de la surface ce qui permet de maintenir une production de biomasse phytoplanctonique significative dans cette couche (voir Figure 1.2-c).

Pour résumer, les **variations du forçage physique** (mélange, stratification et transitions entre les deux situations) sont à l'origine des modifications des champs de lumière et des éléments nutritifs dans les couches superficielles de l'océan. Ces variations impactent la biomasse phytoplanctonique ainsi que sa répartition verticale. Cependant, les relations sous-jacentes ne sont pas encore connues et décrites de manière précise dans toutes les régions océaniques du monde.

1.1.2 La structure des communautés phytoplanctoniques

La **composition de la biomasse phytoplanctonique** est déterminante dans la régulation des flux de carbone à l'échelle globale. Historiquement, diverses tentatives ont été présentées pour caractériser la structure des communautés phytoplanctoniques dans l'océan (Margalef, 1967, Smayda, 1980). Le nombre d'organismes, l'abondance relative des espèces, leurs traits biologiques, leurs tailles ainsi que leurs rôles fonctionnels ont été utilisés tour à tour et/ou conjointement pour décrire la structure des communautés phytoplanctoniques (Reynolds, 2006).

La taille des communautés phytoplanctoniques : un paramètre essentiel

Ce qui est communément appelé phytoplancton peut varier de $0.6\mu m$ pour les cellules les plus petites à des agrégats de plusieurs cellules de taille supérieure à $200\mu m$ (voir Figure 1.3). Ce continuum de tailles peut être divisé selon trois grandes **classes de taille** : le **picophytoplancton** $(0.2 - 2\mu m)$, le **nanophytoplancton** $(2 - 20\mu m)$ et le **microphytoplancton** $(20 - 200\mu m)$.

La **taille des cellules phytoplanctoniques** est considérée comme un facteur important à considérer pour les études écologiques et biogéochimiques (Sieburth *et al.*, 1978) car elle joue un rôle déterminant dans les voies vers lesquelles le **carbone** est dirigé (Michaels et Silver, 1988), l'efficacité avec laquelle il est fixé (Claustre *et al.*, 1997, Hashimoto et Shiomoto, 2002, Uitz *et al.*, 2008) puis à terme exporté vers les couches profondes de l'océan (Briggs *et al.*, 2011, Guidi *et al.*, 2009, Laws *et al.*, 2000). De plus, la taille du phytoplancton influence le taux de croissance du phytoplancton (Geider *et al.*, 1986, Sunda et Huntsman, 1997), ainsi que la structure de la chaine alimentaire océanique (Legendre et Le Fevre, 1991, Moloney et Field, 1991). C'est pourquoi de nombreux **modèles biogéochimiques** en océanographie ont été développés en prenant en compte la taille des communautés phytoplanctoniques (e.g. Aumont *et al.*, 2003, Blackford *et al.*, 2004, Kishi *et al.*, 2007, Marinov *et al.*, 2010, Ward *et al.*, 2012).

La classification des groupes phytoplanctoniques en différentes classes de taille permet d'obtenir une représentation **synthétique et informative** de la composition des communautés phytoplanctoniques. C'est pour cette raison que nous avons choisi d'utiliser cette classification dans les différentes analyses présentées tout au long de de ce manuscrit.

Le picophytoplancton

Les deux groupes pico-phytoplanctoniques les plus abondants sont *Prochlorococcus* $(< 1\mu m)$ et *Synechococcus* $(1 - 2\mu m)$. Ces deux groupes, faisant partie respectivement des cyanobactéries et des prochlorophytes (voir Figure 1.3), coexistent la plupart du temps dans les **eaux chaudes et oligotrophiques des régions tropicales et subtropicales** $(< 40°$ Nord et Sud Partensky *et al.*, 1999, Post, 2006). Ces deux groupes occupent des niches écologiques différentes dans la colonne d'eau : *Synechococcus* occupe la surface et les **couches les moins profondes** tandis que *Prochlorococcus* est

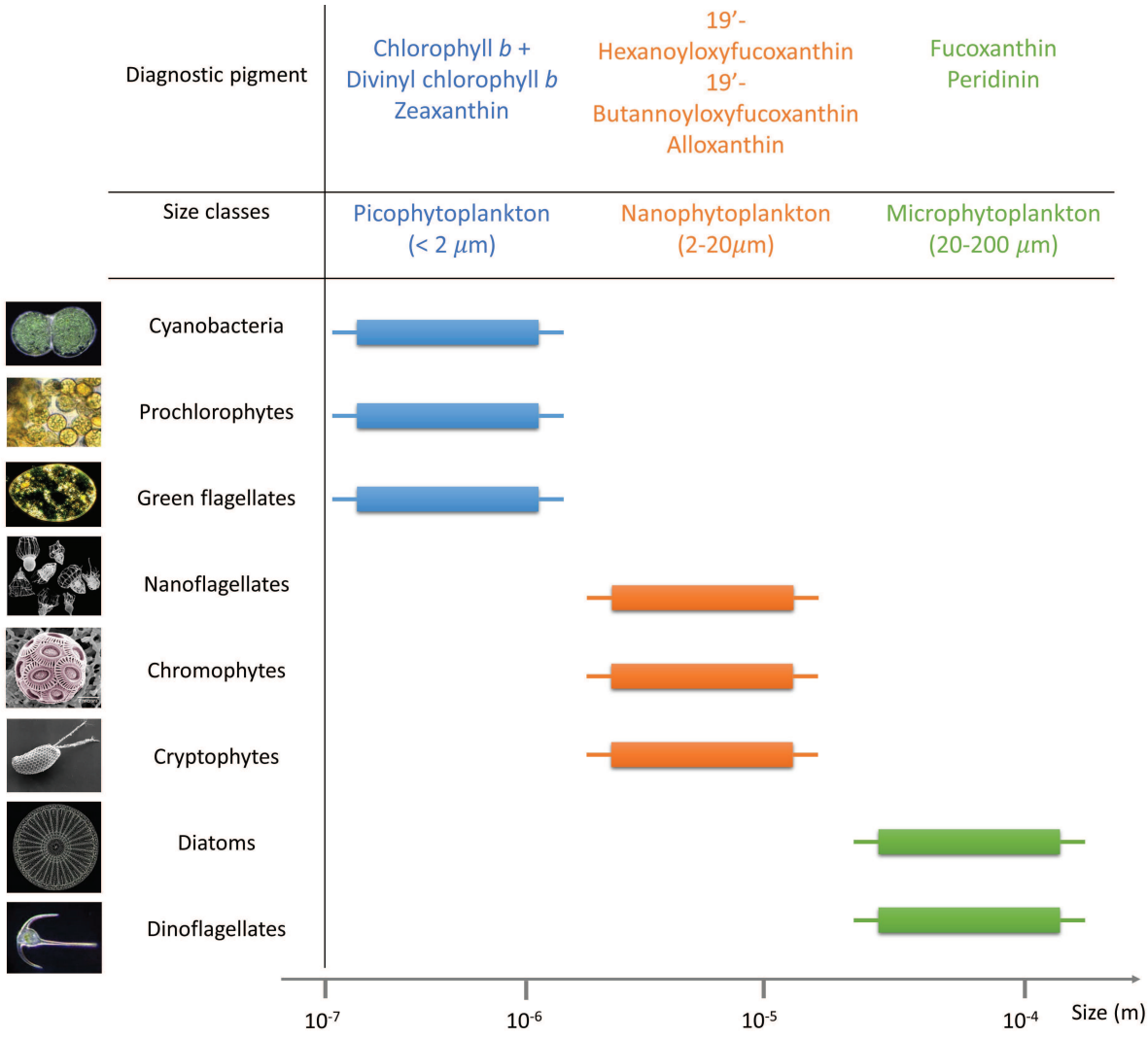


FIGURE 1.3 – *Classes de taille et pigments diagnostiques de 8 taxons phytoplanctoniques communs. Schéma adapté à partir de Vidussi* et al. *(2001).*

principalement observée **en profondeur** (i.e. à la base et sous la couche euphotique, Partensky *et al.*, 1999, Agusti, 2004).

Le nanophytoplancton

Le groupe nanophytoplanctonique, composé notamment de flagellés, chromophytes et cryptophytes (voir Figure 1.3) est un groupe **ubiquiste** dans l'océan global (Liu *et al.*, 2009). Ce groupe domine majoritairement dans les zones équatoriales ainsi que dans les zones d'upwelling. Comme pour le microphytoplancton, ce groupe phytoplanctonique peut dominer des **eaux bien mélangées et riches en nutriments** et pendant les périodes de **bloom** printanier. En effet, dans certaines zones comme l'Atlantique Nord, il est possible d'observer des alternances entre des années caractérisées par

un bloom de microphytoplancton (e.g. diatomées, voir après) et d'autres caractérisées par un bloom de nanophytoplancton (e.g. coccolithophores, Xing *et al.*, 2014).

Le microphytoplancton

Le microphytoplancton est dominé par des groupes phytoplanctoniques généralement de grande taille comme les diatomées ou les dinoflagellés (voir Figure 1.3). Le groupe microphytoplanctonique est souvent observé dans les **systèmes dynamiques, riches en nutriments** comme les zones frontales ou les upwellings et pendant les périodes de **bloom** printanier (Malone, 1980, Chisholm, 1992, Bruland *et al.*, 2001, 2005). Utilisant majoritairement les formes nouvelles de l'azote (NO_3^-) par rapport aux formes régénérées (NH_4^+), ce groupe est à la base de la production primaire dite nouvelle (Dugdale et Goering, 1967, Eppley et Peterson, 1979, Goldman, 1993). Les principales caractéristiques qui font du microphytoplancton un groupe très important dans l'export de carbone à l'échelle globale sont sa grande taille ainsi que sa densité (e.g. frustule siliceuse des diatomées pouvant jouer un rôle de « ballast », Armstrong *et al.*, 2001). Ces caractéristiques leur permettent de sédimenter rapidement, favorisant ainsi **l'exportation du carbone hors de la zone euphotique**, vers les couches profondes de l'océan (Eppley et Peterson, 1979, Michaels et Silver, 1988). Il est important de noter que l'export de carbone peut se faire à travers d'autres processus comme par exemple le « grazing » (i.e. broutage) du phytoplancton (microphytoplancton et autres cellules) par le zooplancton.

1.2 Étude de la biomasse phytoplanctonique à l'échelle globale

1.2.1 Mesures satellitaires « couleur de l'eau »

Du bleu limpide des eaux du Pacifique, au vert profond ou jaune-marron des eaux côtières, jusqu'au rouge des zones affectées par les efflorescences algales toxiques, l'océan est caractérisé par une large **gamme de couleurs**. Cette palette de couleurs dépend directement de la quantité et de la nature du matériel en suspension et dissous dans l'eau de mer. Le vert de certaines eaux océaniques est due à la forte teneur en **chlorophylle *a***, tandis que la couleur marron-jaune des eaux côtières est due à leur forte teneur en matériel dissous et/ou particules minérales.

La **télédetection de la « couleur de l'océan »** permet d'estimer, depuis l'espace, le contenu en

$[Chl]$ des eaux océaniques de surface de manière **quasi-synoptique**. En effet, les mesures satellitales couleur de l'océan (OCR pour « Ocean Color Radiometry », Gordon et McCluney, 1975) sont acquises avec une **haute résolution** spatiale (quelques kilomètres en moyenne), une haute résolution temporelle (journalière) et permettent d'obtenir une couverture pratiquement globale des océans. La télédetection est un outil adapté à l'étude des processus physiques et biogéochimiques régulant la composition et la distribution de la biomasse phytoplanctonique qui sont caractérisés par une forte variabilité spatio-temporelle. La mesure de la $[Chl]$ obtenue par OCR repose sur la dépendance de la couleur de la lumière émergeant de l'océan vis à vis des **pigments** présents dans le phytoplancton (Yentsch, 1960). Dans l'océan ouvert, la **réflectance** (rapport entre éclairement entrant et sortant à la surface de l'eau) mesurée à 440 nm, longueur d'onde du pic d'absorption de la chlorophylle *a*, diminue quand la $[Chl]$ augmente (Morel et Antoine, 1994). C'est en se basant sur cette dépendance que les algorithmes qui permettent d'estimer la $[Chl]$ à la surface des océans ont été développés (e.g. O'Reilly *et al.*, 1998).

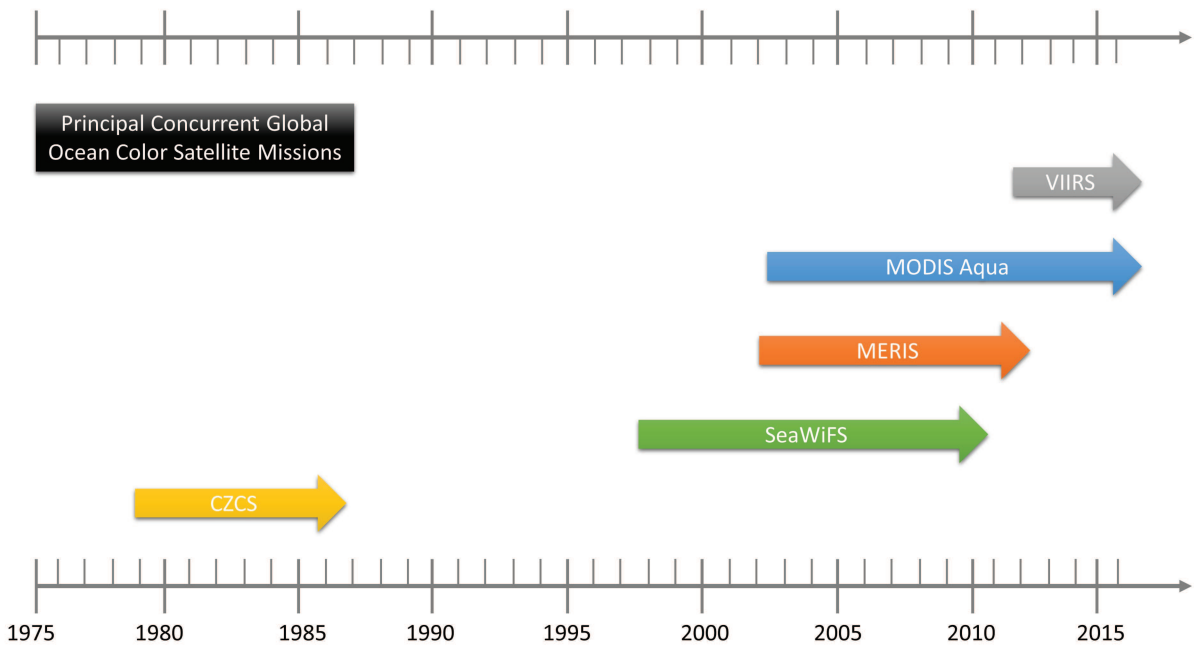


FIGURE 1.4 – *Principales missions globales « couleur de l'océan » de 1975 à de nos jours (satellites LEO : « Low Earth Orbites »). MODIS Aqua et VIIRS sont toujours en mission.*

Depuis la mise en œuvre des techniques satellitales d'observation de la couleur de l'océan, plus de 25 ans d'observations OCR ont été acquises en cumulé et la $[Chl]$ est mesurée à la surface des

océans de manière continue depuis plus de 15 ans (voir Figure 1.4). Le capteur CZCS (Coastal Zone Color Scanner, Feldman *et al.*, 1989) lancé par la NASA (National Aeronautics and Space Administration) a été le premier capteur satellite couleur de l'océan opérationnel d'octobre 1978 à juin 1986. Il a été remplacé 10 ans plus tard par le capteur SeaWiFS (Sea-viewing Wide Fiel-of-view Sensor, Hooker et Esaias, 1993, McClain *et al.*, 2004) qui a fonctionné de septembre 1997 à décembre 2010. Le capteur MERIS (Medium Resolution Imaging Spectrometer, Rast *et al.*, 1999), lancé par l'ESA (European Spatial Agency) a quant à lui été opérationel de mars 2002 à avril 2012. Aujourd'hui les capteurs **MODIS Aqua** (Moderate Resolution Imaging Spectroradiometer, Salomonson *et al.*, 1987, Savtchenko *et al.*, 2004) lancé en 2002 et **VIIRS** lancé en 2011 (Arnone *et al.*, 2012, Visible Infrared Imaging Radiometer Suite,) sont toujours en fonctionnement. MODIS Aqua a notamment permis la continuité des observations avec SeaWiFS (Franz *et al.*, 2005). Ces derniers capteurs qui ont de meilleures résolutions radiométriques spectrales et spatiales sont désormais devenus les outils essentiels de quantification de la biomasse phytoplanctonique océanique (McClain, 2009). Des progrès ont également été réalisés pour dériver d'autres produits tels que le **carbone organique particulaire** (POC pour « Particulate Organic Carbon », e.g. Stramski *et al.*, 1999, Stramska, 2009, Gardner *et al.*, 2006), la distribution de la **taille des particules** (Loisel *et al.*, 2006, Kostadinov *et al.*, 2009) ou encore la **composition des communautés phytoplanctoniques** (Alvain *et al.*, 2005, Ciotti *et al.*, 2002, Sathyendranath *et al.*, 2004).

Malgré la couverture spatio-temporelle **quasi-synoptique** que nous offrent les observations satellitales, les données couleur de l'eau ont leurs limites. En effet, les données OCR ne proviennent pas de mesures *in situ* directes mais sont des **produits dérivées de modèles**. Les données *in situ* sont donc essentielles pour valider les algorithmes d'inversion. De plus, les données couleur de l'eau ne nous renseignent que sur la **couche superficielle de l'océan**, définie comme étant la couche partant de la surface jusqu'à la profondeur dite de pénétration (Z_{pd}, Gordon et McCluney, 1975). Cette couche superficielle ne représente que $1/5^{\text{ème}}$ **de la couche euphotique**, zone où l'on considère que la photosynthèse peut être réalisée par les organismes phytoplanctoniques (Morel et Berthon, 1989). La profondeur de la couche euphotique, Z_e est la profondeur à laquelle l'éclairement est réduit à 1% de sa valeur de surface. Les satellites n'observent donc pas les $4/5^{\text{ème}}$ de la zone productive de l'océan, cette information à l'échelle globale est primordiale pour une meilleure évaluation des stocks de bio-

masse et des flux de carbone. Un des challenges actuels en océanographie est d'étudier en plus de la distribution en surface des océans, la **distribution verticale** de la biomasse phytoplanctonique avec la même résolution spatio-temporelle que les observations satellitaires. Pour résumer, on ne peut plus se contenter d'avoir une vue en 2 dimensions (i.e. 2D) de la biomasse phytoplanctonique en surface, il faut étendre ces observations 2D en profondeur pour obtenir à terme **une vue en trois dimensions** (i.e. 3D) de la biomasse phytoplanctonique océanique.

1.2.2 Mesures *in situ* discrètes

La distribution verticale de la biomasse phytoplanctonique est notamment étudiée via des **mesures *in situ* discrètes**. Ces mesures résultent du prélèvement d'échantillons d'eau de mer collectés à différentes profondeurs et ensuite filtrés sur un bateau lors d'une campagne océanographique. Les filtres sont ensuite analysés en laboratoire, les **mesures discrètes** résultantes sont généralement appelées « données bouteilles » en référence aux bouteilles Niskin utilisées pour prélever l'eau.

Au début de l'océanographie, la méthode traditionnelle utilisée pour identifier le phytoplancton était la **microscopie**. Elle permet de déterminer le type de phytoplancton en se basant sur sa morphologie (Tomas, 1997). De nouveaux outils complémentaires et permettant d'analyser un plus grand nombre d'échantillons ont ensuite émergés pour étudier et quantifier la biomasse phytoplanctonique. Parmi ces outils, la **cytométrie en flux** permet de caractériser les cellules en fonction de leur autofluorescence ainsi que de leurs propriétés de diffusion de la lumière. Cette méthode permet de dériver et quantifier certaines propriétés des cellules comme leur taille, leur pigmentation ou encore leur appartenance taxonomique (e.g. *Prochlorococcus, Synechococcus*, pico-eucaryotes, diatomées, coccolithophores, Olson *et al.*, 1989, Collier, 2000). Les **méthodes moléculaires** permettent quant à elles de séquencer les gènes pour caractériser de manière qualitative la diversité des différentes communautés phytoplanctoniques. **L'analyse pigmentaire** repose sur des techniques variées comme la fluorimétrie *in vitro* (Holm-Hansen *et al.*, 1965), la spectrophotométrie (Lorenzen, 1967) ou la chromatographie (Mantoura et Llewellyn, 1983) qui permettent d'estimer la concentration en chlorophylle *a*. Finalement, **l'analyse chromatographique** (**HPLC** pour « High Performance Liquid Chromatography ») des pigments est certainement l'approche la plus **précise** pour quantifier la concentration en chlorophylle *a* et la plus communément utilisée pour caractériser les communautés phytoplanctoniques

(Jeffrey *et al.*, 1997, Roy *et al.*, 2011).

Les **pigments** présents dans les cellules phytoplanctoniques peuvent être classés en trois groupes majeurs : les chlorophylles (a,b et c), les caroténoïdes (carotènes et leurs dérivés oxygénés connus sous le nom de xanthophylles) et les biliprotéines (phycoerythrines, phycocyanines et allophycocyanines). La **chlorophylle *a*** est un pigment présent dans tous les organismes phototrophes tandis que la distribution des autres pigments varient en fonction des groupes phytoplanctoniques. Certains pigments sont représentatifs d'un taxon et peuvent donc être utilisés comme pigments marqueurs, ou **pigments diagnostiques**, d'un groupe phytoplanctonique donné (voir Figure 1.3, Jeffrey *et al.*, 1997). En plus de quantifier de manière très précise la $[Chl]$, l'HPLC permet donc d'estimer la contribution des différents groupes phytoplanctoniques, notamment en terme de **classes de taille** (i.e. micro-, nano- et pico-pytoplancton, Claustre, 1994, Vidussi *et al.*, 2001, Uitz *et al.*, 2006). L'HPLC a donc l'avantage par rapport à la microscopie, la cytométrie et la génétique, de produire une description **quantitative** de la biomasse des grands groupes taxonomiques.

Sept pigments diagnostiques majeurs sont reconnus comme étant représentatifs de différents groupes phytoplanctoniques : la fucoxanthine, la peridinine, la 19'-hexanoyloxyfucoxanthine, la 19'-butanoyloxyfucoxanthine, l'alloxanthine, la chlorophylle *b* associée au divinyl chlorophylle *b* et la zeaxanthine (voir les abréviations dans la Table 1.1). Les fractions de concentration en chlorophylle *a* associées aux trois classes de taille phytoplanctoniques (f_{micro}, f_{nano} et f_{pico}) sont dérivées selon les équations suivantes (Uitz *et al.*, 2006) :

$$f_{micro} = \frac{1.41[Fuco] + 1.41[Perid]}{\sum DP_w} \tag{1.1}$$

$$f_{nano} = \frac{1.27[Hex-fuco] + 0.35[But-fuco] + 0.60[Allo]}{\sum DP_w} \tag{1.2}$$

$$f_{pico} = \frac{1.01[TChlb] + 0.86[Zea]}{\sum DP_w} \tag{1.3}$$

avec $\sum DP_w$ la somme des concentrations des sept pigments diagnostiques :

$$\begin{aligned}\sum DP_w = 1.41[Fuco] + 1.41[Perid] + 1.27[Hex-fuco] \\ +0.35[But-fuco] + 0.60[Allo] + 1.01[TChlb] + 0.86[Zea]\end{aligned} \tag{1.4}$$

Ensuite, il est facilement possible de dériver la concentration en chlorophylle *a* associée à chaque

groupe phytoplanctonique ($[microChl]$, $[nanoChl]$ et $[picoChl]$) :

$$[microChl] = f_{micro} \cdot [TChl] \tag{1.5}$$

$$[nanoChl] = f_{nano} \cdot [TChl] \tag{1.6}$$

$$[picoChl] = f_{pico} \cdot [TChl] \tag{1.7}$$

avec $[TChl]$ la concentration en chlorophylle *a* associée à la biomasse phytoplanctonique totale. Cette approche est reconnue comme étant un **outil pertinent** et communément utilisée pour étudier la distribution des groupes de taille phytoplanctoniques à **l'échelle globale** (e.g. Aiken *et al.*, 2009, Ras *et al.*, 2008, Uitz *et al.*, 2006).

TABLE 1.1 – *Pigments diagnostiques utilisés dans les Equations 1.1, 1.2,1.3 et 1.4 et leurs abréviations*

Pigments diagnostiques	Abréviations
Fucoxanthine	Fuco
Peridinine	Peri
19'-hexanoyloxyfucoxanthine	Hex-fuco
19'-butanoyloxyfucoxanthine	But-fuco
Alloxanthine	Allo
Chlorophylle *b* + divinyl chlorophylle *b*	TChl*b*
Zeaxanthine	Zea

Si les mesures HPLC permettent une **description précise** du contenu pigmentaire, le coût, le temps et l'expertise que demande cette méthode ne permettent ni une **couverture spatiale globale** de l'océan, ni une **description complète de la colonne d'eau**. En effet, la complexité d'échantillonnage de certaines zones due aux conditions météorologiques (e.g. l'hiver en Atlantique Nord) et/ou à la distance des zones étudiées (e.g. océan Austral) implique un biais dans l'effort d'échantillonnage global (voir Figure 1.5). Ces contraintes ont conduit la communauté scientifique à développer des **méthodes statistiques** pour améliorer la **couverture spatiale et temporelle** de ces données *in situ*. Dans ce contexte, certaines études ont donc été menées pour relier les données satellitales couleur de l'eau, acquises à haute résolution spatio-temporelle, avec les données HPLC afin d'établir une paramétrisation de la distribution verticale de la $[Chl]$ ainsi que des communautés phytoplanctoniques associées pour l'océan global (Uitz *et al.*, 2006, Morel et Berthon, 1989).

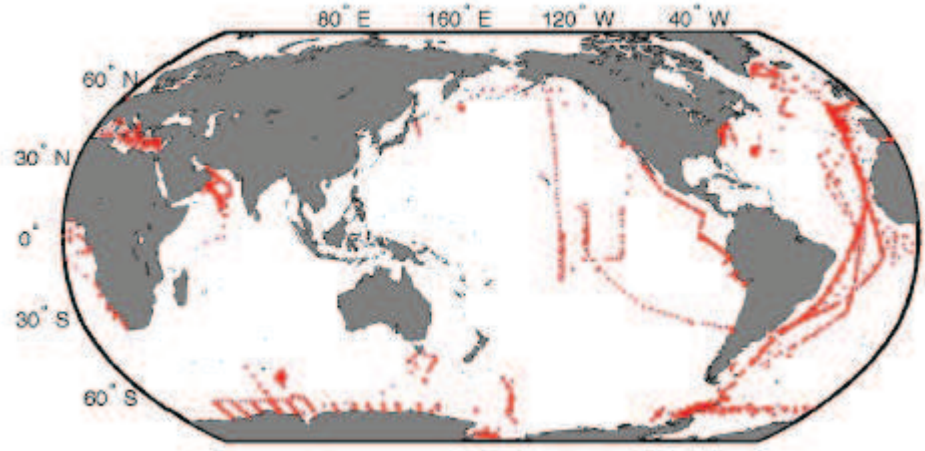


FIGURE 1.5 – *Distribution géographique en surface des 35 634 échantillons de pigments analysés par HPLC de la base de données MAREDAT regroupant toutes les données HPLC pour l'océan global. Source : Peloquin* et al. *(2013).*

1.2.3 Mesures *in situ* à haute fréquence

La **fréquence d'échantillonnage** (temporelle et spatiale) des mesures discrètes *in situ* utilisées pour étudier la distribution verticale de la biomasse phytoplanctonique (voir ci-dessus, Sect. 1.2.2) a été fortement améliorée grâce à des méthodes d'échantillonnage alternatives reposant sur la mesure d'estimateurs **bio-optiques**.

Les constituants biogéochimiques présents dans l'eau de mer interagissent en effet avec la lumière au travers des processus de diffusion et d'absorption. Ainsi, l'étude des **propriétés optiques** peut fournir des informations pertinentes sur l'abondance et la nature des substances et des particules microscopiques en suspension dans l'eau de mer : divers composés chimique dissous, micro-organismes (virus, bactéries hétérotrophes, phytoplancton, etc.) et particules non-vivantes (organiques et minérales). La majorité de ces constituants absorbe et diffuse la lumière d'une façon particulière. Grâce à des **capteurs bio-optiques**, il est désormais possible de mesurer ces propriétés *in situ* de façon **continue** et de **manière non-intrusive**. Ces proxies de la biomasse phytoplanctonique dans l'océan global sont mesurés avec une résolution spatiale, temporelle et verticale plus fine que celles qu'autorisent les méthodes traditionnelles présentées ci-dessus (voir Sect. 1.2.2).

Nous présenterons dans cette partie deux **estimateurs bio-optiques de la biomasse phytoplanctonique** sur lesquels nous avons travaillé au cours de cette thèse : 1) la fluorescence de la chlorophylle *a* et 2) le coefficient de rétrodiffusion particulaire.

La fluorescence de la chlorophylle *a*

La fluorescence de la chlorophylle *a* (notée fluorescence par la suite) est une mesure bio-optique qui permet d'estimer la $[Chl]$ dans l'eau de mer de manière non intrusive (Lorenzen, 1966). Cette mesure repose sur la propriété de **fluorescence de la molécule de chlorophylle *a*** : lorsque celle-ci a absorbé des photons dans la partie **bleue** du spectre, elle passe à un état excitée puis revient à son état initial en réémettant des photons dans la partie **rouge** du spectre (i.e. phénomène de relaxation). La quantité de photons émis pendant ce processus peut être liée à la $[Chl]$ dans la colonne d'eau selon la formule suivante :

$$f = Ea^*\phi[Chl] \tag{1.8}$$

où f est la fluorescence (mole quanta $m^{-3}s^{-1}$) ; E l'éclairement d'excitation (mole quanta $m^{-2}s^{-1}$) ; $[Chl]$ la concentration en chlorophylle *a* ($mg.m^{-3}$) ; a^* le coefficient d'absorption phytoplanctonique spécifique de la chlorophylle *a* $[m^2(mg.m^{-3})^{-1}]$ et ϕ le rendement quantique de fluorescence [mole quanta émise $(\text{mole quanta absorbée})^{-1}$]. La fluorescence est une méthode à **haute résolution**, rapide et peu coûteuse permettant d'acquérir de manière **quasi-continue** la $[Chl]$ sur la dimension verticale. Ce sont ces différents avantages qui font de la fluorescence la méthode la plus utilisée pour estimer la $[Chl]$ et certainement l'une des propriétés bio-optiques les plus mesurées dans l'océan ouvert.

Bien que les profils de fluorescence acquis dans l'océan global soient très nombreux, ils n'ont toujours pas été exploités dans leur globalité car la relation entre la $[Chl]$ et la fluorescence est complexe. En effet, le **rapport** fluorescence/$[Chl]$ n'est pas toujours constant car le coefficient d'absorption phytoplanctonique spécifique de la chlorophylle *a*, a^*, et le rendement quantique de fluorescence, ϕ, (voir Equation 1.8) **varient avec les conditions du milieu** (Falkowski *et al.*, 1985). Le coefficient a^* est dépendant de plusieurs facteurs comme la structure de taille de l'assemblage phytoplanctonique au travers du « package effect » ou effet de discrétisation (Morel et Bricaud, 1981, Sathyendranath *et al.*, 1987) et la contribution en pigments accessoires par rapport à la $[Chl]$ (e.g. Bricaud *et al.*, 2004, 1995). La plus grande source de variabilité de ϕ est due aux changements d'éclairement ambiants, y compris sur des échelles de temps très courtes comme lors d'un passage nuageux (Abbott *et al.*, 1982) ou sur des échelles de temps plus importantes comme la fluctuation diurne (Claustre *et al.*, 1999). Un

exemple bien connu de cette variabilité est le « Non Photochemical Quenching » (NPQ, Kolber et Falkowski, 1993, Muller, 2001). Lorsque la lumière est très intense (e.g. en surface au midi solaire), les cellules phytoplanctoniques déclenchent des mécanismes de photo-protection qui impliquent une diminution de la fluorescence ne correspondant pas à une réelle diminution de $[Chl]$ (Cullen et Lewis, 1995, Xing *et al.*, 2012a). En parallèle de ces variations du ratio fluorescence/$[Chl]$ dues aux conditions du milieu, les différents types de fluorimètres avec des plages spectrales d'excitation et d'émission spécifiques compliquent l'utilisation du signal de fluorescence. Pour toutes ces raisons, il est nécessaire de **calibrer** la fluorescence en $[Chl]$.

Ces dernières années, plusieurs procédures de calibration ont été proposées (détails dans la Sect. 2.1, e.g. Lavigne *et al.*, 2012, Xing *et al.*, 2011). Ces méthodes requièrent toutes des mesures additionnelles (*in situ* ou satellitales). Cependant, très peu de profils de fluorescence sont accompagnés de ces mesures complémentaires dans les bases de données « historiques ». Il reste donc encore à développer une méthode de calibration indépendante de toutes mesures additionnelles et applicable dans n'importe quelles conditions, afin d'utiliser la **globalité** des profils de fluorescence acquis depuis le début des années 60 jusqu'à maintenant.

Le coefficient de rétrodiffusion particulaire

Le coefficient de rétrodiffusion particulaire, $b_{bp}(\lambda)$, est caractéristique des **particules submicroniques** (e.g. détritus, bactéries, virus), c'est-à-dire de 0.1 à $1\mu m$ (Morel et Ahn, 1991), voire jusqu'à $10\mu m$ (Loisel *et al.*, 2007). Les capteurs de rétrodiffusion mesurent $\beta(\theta, \lambda)$ qui est définie comme **l'indicatrice de diffusion** par rapport à la direction de la propagation de la lumière θ à la longueur d'onde λ (notée VSF pour « Volume Scattering Function », $m^{-1}sr^{-1}$). Afin de retrouver la contribution de l'indicatrice de diffusion des particules, β_p, la contribution de l'eau de mer pure, β_{sw} doit être soustraite selon la formule :

$$\beta_p(\theta, \lambda) = \beta(\theta, \lambda) - \beta_{sw}(\theta, \lambda) \tag{1.9}$$

avec β_{sw} dépendante de la température et de la salinité et calculée en utilisant un coefficient de dépolarisation (Zhang *et al.*, 2009). Le coefficient de rétrodiffusion $b_{bp}(\lambda)$ est ensuite déterminé en utilisant un facteur de conversion χ associé à $\beta_p(\theta, \lambda)$ (Boss et Pegau, 2001, Kokhanovsky, 2012, Sullivan et

Twardowski, 2009) :

$$b_{bp}(\lambda) = 2\pi\chi(\beta(\theta,\lambda) - \beta_{sw}(\theta,\lambda)) \tag{1.10}$$

La dépendance spectrale du coefficient de rétrodiffusion est de la forme (Loisel *et al.*, 2006, Kostadinov *et al.*, 2009) :

$$b_{bp}(\lambda) = b_{bp}(\lambda_0) \cdot (\frac{\lambda}{\lambda_0})^{-\gamma} \tag{1.11}$$

avec λ et λ_0 deux longueurs d'onde différentes et γ le coefficient de la dépendance spectrale de b_{bp}. Connaissant b_{bp} à une longueur d'onde, il est donc possible d'en avoir une estimation à une autre longueur d'onde. A partir de l'estimation de γ extraite de données satellitaires couleur de l'eau, Loisel *et al.* (2006) ont montré que les valeurs de γ élevées sont corrélées à la proportion de petites particules (trouvées dans les régions oligotrophes des gyres subtropicaux). Inversement, les faibles valeurs de γ sont caractéristiques des zones les plus productives de l'océan.

En plus d'être sensible aux particules submicroniques (Morel et Ahn, 1991, Loisel *et al.*, 2007), certaines études récentes montrent que le b_{bp} est sensible à la gamme de taille des **cellules phytoplanctoniques** (Behrenfeld *et al.*, 2005, Huot *et al.*, 2007, 2008). En effet, le b_{bp} et la $[Chl]$ sont corrélés au premier ordre. Plusieurs études ont montré que le b_{bp} pouvait être utilisé comme un **estimateur du POC** (e.g. Balch *et al.*, 2001, Boss *et al.*, 2004, Cetinić *et al.*, 2012a, Loisel *et al.*, 2001, 2002, Stramski *et al.*, 1999, 2008). D'autres études ont montré que le b_{bp} pouvait également être utilisé comme un **proxy du carbone phytoplanctonique** (e.g. Behrenfeld *et al.*, 2005, Graff *et al.*, 2015, Martinez-Vicente *et al.*, 2013). Ainsi, b_{bp} apparaît comme étant une **grandeur biogéochimique** importante à mesurer pour estimer les **flux de carbone** dans l'océan global.

Quel est le meilleur estimateur de la biomasse phytoplanctonique ?

Le principal avantage des deux proxies que nous venons de présenter de la biomasse phytoplanctonique (i.e. fluorescence et b_{bp}) est qu'ils peuvent être mesurés facilement grâce à des **capteurs bio-optiques** effectuant des mesures **non intrusives** à haute résolution sur la colonne d'eau, permettant ainsi l'acquisition d'un profil **quasi-continu sur la verticale**.

La fluorescence est actuellement l'estimateur bio-optique le plus utilisé pour estimer la $[Chl]$. Malgré la variabilité du **rapport** $[Chl]$/carbone dépendant fortement des conditions environnemen-

tales (e.g. lumière, nutriments, température, MacIntyre *et al.*, 2002) mais aussi biotiques (statut physiologique, composition de l'assemblage), la $[Chl]$ est depuis longtemps la principale grandeur biogéochimique utilisée pour estimer la biomasse phytoplanctonique à l'échelle globale (e.g. Huot *et al.*, 2007, Henson *et al.*, 2010, Boyce *et al.*, 2014). Depuis quelques années, l'émergence de mesures **alternatives** à partir des propriétés bio-optiques (e.g. b_{bp}) ont offert de nouvelles possibilités pour estimer la biomasse phytoplanctonique (e.g. Behrenfeld et Boss, 2006). Certaines études ont démontré que la $[Chl]$ restait le meilleur estimateur de la biomasse phytoplanctonique (e.g. Huot *et al.*, 2007). Cependant, grâce aux études récemment menées qui permettent d'établir des relations relativement précises entre le carbone phytoplanctonique et certaines propriétés bio-optiques comme le b_{bp}, le débat reste encore ouvert (e.g. $R^2 = 0.69$, Graff *et al.*, 2015).

Échantillonnage des mesures bio-optiques *in situ*

Les capteurs bio-optiques peuvent être installés avec une **sonde CTD** (« Conductivity Temperature Depth ») sur une rosette. Si la fluorescence est mesurée depuis assez longtemps (années 1970) de cette façon, l'intégration des capteurs de b_{bp} sur les CTD reste relativement récente. Quoi qu'il en soit ces profils de fluorescence et de b_{bp} permettent d'apprécier des proxies de la biomasse phytoplanctonique ou des particules avec la même résolution que les mesures des propriétés physiques. Cette avancée a permis d'améliorer significativement la résolution spatio-temporelle d'échantillonnage en comparaison à l'effort d'échantillonnage possible lorsque les mesures sont discrètes (voir Sect. 1.2.2). De plus, durant les campagnes océanographiques, l'acquisition des profils quasi-continus pendant les stations permet de guider les scientifiques à la décision des profondeurs d'échantillonnage des mesures « bouteilles » (e.g. mesure à la profondeur du DCM).

Flotteurs profileurs, programme Argo et émergence du programme Bio-Argo

Dans les années 90, le **programme Argo** dédié à l'observation des propriétés physiques des océans a été lancé (Roemmich *et al.*, 1999). L'objectif était de développer un **réseau global** de 3000 flotteurs profileurs mesurant les **propriétés physiques** des océans depuis la surface jusqu'à 2000 m de profondeur. Le flotteur profileur est un instrument déployé à partir d'un navire océanographique ou bien un navire d'opportunité. Une fois le flotteur déployé, il se déplace **au gré des courants** pendant

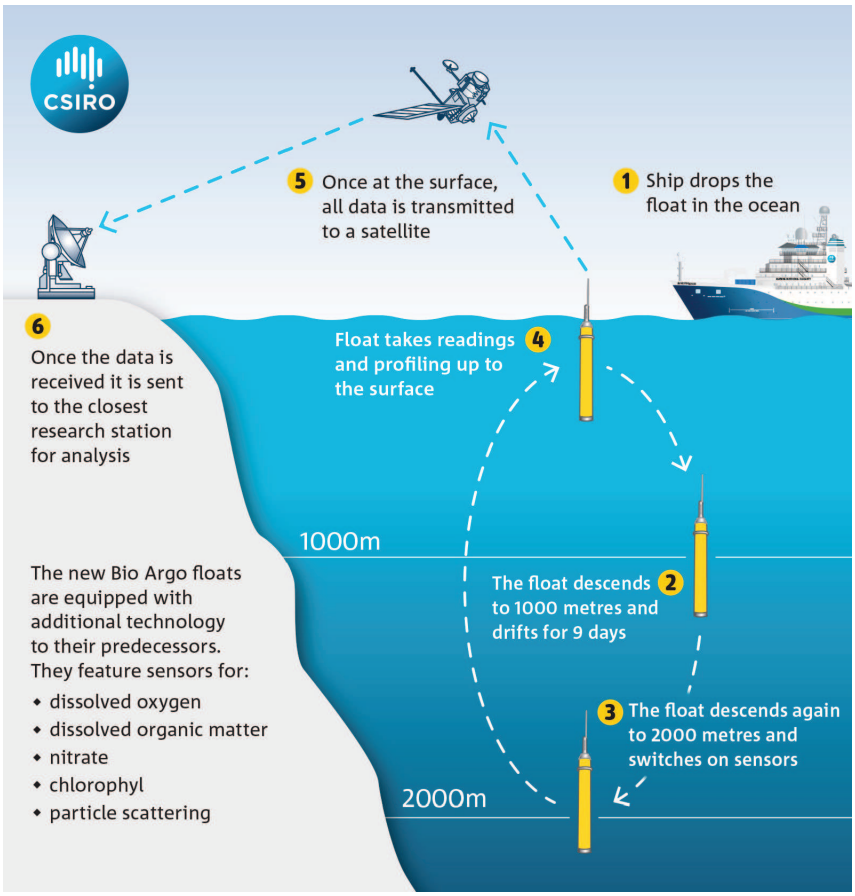


FIGURE 1.6 – *Schéma conceptuel des cycles de mesures d'un flotteur profileur. Source : CSIRO (`https://blogs.csiro.au`). Le cycle d'un flotteur profileur est également décrit dans l'animation `http://www.monoceanetmoi.com/web/index.php/fr/component/content/article/58-fr/ressources/animations/112-le-flotteur`*

sa période d'autonomie. Le flotteur monte et descend dans la colonne d'eau en faisant varier son volume. Le fonctionnement par cycles d'un flotteur profileur peut être résumé ainsi : 1) il reste pendant 10 jours à 1000 m de profondeur (phase de parking) ; 2) pour entamer son cycle de mesures, il descend à 2000 m de profondeur ; 3) il remonte lentement vers la surface en enregistrant des profils de température et de salinité de 2000 m à la surface ; 4) une fois en surface les données sont transmises par satellite aux scientifiques pour analyse. Le flotteur replonge ensuite pour entamer un nouveau cycle de mesures (voir Figure 1.6). Grâce à une connexion Irridium, les informations sont transmises à **double-sens**. Le flotteur envoie les données mesurées et les scientifiques peuvent envoyer certaines directives au flotteur et ainsi contrôler plusieurs paramètres de mission (e.g. profondeur de parking, temps entre deux cycles).

Fin 2007, le projet Argo a atteint son objectif initial avec 3 000 flotteurs en opération qui fournissent chaque année ~100 000 profils de température et de salinité, soit **vingt fois plus** que le nombre

de profils collectés à bord des navires de recherche (voir Figure 1.7-a). Ces ~ 100 000 profils Argo contribuent pour 95% de l'ensemble des profils hydrologiques acquis annuellement. Ces observations, accessibles **quasiment en temps réel**, couvrent **toutes les saisons et toutes les zones géographiques**, y compris les hautes latitudes très peu accessibles aux navires océanographiques en hiver. Par exemple, Argo a acquis plus de données au Sud de 30° Sud durant un seul hiver austral que durant toute la période pré-Argo (Roemmich *et al.*, 2009a).

Devant le succès incontestable de ce projet, les océanographes biogéochimistes se sont lancé le défi de mettre en place un réseau d'observation comparable. Ainsi, des **capteurs biogéochimiques** consommant peu d'énergie ont été intégrés aux flotteurs profileurs. Cette nouvelle génération de flotteurs (i.e. **flotteurs Bio-Argo**) permet l'acquisition *in situ* à haute résolution de **variables biogéochimiques** fondamentales comme : 1) la concentration en **oxygène dissous** (Körtzinger *et al.*, 2004), 2) la concentration en **nitrates** (Riser et Johnson, 2008), 3) les propriétés **bio-optiques** (i.e. fluorescence de la chlorophylle, coefficient de rétrodiffusion particulaire, coefficient d'atténuation et fluorescence du CDOM pour « Color Dissolved Organic Material », Claustre *et al.*, 2010b), 4) la **radiométrie** (e.g. éclairement incident et lumière disponible pour la photosynthèse PAR pour « Photosynthetically Available Radiation » Xing *et al.*, 2014) et 5) la mesure du **pH** (capteurs intégrés très récemment). Nous n'utiliserons pendant cette thèse seulement les flotteurs Bio-Argo mesurant les propriétés **bio-optiques**.

Depuis 2008, les flotteurs Bio-Argo mesurent *in situ* plusieurs estimateurs bio-optiques de la biomasse phytoplanctonique, notamment dans des régions océaniques très éloignées (voir Figure 1.7-b, Bishop *et al.*, 2002, Boss *et al.*, 2008, Bishop et Wood, 2009, Whitmire *et al.*, 2009, Claustre *et al.*, 2010b). Les mesures provenant de ces plateformes autonomes permettent de couvrir une large gamme d'échelles spatiales (méso-échelle à régionale) et temporelles (diurne, saisonnière, annuelle et pluriannuelle) encore non étudiées complétant ainsi les dispositifs plus traditionnels d'échantillonnage reposant jusqu'à maintenant majoritairement sur des campagnes à bord de navires.

Autres plateformes autonomes bio-optiques

D'autres types de plateformes autonomes mesurent les propriétés bio-optiques *in situ* de la colonne d'eau avec une haute résolution spatio-temporelle.

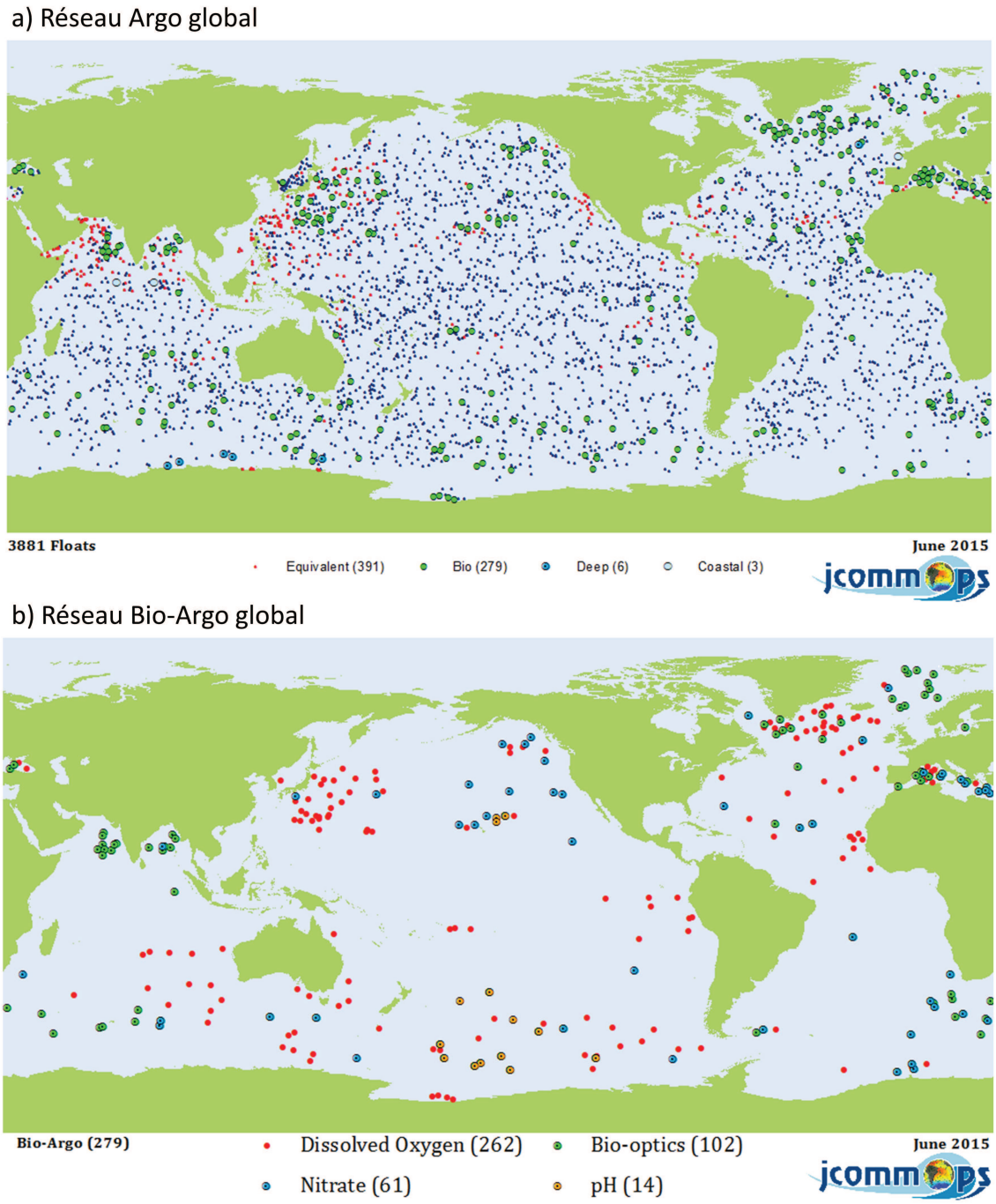


FIGURE 1.7 – *Distribution spatiale en Juin 2015 des flotteurs profileurs des programmes : a) Argo et b) Bio-Argo. Source : www.jcommops.org.*

Les **éléphants de mer** équipés d'un fluorimètre permettent l'acquisition en plus des données hydrologiques (profils température/salinité) de profils de fluorescence à chacune de leurs plongées et remontées (Guinet *et al.*, 2012). Ces données ont permis de mieux étudier la distribution verticale de la biomasse phytoplanctonique dans l'océan Austral (région de Kerguelen), où peu de données *in situ*

étaient accessibles (Xing *et al.*, 2012a).

Les **gliders** (i.e. planeurs sous-marins) sont des plateformes autonomes se déplaçant dans la colonne d'eau sur le même principe (changement de volume) que les flotteurs profileurs. Leur spécificité est qu'ils peuvent être dirigés par le scientifique pendant sa mission. Ainsi, ce type de plateforme est généralement utilisé pour étudier à haute résolution les propriétés hydrologiques et biogéochimiques correspondant à des échelles allant du bassin à la sub-mésoéchelle (e.g. Niewiadomska *et al.*, 2008). Etant donné que les gliders mesurent les propriétés bio-optiques à très haute fréquence sur des périodes relativement courtes par rapport au flotteur (quelques mois) et des zones très ciblées, nous n'avons pas utilisé les données provenant de ce type de plateformes qui auraient pu biaiser nos analyses qui se voulaient globales.

Les plateformes autonomes mesurant la fluorescence (i.e. flotteurs Bio-Argo équipés de capteurs bio-optiques et éléphants de mer équipés de fluorimètres) ont permis d'obtenir plus de 8 000 profils de fluorescence en 2 ans, ce qui correspond à $1/6^{\text{ème}}$ de tous les profils acquis avant le lancement du programme Bio-Argo (voir Figure 1.8, Sauzède *et al.*, 2015d). Il est primordial que des méthodes soient développées pour intégrer cette nouvelle masse d'informations dans les analyses globales.

1.3 Objectifs de l'étude et description des axes de travail

Dans le contexte actuel, il est devenu essentiel de mieux quantifier et estimer les incertitudes sur les termes sources et puits de carbone dans l'océan global. En amont, avoir une **vue en trois dimensions de la biomasse phytoplanctonique et de la composition des communautés phytoplanctoniques** à l'échelle globale et régionale devient alors incontournable.

D'une part, les **données satellites** nous renseignent sur la distribution phytoplanctonique à la surface des océans de manière quasi-synoptique. Cependant, ces mesures manquent une grande partie de la zone productive de l'océan. Pour étudier la distribution verticale de la biomasse phytoplanctonique de manière précise, **l'étude pigmentaire par HPLC** est la méthode de référence. De plus, cette méthode permet d'identifier les grandes communautés phytoplanctoniques et de classifier la biomasse selon trois classes de taille. Cependant, cette méthode ne permet ni d'avoir une couverture spatiale globale, ni une description complète de la colonne d'eau. Nous avons vu que désormais

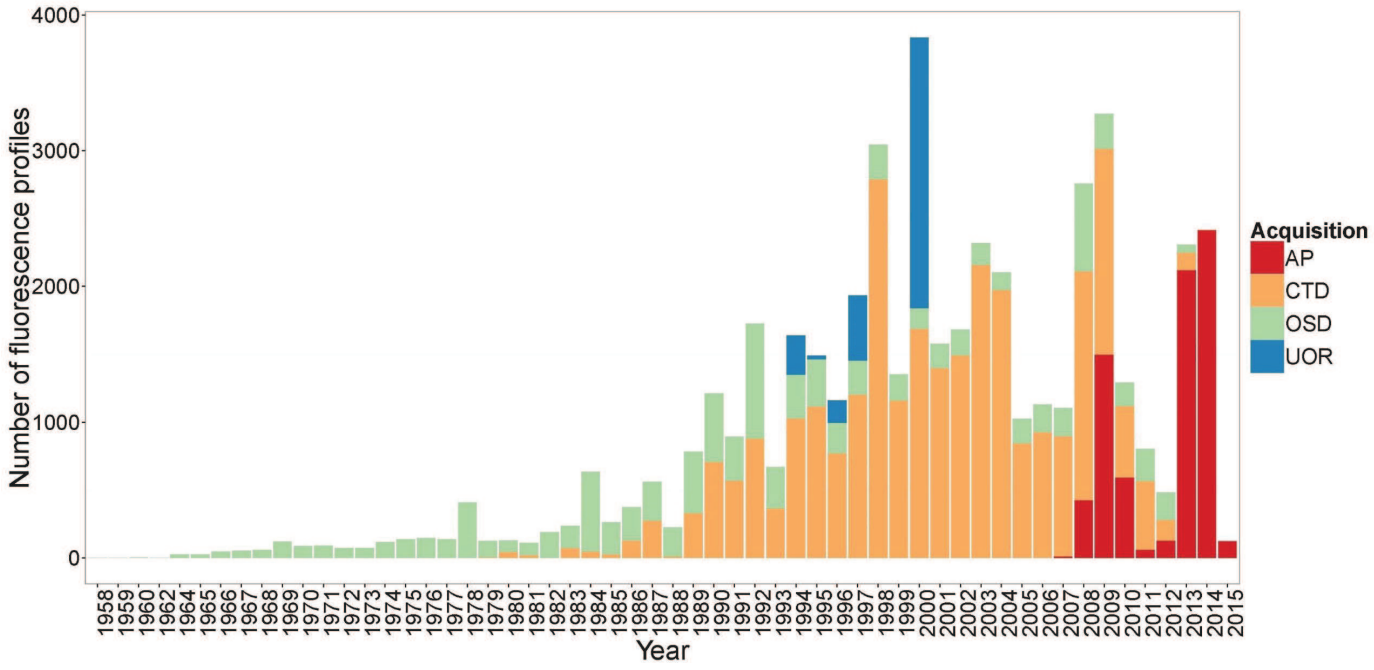


FIGURE 1.8 – *Nombre de profils de fluorescence acquis par année en fonction de différents mode d'acquisition : plateformes autonomes (AP pour « Autonomous Platforms », correspond aux flotteurs Bio-Argo mesurant les propriétés bio-optiques et aux éléphants de mer équipés d'un fluorimètre), CTD (sonde CTD équipée d'un fluorimètre), OSD (OSD pour « Ocean Station Data », correspond à des données bouteilles : profils à basse résolution verticale) et UOR (UOR pour « Undulating Oceanographic » Recorder, correspond aux profils acquis par un poisson ondulant derrière le navire océanographique). Source : Sauzède* et al. *(2015d).*

certaines grandeurs descriptives de la biomasse phytoplanctonique peuvent être acquises *in situ* de manière continue à partir de **capteurs bio-optiques**. Ces capteurs pouvant être montés sur des plateformes autonomes, le nombre de profils échantillonnés a augmenté de manière quasi-exponentielle ces dernières années. En raison des différentes contraintes que nous avons présentées, les différents estimateurs bio-optiques de la biomasse phytoplanctonique mesurés *in situ* (i.e. fluorescence et b_{bp}) à **haute résolution spatio-temporelle** n'ont pas encore été intégrés et exploités dans les analyses globales. L'objectif de cette thèse est de développer des méthodes adaptées permettant de tirer profit de ces bases de données très denses qui sont en train de se constituer, avec comme fil directeur la **reconstruction tridimensionnelle** (i.e. en 3D) de la **biomasse phytoplanctonique et de la composition des communautés phytoplanctoniques** à l'échelle globale (voir Figure 1.9).

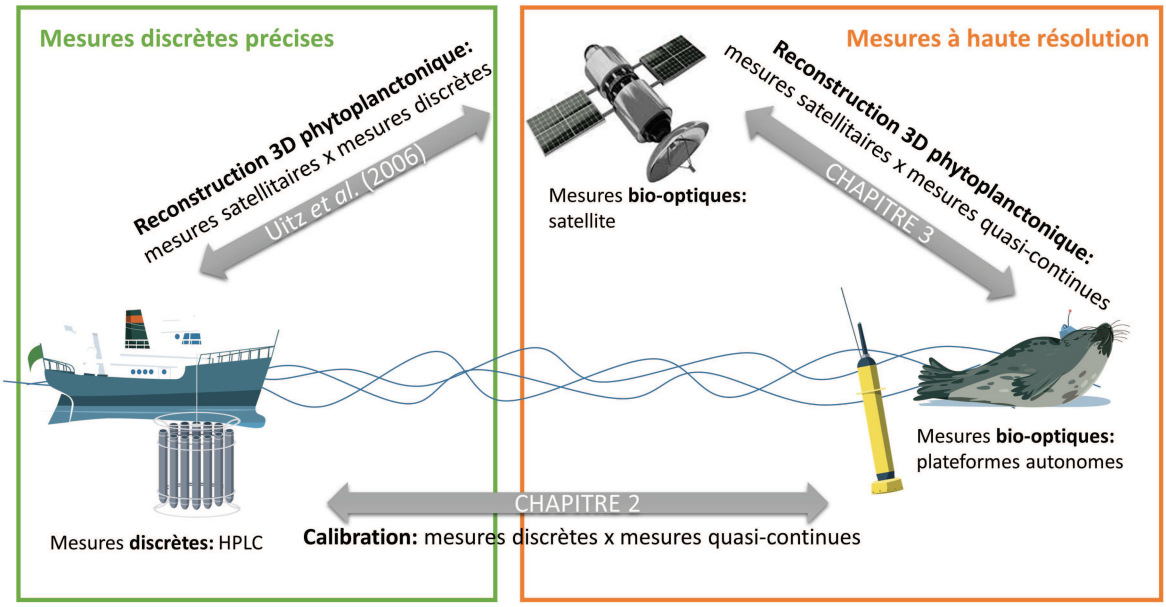


FIGURE 1.9 – *Schéma conceptuel des différents axes de travail présentés dans cette thèse.*

Vers le développement d'une climatologie 3D des communautés phytoplanctoniques pour l'océan global

Nous avons vu plus haut que la **fluorescence** de la chlorophylle *a* est sûrement la propriété bio-optique la plus mesurée dans l'océan, des milliers de profils sont donc stockés dans différentes bases de données (voir Figure 1.8). Cependant, comme nous l'avons également observé, à cause de la variabilité du rapport fluorescence/$[Chl]$, la fluorescence doit être calibrée en $[Chl]$. Il existe un nombre considérable (et croissant) de profils de fluorescence stockés dans des bases de données anciennes (e.g. acquises pendant des campagnes océanographiques) mais aussi dans des bases de données de « nouvelle génération » hébergeant les données acquises par les plateformes autonomes. Dans le premier axe de cette thèse, nous proposons une **méthode de calibration** permettent de rendre **intéropérable** tous ces profils de fluorescence accessibles et ainsi de pouvoir les utiliser dans leur globalité. Cette méthode permettra une meilleure **exploitation du potentiel** des mesures de fluorescence et pourra être utilisée dans le but d'obtenir une base de données **homogène** de $[Chl]$. Nous allons voir par la suite qu'il sera également possible de dériver la composition des communautés phytoplanctoniques à partir d'un profil de fluorescence. L'application de cette méthode permettra d'utiliser les profils de fluorescence acquis à l'échelle globale pour en dériver des premières climatologies en 3D

de la composition des communautés phytoplanctoniques et de leur biomasse pigmentaire.

Relier les données couleur de l'eau à la distribution verticale de la biomasse phytoplanctonique

Comme nous l'avons évoqué précédemment, Uitz *et al.* (2006) ont établi une relation permettant de relier les données couleur de l'eau à la distribution verticale des communautés phytoplanctoniques. Cependant, cette étude était basée sur des données discrètes (i.e. données HPLC) relativement éparses pour l'océan global. Désormais, nous pouvons envisager d'aller plus loin et de renforcer ce types d'approches en utilisant dans leur globalité le nombre de profils bio-optiques conséquent accessible à partir des plateformes autonomes. L'objectif des études que nous allons mener sera de fusionner les mesures bio-optiques mesurées à la fois par les **satellites** et par les **flotteurs profileurs** accessibles à haute résolution spatio-temporelle (en comparaison aux données bouteilles discrètes) afin d'obtenir à terme une **vision tridimensionnelle** des estimateurs bio-optiques de la biomasse phytoplanctonique pour l'océan global.

Ces deux premiers points correspondent aux **deux objectifs principaux** que nous nous sommes fixés dans le cadre de ce travail de thèse : 1) intégrer la fluorescence dans une étude globale afin d'obtenir une base de données pouvant être utilisée à terme pour obtenir une vue 3D de la distribution de la biomasse phytoplanctonique ainsi que de l'assemblage des communautés phytoplanctoniques associée à l'échelle globale et 2) établir une relation globale permettant de prédire la distribution verticale de la biomasse phytoplanctonique à partir des données couleur de l'océan. Ces deux objectifs ont respectivement fait l'objet des chapitres 2 et 3 de cette thèse. Finalement, dans le chapitre 4, les outils développés sont testés et appliqués à une région spécifique afin d'y étudier la variabilité saisonnière de la biomasse phytoplanctonique ainsi que de la production primaire. Cet exemple d'application vise à illustrer le fort potentiel des méthodes développées pendant cette thèse.
Ces études ont fait l'objet des publications suivantes :

– **R. Sauzède**, H. Claustre, C. Jamet, J. Uitz, J. Ras, A. Mignot, et F. D'Ortenzio (2015). Retrieving the vertical distribution of chlorophyll a concentration and phytoplankton community composition from in situ fluorescence profiles : A method based on a neural network with potential for global-scale applications. *J. Geophys. Res. Oceans*, **120**, doi :10.1002/2014JC010355.

– H. Lavigne, F. D'Ortenzio, M. Ribera D'Alcalà, H. Claustre, **R. Sauzède** et M. Gacic (2015). On the vertical distribution of the chlorophyll a concentration in the Mediterranean Sea : a basin-scale and seasonal approach. *Biogeosciences*, **12** :5021-5039, doi :10.5194/bg-12-5021-2015.

– **R. Sauzède**, H. Lavigne, H. Claustre, J. Uitz, C. Schmechtig, F. D'Ortenzio, C. Guinet et S. Pesant (2015). Vertical distribution of chlorophyll a concentration and phytoplankton community composition from in situ fluorescence profiles : a first database for the global ocean. *Earth System Science Data*.

– **R. Sauzède**, H. Claustre, J. Uitz, C. Jamet, G. Dall'Olmo, F. D'Ortenzio, B. Gentili, A. Poteau et C. Schmechtig (Soumis).A neural network-based method for merging ocean color and Argo data to extend surface bio-optical properties to depth : Particulate backscattering coefficient retrieval. *J. Geophys. Res. Oceans*.

– **R. Sauzède**, H. Claustre, J. Uitz, C. Jamet, F. D'Ortenzio, A. Poteau et C. Schmechtig (En preparation).A neural network for merging ocean color and Argo data to extend surface bio-optical properties to depth : Chlorophyll a concentration and phytoplankton community composition retrieval.

1.4 Méthodologie : Approche neuronale

Un problème qui se pose actuellement en océanographie est la définition et la mise en place de **méthodes numériques performantes et adaptées** pour le traitement des grosses bases de données qui sont en train de s'assembler. Afin de répondre aux différentes problématiques exposées dans la Sect. 1.3, nous avons choisi d'utiliser un outil très utilisé dans le domaine des bio-statistiques et de l'écologie en général depuis plusieurs années : les **réseaux de neurones**. Nous avons opté pour cette méthodologie statistique car les réseaux de neurones sont des **fonctions approximatrices** très flexibles, qui permettent de modéliser n'importe quel jeu de données (Marzban, 2009), et en particulier des jeux de données très **complexes et bruitées**. En effet, cet outil s'avère être très puissant pour établir des modèles, spécialement quand les relations sous-jacentes sont inconnues et non triviales.

1.4.1 Un peu d'histoire

C'est dans les années 1940, quand les mathématiciens et les biologistes ont commencé à travailler ensemble pour modéliser le cerveau humain que les premières méthodologies d'intelligence artificielle sont nées. Le concept de réseaux de neurones a vu le jour en 1943, décrit par McCulloch et Pitts (1943). C'est dans les années 80, quand la puissance du calcul informatique commence à être suffisante pour le développement de telles méthodes, que le développement des méthodologies de réseaux de neurones prend une orientation plus **mathématique et statistique**. Des publications clés mettent alors en avant la puissance des principaux types de réseaux de neurones : les **cartes auto-organisatrices de Kohonen** (Kohonen, 1990), les **réseaux interconnectés** (Hopfield, 1982, 1984) et enfin les **perceptrons multicouches** (Rumelhart *et al.*, 1986, 1988). Cet outil statistique a trouvé des applications dans de nombreux domaines liés à l'écologie et aux **bio-statistiques**. En océanographie, le potentiel des réseaux de neurones a déjà été largement testé, notamment pour estimer divers produits ou **grandeurs biogéochimiques** à la surface des océans comme le coefficient d'atténuation diffuse (K_{d490}, Jamet *et al.*, 2012), la pression partielle du dioxyde de carbone (pCO_2, Friedrich et Oschlies, 2009, Telszewski *et al.*, 2009), la concentration en pigments phytoplanctoniques (Gross *et al.*, 2000) ou bien encore différents types fonctionnels de phytoplancton (Bricaud *et al.*, 2007, Raitsos *et al.*, 2008, Ben Mustapha *et al.*, 2013, Palacz *et al.*, 2013). En effet, l'emploi des outils statistiques plus « classiques » pose problème dans ce type de disciplines à cause du nombre et de la dimension des données, ainsi que de la non-linéarité des problèmes traités. Au vu de leurs très bons résultats en prédiction de réponses biologiques, nous avons choisi dans cette thèse d'utiliser les perceptrons multicouches (**MLP** pour « Multi-Layered Perceptron ») afin de développer les méthodologies qui répondront aux problématiques soulevées dans cette thèse (voir Sect. 1.3).

1.4.2 Modèle général d'un MLP

Un réseau de neurones est composé comme son nom l'indique de plusieurs neurones permettant d'échanger des informations à travers les **connexions** qui les relient. Les connexions sont directionnelles et chacune d'elles est définie par un **poids** spécifique. Ainsi, l'information est transmise de manière unidirectionnelle du neurone j vers le neurone i, affectée d'un coefficient pondérateur $w_{i,j}$

(i.e. poids de la connexion).

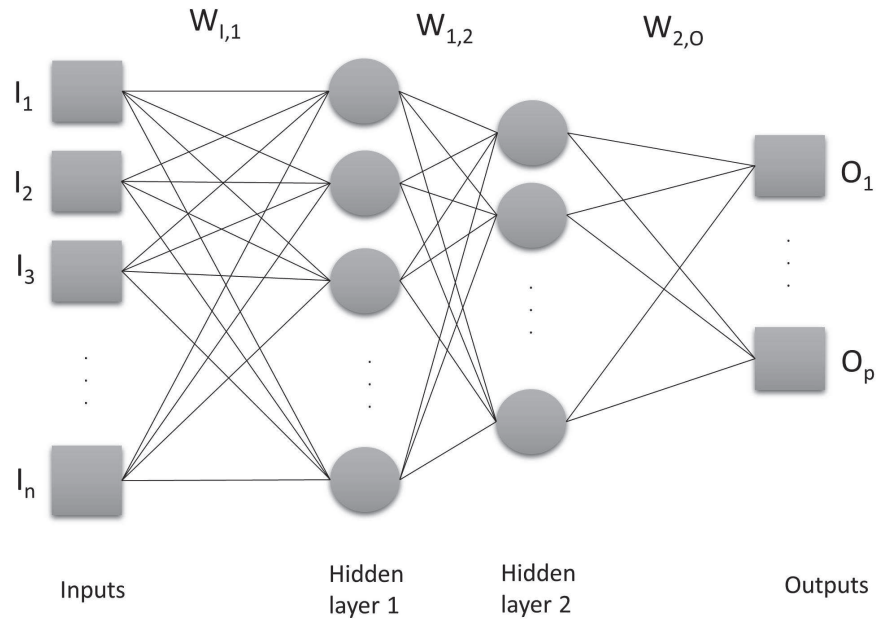


FIGURE 1.10 – *Schéma d'un perceptron multicouche (MLP) à n entrées et p sorties. Chaque droite reliant les différents neurones est définie par un poids $w_{i,j}$. Les W représentent les matrices contenant les poids $w_{i,j}$ entre chaque couche.*

De manière générale, les MLPs sont des réseaux de neurones organisés en **couches successives** (Bishop, 1995). En plus des couches dites d'entrée et de sortie, le réseau peut être composé d'une ou plusieurs **couches cachées**. Le schéma d'organisation d'un MLP à deux couches cachées est présenté sur la Figure 1.10. Les connexions sont toujours dirigées de la couche inférieure à la couche supérieure. Choisir **l'architecture** d'un MLP, appelée également topologie, revient à fixer le nombre de couches cachées ainsi que le nombre de neurones par couches.

La première couche du réseau, qui correspond à la **couche d'entrée** est définie par un vecteur contenant n neurones ($I = i_1, \ldots, i_n$). Cette couche d'entrée correspond aux données étudiées. De la même façon, les p sorties calculées par le réseau de neurones sont rassemblées dans le **vecteur de sortie** $O = o_1, \ldots, o_p$. Le transfert de l'information à travers le réseau de neurones (entre les entrées I et les sorties O) peut être formulé de la manière suivante :

$$O = f(W_{2,O} f(W_{1,2} f(W_{I,1} I))) \tag{1.12}$$

avec $W_{I,1}$, $W_{1,2}$ et $W_{2,O}$ les **matrices de poids** contenant les connexions respectivement entre les

neurones de la couche d'entrée I et de la première couche cachée, des deux couches cachées, et enfin, de la seconde couche cachée et de la couche de sortie O. Dans cette équation (1.12), f est une **fonction sigmoïde** définie selon la formule :

$$f(x) = A \cdot \frac{\exp^{\alpha x} - 1}{\exp^{\alpha x} + 1} \tag{1.13}$$

où A et α sont deux constantes. De manière à assurer un comportement quasi-linéaire entre -1 et 1, les valeurs de α et A sont traditionnellement fixées respectivement à 1.7159 et 4/3.

Ainsi, grâce à cette fonction de transfert, l'information se propage dans le réseau de neurones de proche en proche, d'une couche vers les couches supérieures. Cette partie du calcul est appelée **propagation avant**, en opposition au calcul effectué par rétro-propagation (présentée ci-après). La propagation avant permet donc de récupérer le vecteur O en sortie à partir du vecteur I d'entrée. Un MLP définit donc une fonction de R^n dans R^p.

Pour un MLP d'architecture fixé, la fonction définie par le réseau dépend des valeurs des poids $W = [w_{i,j}]$ où (i, j) donnent toutes les connexions du MLP. Une architecture génère donc une famille de fonctions :

$$\begin{gathered} R^n \to R^p \\ i \to o = f(I, W) \end{gathered} \tag{1.14}$$

Plusieurs résultats théoriques ont été établis concernant la capacité des MLPs à être des approximateurs universels de fonctions (Cybenko, 1989, Funahashi, 1989, Hornik *et al.*, 1989, Poggio et Girosi, 1990, White, 1990). Il a ainsi été démontré que toute fonction continue de R^n dans R^p peut être approximée sur un sous-ensemble de R^n par une fonction définie par un MLP à une seule couche cachée. Plusieurs études ont tout de même montré l'intérêt de considérer des MLPs à **deux couches cachées** (Lippmann, 1987, Sontag, 1992). Cependant, ces résultats théoriques ne permettent pas de déterminer le nombre de neurones nécessaires pour approximer une fonction donnée. La détermination de l'architecture optimale se fait en utilisant les **théories statistiques**.

Approximer une fonction T telle que $O = T(I)$ à partir de $F(I, W)$ revient à rechercher la

fonction $f \in F(I, W)$ telle que :

$$f(i) = f(I, W) \simeq T(i) \qquad \forall \vec{i} \in A \tag{1.15}$$

où I est le vecteur d'entrées du MLP, A est un espace $\in R^n$ et W représente l'ensemble des poids du réseau. Finalement, le problème est de déterminer les valeurs de W qui permettent d'approximer au mieux la fonction de régression que l'on recherche. Cette opération, qui correspond à l'estimation des paramètres en statistique, s'appelle la **théorie de l'apprentissage**.

1.4.3 Les différentes phases du développement d'un MLP

Il existe deux grandes phases dans la création d'un MLP : 1) la **phase d'apprentissage et de validation** et 2) la **phase de test**. La première phase permet de concevoir le réseau de neurones en utilisant 80% des données de la base de données initiale. La dernière phase teste l'efficacité du MLP à partir d'un sous-ensemble indépendant comprenant les 20% restants des données initiales (choisies aléatoirement). Cette phase détermine l'efficacité du MLP et valide le MLP en évaluant sa capacité de généralisation.

Transformation des données d'apprentissage

En écologie et plus particulièrement en océanographie, il n'est pas rare que les MLPs aient à traiter des données d'entrées de natures différentes et donc **d'ordres de grandeurs différents**. Similairement, les sorties du réseau de neurones qui ont des ordres de grandeurs différents sont numériquement difficiles à gérer. Pour que le MLP prenne en compte toutes les variables du problème de manière uniforme, il est nécessaire que chaque variable ait le même **poids numérique**. Les paramètres A et α de l'équation 1.13 sont définis de manière à ce que le domaine utile de la fonction soit $[-1; 1]$. Ainsi, les entrées et les sorties du MLP ($x_{i,j}$) sont **centré-réduites** selon la formule suivante :

$$x_{i,j} = \frac{2}{3}.\frac{x_{i,j} - \bar{x_{i,j}}}{\sigma_{i,j}} \tag{1.16}$$

où $x_{i,j}$ est la $j^{\text{ème}}$ dimension de la $i^{\text{ème}}$ donnée. $\bar{x_{i,j}}$ est la moyenne de la $j^{\text{ème}}$ dimensions et $\sigma_{i,j}$ son écart-type. Le facteur $\frac{2}{3}$ permet de ramener environ 80% des données entre -1 et 1.

Phase d'apprentissage et validation

L'algorithme d'apprentissage consiste à déterminer les poids du réseau de neurones à l'aide d'un **processus itératif** défini de la manière suivante :

$$w_{i,j}(t+1) = w_{i,j}(t) + \Delta w_{i,j}(t) \tag{1.17}$$

Avec $w_{i,j}(t+1)$ le poids qui relie le neurone j au neurone i à l'itération $t+1$. Cette valeur dépend de la valeur du poids à l'itération précédente $w_{i,j}(t)$. L'initialisation des poids du réseau (i.e. à $t = 0$) se fait aléatoirement entre -1 et 1. La **modification des poids** se fait à travers le terme $\Delta w_{i,j}(t)$, qui est déterminé en fonction d'une fonction de coût. L'apprentissage du MLP consiste à **minimiser cette fonction de coût** en utilisant une méthode de gradient. L'algorithme le plus connu, qui est utilisé dans cette étude, est l'algorithme de **rétro-propagation du gradient** (Bishop, 1995, Haykin, 1998). Pour résumer, l'algorithme de rétro-propagation peut être divisé en deux phases :

- Propagation en avant d'un stimuli p (motif d'entrainement en entrée) à travers le MLP afin de générer les sorties
- Rétro-propagation des erreurs basée sur la règle des deltas (règle d'apprentissage fondée sur la descente de gradient). Il y a donc propagation en arrière des erreurs des sorties à travers le MLP pour modifier les poids dans le sens opposé au gradient de l'erreur.

Durant la phase d'apprentissage, un phénomène appelé **sur-apprentissage** peut survenir. Ceci est dû au fait que le réseau de neurones va « trop loin » et apprend par cœur les données à restituer pendant l'apprentissage. Il est alors **mauvais en généralisation** puisqu'il n'aura pas appris la relation sous-jacente aux données mais juste une relation qui lui permet de retrouver avec précision les données d'apprentissage (voir Figure 1.11). Le sur-apprentissage peut être évité en limitant le degré de liberté du réseau par rapport à la taille de l'ensemble d'apprentissage (il faut avoir 10 à 20 fois plus de données que de poids à ajuster dans un réseau) et en arrêtant l'apprentissage avant la fin de la convergence si besoin. La **complexité** du MLP ainsi que le **temps d'arrêt de l'apprentissage** se déterminent grâce à la méthode de validation croisée (cross-validation, Bishop, 1995).

Cette méthode consiste à utiliser un sous-ensemble de la base de données d'apprentissage comme base de données de validation du MLP. Lorsqu'il y a sur-apprentissage, les performances du MLP

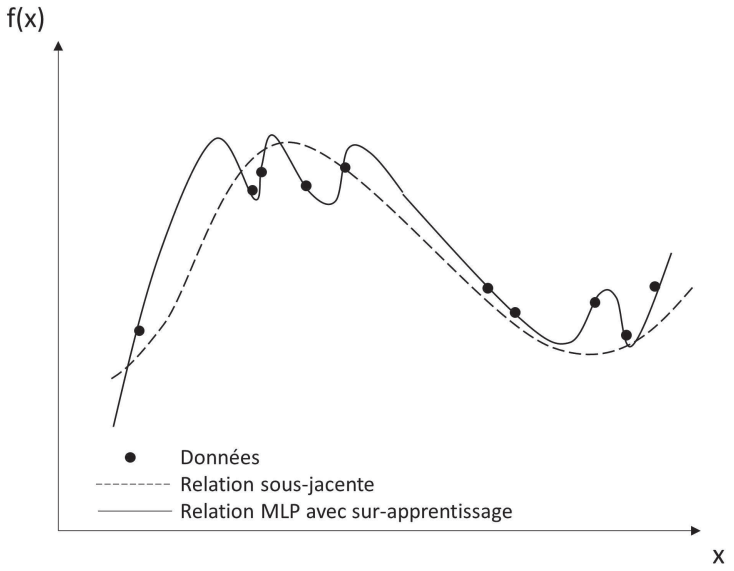


FIGURE 1.11 – *Exemple de sur-apprentissage d'un MLP approximant une fonction $f(x)$ à partir de données réelles.*

continuent à s'améliorer sur l'ensemble d'apprentissage tandis qu'elles se dégradent sur l'ensemble de validation (voir Figure 1.12). L'itération à laquelle l'erreur de validation commence à augmenter correspond au **minimum de l'erreur** faite sur l'ensemble de validation, c'est donc le moment optimal pour arrêter l'apprentissage.

Phase de test

Les performances réelles du réseau de neurones sont évaluées à partir d'un troisième ensemble de données, indépendant des ensembles d'apprentissage et de validation : c'est la base de données dite de test. Cet ensemble sert à vérifier que le MLP généralise bien et que la performance du modèle ne dépend pas de l'ensemble de validation.

La performance du MLP en phase de test est alors déterminée à partir de trois paramètres statistiques principaux :

– le coefficient de détermination, R^2, du modèle linéaire établi entre les valeurs calculées par le réseau de neurones, y_{cal}, et les valeurs réelles observées, y_{obs} :

$$R^2 = \frac{1}{N}\sum_{i=1}^{n}\frac{(y_{obs} - \bar{y_{obs}}) \cdot (y_{cal} - \bar{y_{cal}})}{\sigma_{y_{obs}} \cdot \sigma_{y_{cal}}} \tag{1.18}$$

où N est le nombre de données, $\bar{y_{obs}}$ et $\bar{y_{calc}}$ sont respectivement les moyennes des valeurs ob-

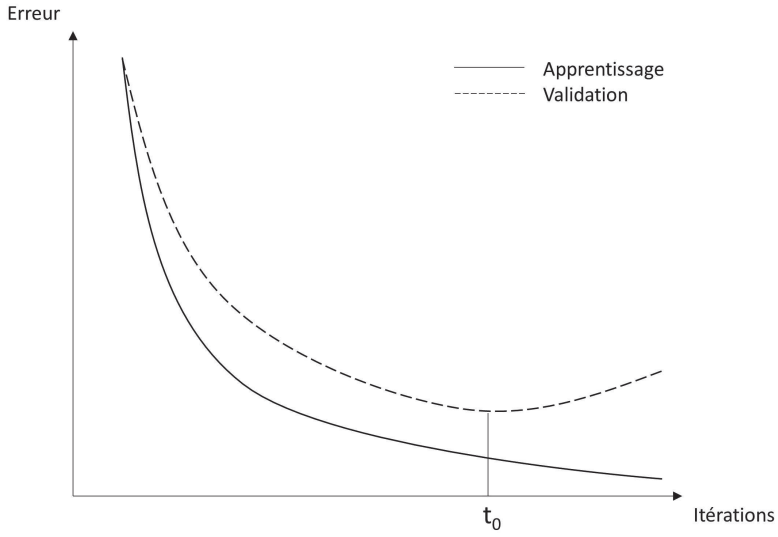


FIGURE 1.12 – *Validation croisée : évolution des erreurs d'apprentissage et de validation au cours des itérations pendant l'apprentissage d'un MLP. L'itération t_0 correspond à l'itération à laquelle l'apprentissage doit être arrêté pour éviter le phénomène de sur-apprentissage.*

servées (réelles) et calculées par le MLP et $\sigma_{y_{obs}}$ et $\sigma_{y_{cal}}$ leurs écart-types. Le R^2 nous renseigne sur la corrélation entre les valeurs prédites et observées.

– la pente de ce modèle linéaire qui nous renseigne sur le biais entre les valeurs prédites et observées

– un dernier paramètre nous renseignant sur l'erreur absolue de prédiction, MAPD, pour « Median Absolute Percent Difference », calculé selon la formule suivante :

$$MAPD = \text{médiane}(\frac{|y_{calc} - y_{obs}|}{y_{obs}}) \tag{1.19}$$

Ce sont principalement ces trois différents paramètres statistiques qui seront utilisés afin de valider les modèles neuronaux établis tout au long de cette thèse. De plus, pour comparer nos modèles aux modèles déjà existants, nous utiliserons la racine carrée de l'erreur quadratique (RMSE pour « Root Mean Squared Error »). La RMSE est calculée selon la formule suivante :

$$RMSE = \sqrt{\frac{\sum_{k=1}^{N}(y_{obs,k} - y_{calc,k})^2}{N}} \tag{1.20}$$

avec $y_{obs,k}$ la $k^{\text{ème}}$ valeur réelle observée et $y_{calc,k}$ la $k^{\text{ème}}$ valeur prédite par le modèle. La RMSE la plus faible correspond au modèle le plus réaliste.

CHAPITRE 2

Vers le développement d'une climatologie 3D des communautés phytoplanctoniques pour l'océan mondial

Sommaire

2.1 Introduction 35

2.2 Calibration de la fluorescence en concentration de chlorophylle *a* totale et détermination de la composition des communautés phytoplanctoniques associées 40

2.2.1 Résumé de l'étude 40

2.2.2 Article publié dans *Journal of Geophysical Research* : Retrieving the vertical distribution of chlorophyll *a* concentration and phytoplankton community composition from *in situ* fluorescence profiles : A method based on a neural network with potential for global-scale applications 41

2.2.3 Conclusion et perspectives de l'étude 62

2.3 Création d'une base de données de la distribution verticale de la composition des communautés phytoplanctoniques pour l'océan global 62

2.3.1 Résumé de l'étude 62

2.3.2 Article publié dans *Earth System Science Data* : Vertical distribution of chlorophyll *a* concentration and phytoplankton community composition from *in situ* fluorescence profiles : a first database for the global ocean 64

2.3.3 Résultats complémentaires 78

2.3.4 Exemple d'utilisation de la base de données : une étude spécifique à la Méditerranée 80

2.3.5 Conclusion et perspectives de l'étude 82

2.4 Conclusion générale 83

2.1 Introduction

Une **climatologie** se définit par l'étude de la succession des conditions d'un paramètre donné sur de longues périodes de temps. Les climatologies permettent donc de décrire la distribution d'un paramètre **en moyenne sur une période relativement longue** en fonction d'une certaine **échelle de temps** (e.g. climatologies mensuelles, saisonnières, interannuelles, annuelles, etc.). Par exemple, les climatologies mensuelles et saisonnières résulteront de la moyenne de toutes les données (donc provenant de toutes les années disponibles) respectivement par mois ou par saisons. Une climatologie annuelle intègre absolument toutes les données accessibles de la base de données tandis qu'une climatologie interannuelle est calculée en différenciant les données provenant de chaque année.

Une climatologie peut avoir plusieurs **dimensions** : 2, 3 ou 4 dimensions (i.e. 2D, 3D ou 4D). La climatologie caractérisé par 2 dimensions est moyenné en fonction d'une grille x, y. En océanographie, une climatologie en 2D correspond donc aux valeurs moyennes d'un paramètre pour une certaine couche de l'océan (e.g. climatologies en surface, à 100m ou climatologie d'un paramètre intégré de la surface à 200m). Une climatologie en 3D intègre en plus de x, y la profondeur (i.e. z). Enfin, une climatologie en 4D intègre le temps (i.e. t) en plus des dimensions spatiales x, y, z.

Les climatologies représentent un outil indispensable utilisé par la communauté océanographique dans le but de **résumer** une masse d'informations relatives à un paramètre spécifique et ainsi mettre en évidence plus facilement des **tendances moyennes**. En effet, elles peuvent servir de référence en comparaison avec des « images » plus précises du système et ainsi permettre de détecter certaines tendances. Les climatologies sont également très utiles pour le **contrôle qualité** des données *in situ* car elles apportent une information sur les **gammes de valeurs** attendues pour une région et une période donnée. De plus, les climatologies fournissent des valeurs moyennes permettant de remplacer des données manquantes dans les modèles. En effet, elles sont largement utilisées en **modélisation** pour initialiser ou valider les données d'entrée et de sortie de certains modèles biogéochimiques (e.g. Aumont *et al.*, 2003). Il a également été montré que les climatologies étaient très utiles au développement des modèles basés sur l'assimilation de données (Brasseur *et al.*, 2009).

La principale climatologie de $[Chl]$ à l'échelle globale actuellement utilisée par la communauté océanographique est celle publiée dans le **World Ocean Atlas** 2001 (WOA, Conkright *et al.*, 2002).

Cette climatologie relativement ancienne et non réactualisée présente d'importantes **limitations**. En effet, les climatologies de la $[Chl]$ du WOA disponibles sont soit 1) des climatologies 2D de **surface** saisonnières (le temps est représenté mais pas la verticale) ou bien des climatologies annuelles en 3D mais relativement **mal résolues sur la verticale** (la verticale est intégrée mais pas le temps). Malheureusement, la résolution verticale de ces climatologies annuelles, basée sur des **profondeurs standards** (i.e. 7 profondeurs disponibles de 0 à 100 m pour toute la colonne d'eau), ne permet pas d'estimer de manière fine la forme du profil vertical de $[Chl]$, ni de modéliser certaines couches dans la colonne d'eau où l'on sait qu'il subsiste des quantités significatives de biomasse chlorophyllienne (e.g. « Deep Chlorophyll Maximum », DCM, des grands gyres sub-tropicaux). Cette climatologie souffre également d'un manque important de données *in situ*, notamment en profondeur, ce qui a conduit à une **forte interpolation** des données dans l'espace (voir l'exemple de la distribution spatiale des données du WOA pour 4 des profondeurs standards sur la Figure 2.1). Ce manque de données peut s'expliquer par le coût des mesures de $[Chl]$ utilisées pour établir cette climatologie. En effet, la climatologie du WOA est basée sur des mesures dites "bouteilles" qui ne peuvent être obtenues que lors des campagnes océanographiques et qui permettent d'obtenir seulement des profils verticaux discrets de $[Chl]$. Ces données "bouteilles" sont issues d'échantillons d'eau de mer collectés à différentes profondeurs, ensuite filtrés avant l'analyse de leur pigment préalablement extrait par un solvant (voir les détails dans la Sect. 1.2.2 du chapitre 1). Les analyses reposent sur des techniques variées que ce soit la fluorimétrie *in vitro* (Holm-Hansen *et al.*, 1965), la spectrophotométrie (Lorenzen, 1967) ou la chromatographie (Mantoura et Llewellyn, 1983). Les mesures par chromatographie liquide à haute performance (**HPLC** pour « High Performance Liquid Chromatography ») sont désormais utilisées comme données de **référence** pour estimer la $[Chl]$ (Hooker *et al.*, 2009, Mantoura et Llewellyn, 1983, Peloquin *et al.*, 2013, Claustre *et al.*, 2004).

Face aux limites des climatologies actuelles de $[Chl]$, une base de données qui pourra être utilisée à terme pour créer une **nouvelle climatologie en 3D** et à terme en 4D de la $[Chl]$ est présentée et détaillée dans ce chapitre. Cette base de données globale est constituée de **profils de fluorescence** *in vivo* de la chlorophylle *a* (notée fluorescence par la suite) en lieu et place des données « bouteilles » sur lesquelles la climatologie du WOA est basée.

La mesure de la $[Chl]$ par fluorescence *in vivo* (Lorenzen, 1966) repose sur la propriété de fluores-

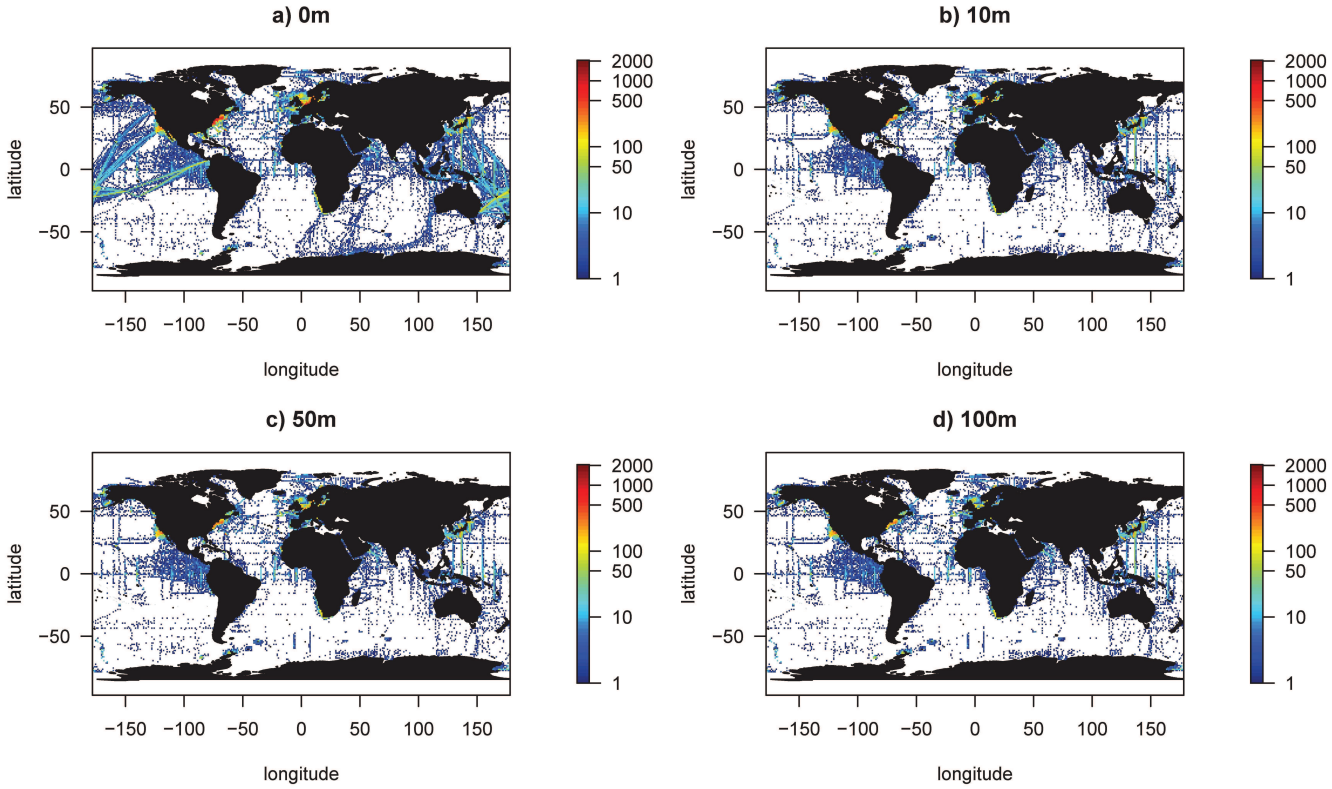


FIGURE 2.1 – *Distribution des données « bouteilles » sur lesquelles reposent les climatologies du World Ocean Atlas pour 4 des 7 profondeurs standards à disposition (Conkright* et al.*, 2002) : a) données de surface, b) données à 10 m, c) données à 50 m et d) données à 100 m.*

cence de la molécule de chlorophylle *a* (voir détails dans la Sect. 1.2.3 du chapitre 1). Le capteur de fluorescence (i.e. fluorimètre) effectue des mesures **non intrusives** à haute fréquence, permettant ainsi l'acquisition d'un profil **quasi-continu sur la verticale** (i.e. environ une mesure tous les mètres). Ces capteurs peuvent être couplés à une sonde CTD sur une rosette ou sur des plateformes autonomes. Grâce à la relative **facilité d'acquisition** de la fluorescence, le nombre de profils à disposition a déjà augmenté significativement depuis quelques années (et de plus en plus de mesures sont attendues à l'avenir).

Pour résumer, la fluorescence est une méthode à haute résolution, rapide et peu coûteuse permettant d'acquérir de manière quasi-continue la $[Chl]$ sur la dimension verticale. Ce sont ces différents avantages qui font de la fluorescence la méthode la plus utilisée pour estimer la $[Chl]$ et certainement l'un des paramètres biogéochimiques les plus mesurés dans l'océan ouvert. Cependant, malgré ces nombreux avantages, les profils de fluorescence n'ont toujours pas été intégrés dans les analyses globales (e.g. création d'une climatologie 3D de la $[Chl]$ à partir de profils de fluorescence). En effet, le principal inconvénient de la fluorescence est que le rapport **fluorescence/[Chl]** n'est pas toujours

constant (voir détails dans la Sect 1.2.3), d'où l'importance cruciale d'associer ces mesures à des **calibrations**.

Cette calibration peut tout d'abord être effectuée localement lorsqu'un fluorimètre est utilisé à bord d'un navire océanographique (associé à la sonde CTD) et si des prélèvements discrets sont réalisés **simultanément** puis analysés ultérieurement (e.g. HPLC). Les régressions linéaires entre la fluorescence et la $[Chl]$ estimées permettent alors de calibrer le fluorimètre (Morel et Maritorena, 2001). Cependant, cette calibration ne peut être appliquée aux données de fluorescence acquises par les **plateformes autonomes** puisque les données HPLC concomitantes aux profils de fluorescence ne sont pas disponibles mis à part éventuellement pour le premier profil associé au déploiement de la plateforme. De plus, il n'y a *a priori* aucun moyen de récupérer le capteur pour une « post-calibration » afin d'évaluer une éventuelle dérive du capteur de fluorescence par rapport à sa première mesure. Ces dernières années, des procédures **alternatives** de calibration ont donc été proposées.

La première méthode de calibration développée dans le but de s'affranchir des mesures HPLC requiert la mesure simultanée de la fluorescence et de **l'irradiance descendante** à $490nm$. La combinaison de ces deux mesures, associée à l'utilisation d'un **modèle bio-optique** permet ensuite de dériver le facteur de calibration de la fluorescence en $[Chl]$ (Xing *et al.*, 2011). Une seconde méthode utilise la $[Chl]$ de surface estimée par **satellite** afin d'en déduire plusieurs paramètres permettant de calibrer le profil de fluorescence (Lavigne *et al.*, 2012). Ces deux méthodes de calibration ont l'avantage d'être « forcées » par une mesure additionnelle (e.g. irradiance et mesure satellite). Cependant, la majorité des profils de fluorescence n'ont pas de **mesures additionnelles** accessibles. En effet, la première méthode nécessite l'acquisition systématique du profil d'irradiance associé à la fluorescence, or l'irradiance ne peut pas toujours être mesurée (e.g. profil de fluorescence acquis pendant la nuit). Les principaux inconvénients de la seconde méthode de calibration sont qu'elle n'est applicable qu'à partir du lancement des premiers satellites « couleur de l'eau » (i.e. CZCS puis SeaWiFS, voir Sect. 1.2.1 du chapitre 1) et qu'elle nécessite un temps clair afin de récupérer les données satellites associées au profil (i.e. pas de matchup s'il y a une couverture nuageuse ou pendant la nuit).

Une autre méthode de calibration tente d'extraire les **caractéristiques de la forme du profil de fluorescence** afin d'estimer la $[Chl]$ (Mignot *et al.*, 2011). Cette calibration a pour inconvénient de n'être applicable que sur les deux formes caractéristiques des profils de $[Chl]$ (i.e. stratifiée ou

mélangée).

Ces méthodes ont chacune leur avantage et leur domaine d'application mais face à leurs différentes limites une nouvelle méthode de calibration **globale**, **indépendante de toutes mesures additionnelles** et applicable dans n'importe quelles conditions nous semblait nécessaire. Nous avons donc développé une méthode de calibration basée sur la **seule forme du profil de fluorescence** (Sauzède *et al.*, 2015a). Cette méthode a pour but de **réconcilier et harmoniser** les données les plus « anciennes » de fluorescence *in vivo* (acquises depuis 1958 à partir des campagnes océanographiques) avec les nouvelles données acquises par les plateformes autonomes dans le but de pouvoir les utiliser dans leur globalité. De plus, cette méthode permet non seulement de calibrer la fluorescence en $[Chl]$ mais aussi d'en dériver **la composition des communautés phytoplanctoniques**.

Comme souligné dans la Sect. 1.1.2 du chapitre 1, la nature des communautés phytoplanctoniques est une variable indispensable à prendre en compte dans les études concernant l'impact du changement environnemental sur les **flux de carbone** (Le Quere *et al.*, 2005). En effet, la **taille du phytoplancton** influence des processus clés de la physiologie du phytoplancton (Geider *et al.*, 1986, Platt et Denman, 1977, 1978, Sunda et Huntsman, 1997) et impacte donc les cycles biogéochimiques du carbone. Plusieurs modèles biogéochimiques sont d'ailleurs basés sur le fractionnement de la communauté phytoplanctonique en plusieurs classes de taille (e.g. Aumont *et al.*, 2003, Blackford *et al.*, 2004, Kishi *et al.*, 2007, Marinov *et al.*, 2010, Ward *et al.*, 2012). En obtenant les **indices de taille** phytoplanctonique à partir de données acquises à haute fréquence (i.e. fluorescence), des bases de données contenant une masse d'information précieuse pourront être assemblées. Ainsi, les climatologies en 3D de la composition des communautés phytoplanctoniques en association à la $[Chl]$ obtenue à partir des profils de fluorescence via la méthode développée, pourraient notamment s'avérer relativement utile à l'initialisation et la validation de ces modèles biogéochimiques.

Dans un premier temps, la méthode de calibration spécifiquement développée dans le but **d'harmoniser** des profils de fluorescence de **sources diverses et hétérogènes** est présentée. Par la suite, nous appliquons cette méthode de calibration à une **base de données dite globale**. Celle-ci est constituée par l'ensemble des profils de fluorescence accessibles, notamment dans les bases de données en ligne. Nous avons regroupé et qualifié tous ces profils avant de les calibrer en utilisant notre méthode. Dans la seconde partie de ce chapitre, les caractéristiques biogéochimiques (i.e. distribution spatiale

de la concentration en chlorophylle a et des communautés phytoplanctoniques associées) de cette base de données sont présentées puis discutées.

2.2 Calibration de la fluorescence en concentration de chlorophylle *a* totale et détermination de la composition des communautés phytoplanctoniques associées

2.2.1 Résumé de l'étude

Une méthode basée sur une **approche neuronale** a été développée dans le but de déterminer la distribution verticale de (1) la **concentration en chlorophylle *a*** ($[Chl]$) et (2) la **composition taxonomique des grands groupes phytoplanctoniques** classés sur la base de leur taille (i.e. microphytoplancton, nanophytoplancton et picophytoplancton). Cette méthode (**FLAVOR** pour "Fluorescence to Algal communities Vertical distribution in the Oceanic Realm") utilise comme principale donnée d'entrée la **forme du profil de fluorescence** à laquelle sont associées les informations concernant la date d'acquisition ainsi que la géolocalisation du profil. Le choix d'avoir ainsi limité les entrées du réseau de neurones est essentiellement guidé par l'objectif final qui est de pouvoir **harmoniser**, sur des bases quantitatives (concentration en chlorophylle *a*), les bases de données les plus anciennes (disposant de très peu de données ancillaires) avec les bases de données plus récentes, et bien évidemment celles susceptibles d'être constituées à l'avenir.

Le réseau de neurones est entrainé et validé avec une base de données comprenant 896 profils verticaux *in situ* de **pigments** déterminés par chromatographie liquide à haute performance (HPLC) concomitant avec le profil de fluorescence. Ces profils ont été collectés dans le cadre de 22 campagnes océanographiques de 1991 à 2012 et sont **représentatifs de la diversité de profils** potentiellement observables dans l'océan mondial en termes de conditions hydrologiques, biogéochimiques et trophiques associées. Il est donc attendu que la méthode ici développée ait un **domaine d'application** couvrant la plupart des eaux océaniques.

FLAVOR est validée grâce aux données de $[Chl]$ et des indices de taille des communautés phytoplanctoniques associés provenant d'un **jeu de données indépendant** (voir les Figures 5, 7, 8 et 9 du

papier ci-après). Ce jeu de données correspond à 20% des profils de la base de données initiale choisis aléatoirement dans plusieurs catégories trophiques, ce qui permet de représenter les différentes conditions biogéochimiques pouvant être rencontrées dans l'océan et de s'assurer du fait que la méthode est **applicable dans la plupart des eaux océaniques**. La méthode apparaît comme étant relativement robuste spatialement et temporellement. FLAVOR restitue la $[Chl]$ avec une erreur médiane de 40% et la $[Chl]$ associée au micro-, nano- et pico-phytoplancton avec respectivement 46%, 35% et 40% d'erreur (voir Table 4 du papier ci-dessous).

Pour illustrer le **potentiel** de cette méthode, nous l'avons ensuite appliquée à des mesures *in situ* provenant de la station BATS ("Bermuda Atlantic Time Series Study") dans le but de générer des **climatologies mensuelles** de $[Chl]$ et des indices de taille phytoplanctoniques associés. Les climatologies résultantes sont très prometteuses comparées aux climatologies basées sur les données *in situ* HPLC disponibles (voir les Figures 10 et 11 du papier ci-après). En effet, l'évolution mensuelle à la fois de la distribution verticale et des valeurs quantitatives de $[Chl]$ est bien reproduite grâce à la méthode proposée. De plus, cette étude indépendante permet de confirmer que la base de données d'entrainement du réseau de neurones, qui n'inclue pas la zone de BATS, est représentative des situations biogéochimiques potentiellement observables dans l'océan mondial.

2.2.2 Article publié dans *Journal of Geophysical Research* : Retrieving the vertical distribution of chlorophyll *a* concentration and phytoplankton community composition from *in situ* fluorescence profiles : A method based on a neural network with potential for global-scale applications

AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Oceans

RESEARCH ARTICLE

10.1002/2014JC010355

Key Points:
- Chlorophyll *a* concentration is retrieved from fluorescence profiles
- Phytoplankton communities size indices are retrieved from fluorescence profiles
- A neural network is developed with a potential for global-scale applications

Supporting Information:
- Readme
- Figure S1
- Figure S2
- Figure S3
- Figure S4
- Figure S5
- Figure S6
- Figure S7
- R and Matlab codes of the proposed algorithm

Correspondence to:
R. Sauzède,
sauzede@obs-vlfr.fr

Citation:
Sauzède, R., H. Claustre, C. Jamet, J. Uitz, J. Ras, A. Mignot, and F. D'Ortenzio (2015), Retrieving the vertical distribution of chlorophyll a concentration and phytoplankton community composition from in situ fluorescence profiles: A method based on a neural network with potential for global-scale applications, *J. Geophys. Res. Oceans*, *120*, doi:10.1002/2014JC010355.

Received 30 JUL 2014
Accepted 17 DEC 2014
Accepted article online 5 JAN 2015

Retrieving the vertical distribution of chlorophyll a concentration and phytoplankton community composition from in situ fluorescence profiles: A method based on a neural network with potential for global-scale applications

R. Sauzède[1,2], H. Claustre[1,2], C. Jamet[3], J. Uitz[1,2], J. Ras[1,2], A. Mignot[4], and F. D'Ortenzio[1,2]

[1]Laboratoire d'Océanographie de Villefranche, CNRS, UMR7093, Villefranche-Sur-Mer, France, [2]Université Pierre et Marie Curie-Paris 6, UMR7093, Laboratoire d'océanographie de Villefranche, Villefranche-Sur-Mer, France, [3]Laboratoire d'Océanologie et de Géosciences, UMR8187, ULCO/CNRS, Wimereux, France, [4]Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

Abstract A neural network-based method is developed to assess the vertical distribution of (1) chlorophyll *a* concentration ([Chl]) and (2) phytoplankton community size indices (i.e., microphytoplankton, nanophytoplankton, and picophytoplankton) from in situ vertical profiles of chlorophyll fluorescence. This method (FLAVOR for Fluorescence to Algal communities Vertical distribution in the Oceanic Realm) uses as input only the shape of the fluorescence profile associated with its acquisition date and geo-location. The neural network is trained and validated using a large database including 896 concomitant in situ vertical profiles of High-Performance Liquid Chromatography (HPLC) pigments and fluorescence. These profiles were collected during 22 oceanographic cruises representative of the global ocean in terms of trophic and oceanographic conditions, making our method applicable to most oceanic waters. FLAVOR is validated with respect to the retrieval of both [Chl] and phytoplankton size indices using an independent in situ data set and appears to be relatively robust spatially and temporally. To illustrate the potential of the method, we applied it to in situ measurements of the BATS (Bermuda Atlantic Time Series Study) site and produce monthly climatologies of [Chl] and associated phytoplankton size indices. The resulting climatologies appear very promising compared to climatologies based on available in situ HPLC data. With the increasing availability of spatially and temporally well-resolved data sets of chlorophyll fluorescence, one possible global-scale application of FLAVOR could be to develop 3-D and even 4-D climatologies of [Chl] and associated composition of phytoplankton communities. The Matlab and R codes of the proposed algorithm are provided as supporting information.

1. Introduction

Phytoplankton is an essential component in marine biogeochemical cycles and ecosystems. The assessment of its spatio-temporal distribution and variability across the global ocean is thus an important objective in biological oceanography. Methods have been developed for monitoring oceanic phytoplankton distribution (horizontally and vertically) with the objective to continuously increase the spatio-temporal resolution of data acquisition. These methods often rely on the estimation of the chlorophyll *a* concentration which is a universal proxy for phytoplankton biomass. Remote sensing of Ocean Color Radiometry (OCR) offers a unique way to map quasi-synoptically chlorophyll *a* concentration at the ocean surface. Through this technique, a wide range of applications has been developed, leading to a better understanding of phytoplankton dynamics in the upper ocean [*McClain*, 2009; *Siegel et al*., 2013]. However, the use of remote sensing of OCR provides chlorophyll *a* concentration only for the surface layers of the ocean [*Gordon and McCluney*, 1975], representing one-fifth of the so-called euphotic layer where phytoplankton photosynthesis takes place [*Morel and Berthon*, 1989]. The vertical distribution of phytoplankton thus escapes from this remote detection. While in situ vertical profiles of chlorophyll *a* concentration are determined with the best accuracy by high-performance liquid chromatography (HPLC) [*Claustre et al*., 2004; *Peloquin et al*., 2013], this method is not compatible with highly repetitive measurements.

Introduced by *Lorenzen* [1966], measurement of in vivo fluorescence of chlorophyll *a* is a nonintrusive technique that allows the direct in situ assessment of chlorophyll *a* concentration. Besides the dissolved oxygen,

© 2015. American Geophysical Union. All Rights Reserved.

chlorophyll *a* fluorescence is certainly the most measured biogeochemical property in the global ocean. The in vivo fluorescence of chlorophyll *a* can be considered as a proxy for chlorophyll *a* concentration with some unavoidable limitations such as the variability of the fluorescence-to-chlorophyll *a* concentration ratio as a function of physiological constraints or community composition [*Kiefer*, 1973; *Falkowski et al*., 1985; *Cunningham*, 1996]. While in vivo fluorescence is an imperfect proxy, it presents the advantage that it can be easily measured thanks to miniature in situ sensors. Thus, the inherent weaknesses of fluorometric measurements are largely compensated by their cost effective acquisition that enables numerous data to be gathered. This is now especially true considering that, besides oceanic cruises, in vivo fluorescence is accessible through autonomous platforms (floats, gliders, animals), allowing a global fluorescence database to be progressively assembled [*Claustre et al*., 2010a, 2010b].

On a case-by-case basis (e.g., an oceanographic cruise), a fluorescence database can be easily and accurately converted into a chlorophyll *a* concentration database, for example, thanks to simultaneous HPLC determination [*Claustre et al*., 1999]. However, with the goal of developing large-scale chlorophyll *a* concentration databases from the merging of different fluorescence data from diverse origins (e.g., sensors, platforms), their consistency and interoperability will become a critical issue. First, the expected exponential growth of fluorescence profile acquisition in the near future (more and more autonomous platforms will be deployed) will stimulate regular updates of these "super" databases. Hence, methods have to systematically add new data in a way that does not compromise the reliability and the quality of the existing database. Second, as some platforms (e.g., floats, animals) are not necessarily recovered, the initial sensor calibration, if any, would represent the only reference for the whole acquisition period (sometimes extending over several years). This can represent a bias because these platforms acquire fluorescence profiles in conditions where the phytoplankton communities and their physiological state may have drastically changed in comparison to the conditions prevailing at the time of platform deployment and initial sensor calibration. Robust methods that allow the retrieval of chlorophyll *a* concentration from fluorescence measurements without any regular or simultaneous HPLC measurement will need to be developed.

Several alternative methods have already been proposed for calibrating fluorescence profiles, which partly circumvent the above mentioned issues. *Lavigne et al*. [2012] use the satellite remotely detected surface chlorophyll *a* concentration as a way to scale the whole fluorescence vertical profile to this reference surface value. While this method presents the advantage of allowing the interoperability of data sets from different origins and locations, it is obviously not applicable to situations where no concurrent satellite data are available (SeaWiFS was launched in 1997). Furthermore, it implicitly postulates that the satellite-derived surface chlorophyll *a* concentration is the "accurate" reference, an assumption that can be challenged, especially in some oceanic regions [e.g., *Bricaud et al*., 2002; *Johnson et al*., 2013]. *Mignot et al*. [2011] have shown that information related to chlorophyll *a* concentration is embedded in the shape of the fluorescence profile. For example, it is intuitive that the deeper the Deep Chlorophyll maximum (DCM), the lower the surface chlorophyll *a* concentration [see also, *Morel and Bertrhon*, 1989; *Uitz et al*., 2006]. *Mignot et al*. [2011] thus proposed a calibration of the fluorescence profile based on its shape. The prerequisite of the method is that the profile shape has to be a priori categorized either into a stratified type (modeled by a Gaussian) or a mixed type (modeled by a sigmoid). Clearly, the advantage of this method is that it does not require any complementary information. Its limit lies in the need for an initial classification of the fluorescence profiles, which may be complicated for those that do not clearly belong to one of the two categories.

The primary objective of the present study relies on, and further extends, the approach of *Mignot et al*. [2011]. It aims at developing a self-consistent calibration method of the fluorescence profile essentially based on its shape, and which requires minimum additional information or a priori knowledge. The choice of limiting the use of additional information is essentially guided by the long-term objective of this study, which is to reconcile the oldest databases (assembled during the 1970s when the first fluorescence profiles were recorded with no or scarce ancillary data) with the most recent as well as future databases (derived from autonomous platforms). Besides the fluorescence profile, its geo-location represents robust additional information that is systematically present (as metadata) and potentially useful. At first order (e.g., on a global scale), the geo-location indeed intrinsically embeds information relative to the trophic status (amount of biomass, e.g., spring bloom in the North Atlantic) or hydrographic conditions (stratified versus mixed, e.g., subtropical gyres versus upwellings).

© 2015. American Geophysical Union. All Rights Reserved.

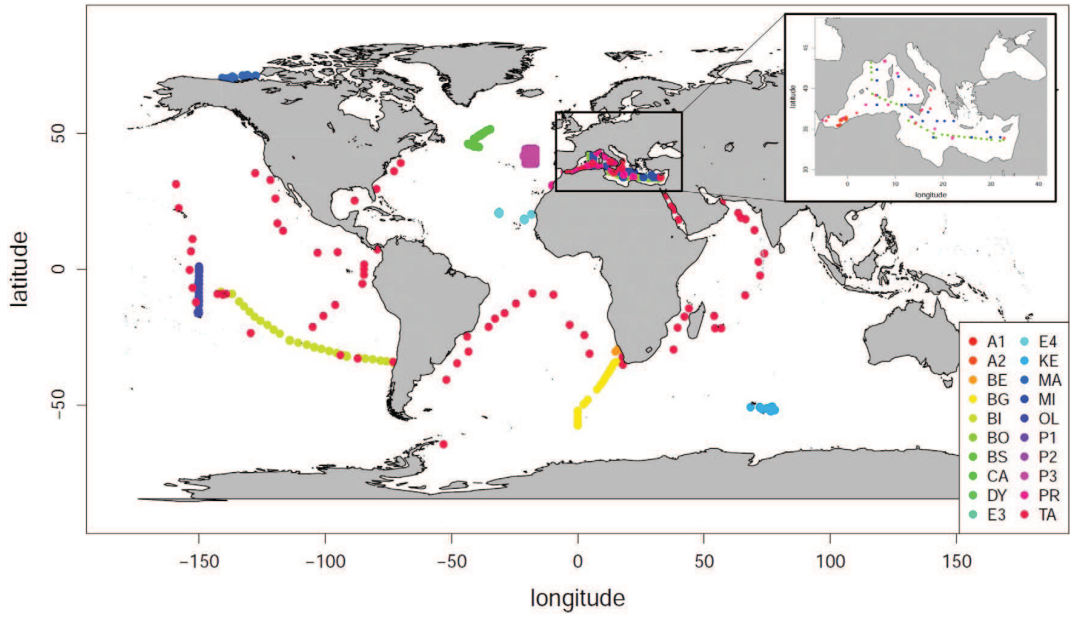


Figure 1. Geographic distribution of the 896 stations used in the present study. For these stations, sampling for HPLC pigment was simultaneous to the acquisition of the fluorescence profile.

Artificial neural networks (ANNs) are approximate functions of any data sets [*Marzban*, 2009]. They represent powerful methods to develop models especially when the underlying relationships are unknown, which is typically the case here with the fluorescence profile shape and the chlorophyll *a* concentration. In particular, multilayered perceptrons (MLPs) are universal approximators of any differentiable and continuous function [*Hornik et al.*, 1989] which are well adapted to ecological data sets having often nonlinear spatially and temporally complex and noisy distributions [*Lek and Guégan*, 1999]. In oceanography, such methods have been developed and used for the retrieval of various products such as the diffuse attenuation coefficient [*Jamet et al.*, 2012], the partial pressure of carbon dioxide (pCO_2) surface distribution [*Friedrich and Oschlies*, 2009; *Telszewski et al.*, 2009], the surface phytoplankton pigment concentration [*Gross et al.*, 2000] or the surface phytoplankton functional types [*Bricaud et al.*, 2007; *Raitsos et al.*, 2008; *Ben Mustapha et al.*, 2013; *Palacz et al.*, 2013]. These methods thus appear well adapted to the type of problem identified here, i.e., the retrieval of a calibrated chlorophyll *a* concentration profile from the shape and geo-location of a fluorescence profile.

In most studies dealing with the possible impact of environmental changes on oceanic carbon fluxes, the nature of phytoplankton communities (e.g., performing regenerated versus new production, small versus large phytoplankton) is an essential variable to account for [e.g., *Le Quere et al.*, 2005]. Ongoing efforts are underway to synthesize the historical knowledge on phytoplankton taxonomy in gathered databases [e.g., *Buitenhuis et al.*, 2012; *Leblanc et al.*, 2012]. These data sets, nevertheless, remain rather sparse and the possibility to regularly improve and update them is weak simply because there are less taxonomic experts than before. Alternative procedures thus need to be developed. *Uitz et al.* [2006] have shown that phytoplankton size indices can be derived from High-Performance Liquid Chromatography (HPLC) pigment analysis [see also *Peloquin et al.*, 2013]. These indices can be related to surface chlorophyll *a* concentration, so that explicit relationships can be observed between the phytoplankton biomass (e.g., chlorophyll *a* concentration), the phytoplankton communities (e.g., phytoplankton size indices) and their vertical distribution.

In summary, the twofold objective of the present study is to retrieve, from the sole knowledge of the fluorescence profile shape and geo-location, the vertical profile of (1) chlorophyll *a* concentration and (2) phytoplankton community size indices. To address this objective, we developed a MLP-based method, hereafter referred to as FLAVOR for Fluorescence to Algal communities Vertical distribution in the Oceanic Realm. The MLP is trained and validated using an in situ database that contains 896 continuous vertical fluorescence

© 2015. American Geophysical Union. All Rights Reserved.

Table 1. Abbreviations Used in the Present Study and Their Significance

Abbreviations	Significance
[TChl]	Chlorophyll *a* concentration associated to the total phytoplankton biomass (mg m^{-3})
fmicro	Fraction of chlorophyll *a* associated to microphytoplankton
fnano	Fraction of chlorophyll *a* associated to nanophytoplankton
fpico	Fraction of chlorophyll *a* associated to picophytoplankton
[microChl]	Total chlorophyll *a* concentration associated to microphytoplankton (mg m^{-3})
[nanoChl]	Total chlorophyll *a* concentration associated to nanophytoplankton (mg m^{-3})
[picoChl]	Total chlorophyll *a* concentration associated to picophytoplankton (mg m^{-3})
[Chl]	Chlorophyll *a* concentration (mg m^{-3}) referring either to [TChl] or [microChl], [nanoChl], and [picoChl]
$[TChl]_{Z_0}$	[TChl] integrated from the surface up to the depth Z_0 (mg m^{-2})
z	Geometrical depth (m)
Z_0	Depth at which the fluorescence begins to be constant with depth (m)
Z_e	Euphotic layer depth (m)
Z_m	Mixed layer depth (m)
ζ	Depth normalized with respect to Z_0, $\zeta = z/Z_0$, dimensionless
Fuco	Fucoxanthin (mg m^{-3})
Perid	Peridinin (mg m^{-3})
Hex-fuco	19′-Hexanoyloxyfucoxanthin (mg m^{-3})
But-fuco	19′-Butanoyloxyfucoxanthin (mg m^{-3})
Allo	Alloxanthin (mg m^{-3})
TChl*b*	Chlorophyll *b* + divinyl Chlorophyll *b* (mg m^{-3})
Zea	Zeaxanthin (mg m^{-3})
Lon	Longitude (°E)
Day	Day of the year
Lon_{rad}	Longitude transformed in radians
Day_{rad}	Day transformed in radians
fluo	The fluorescence (relative units)
$fluo_{norm}$	The normed fluorescence (dimensionless)
$[Chl]_{HPLC}$	The [Chl] values of reference estimated by HPLC (mg m^{-3})
$[Chl]_{MLP}$	The 10 discrete values of [Chl] returned by the MLP (mg m^{-3})
$[Chl]_{cal}$	The fluorescence profile calibrated in [Chl] (mg m^{-3})
α	The calibration coefficient, such as $[TChl]_{cal} = \alpha \, . \, fluo$
MAPD	Median absolute percent difference (%)
a	The slope of the linear regression between $[Chl]_{MLP}$ or $[Chl]_{cal}$ and the reference values of [Chl] estimated by HPLC

profiles acquired simultaneously to HPLC pigment determinations at selected depths. This database is representative of the range of trophic and oceanographic conditions prevailing in the global open ocean. Therefore, we expect that any potential conclusion of the present study will be of general applicability to the global ocean.

2. Data Presentation and Processing

2.1. Database of Vertical Profiles of Chlorophyll Fluorescence and HPLC Pigments

The present study makes use of an extensive database of concurrent vertical profiles of chlorophyll fluorescence (fluo, see abbreviations in Table 1) and phytoplankton pigments determined by HPLC. This data set is an extension of the one used by *Mignot et al.* [2011]. The data were collected at 896 stations sampled during 22 open ocean cruises between 1991 and 2012 (Table 2), which took place in a wide variety of oceanic regions. Most of the data were collected in the Mediterranean Sea (39%) and the Atlantic Ocean (36%), 17% of the data are from the Pacific Ocean, 3% from the South Ocean, 3% from the Arctic Ocean, and 2% from the Indian Ocean (Figure 1).

The chlorophyll *a* in vivo fluorescence profiles acquired during the 22 cruises were obtained using a fluorometer mounted on a CTD-rosette. We applied a quality control to each fluorescence profile similarly to *D'Ortenzio et al.* [2010] in order to remove aberrant data caused by electronic noise.

Samples for HPLC pigment determinations were acquired at discrete depths (approximately 10 data points per profile) and analyzed according to the method described by *Claustre* [1994] for the EUMELI cruises, *Vidussi et al.* [1996] for the cruises that occurred prior to 2004, and *Ras et al.* [2008] for all other cruises after 2004. The concentration of total chlorophyll *a* as determined by HPLC actually refers to the concentration of the so-called total chlorophyll *a* ([TChl]) which is the sum of the concentrations of monovinyl-chlorophyll *a*, divinyl-chlorophyll *a*, chlorophyllide *a* and the allomeric and epimeric forms of chlorophyll *a*. [TChl] is an estimate of the total phytoplankton biomass. In addition to chlorophyll *a*, the HPLC technique enables

© 2015. American Geophysical Union. All Rights Reserved.

Table 2. Abbreviations, Location, Period, Number of Stations (After Quality Control) and References or PI for Each Cruises/Projects Used in the Data Set

Cruise/Project	Abbreviation	Location	Period	Number of Stations	References or PI
Almofront1	A1	Alboran Sea	Apr–May 1991	18	*Claustre et al.* [1994] and *Peloquin et al.* [2013]
Almofront2	A2	Alboran Sea	Nov 1997–Jan 1998	37	*Claustre et al.* [2000], *Uitz et al.* [2006], and *Peloquin et al.* [2013]
Bencal	BE	Benguela upwelling	Oct 2002	5	*Morel et al.* [2006] and *Peloquin et al.* [2013]
Biosope	BI	South Pacific	Austral Summer 2004	60	*Ras et al.* [2008] and *Peloquin et al.* [2013]
Bonus Good Hope	BG	Southern Ocean	Feb–Mar 2008	23	PI: Claustre and Ras
Boum	BO	Mediterranean Sea	Jun–Jul 2008	28	*Crombet et al.* [2011] and *Peloquin et al.* [2013]
Boussole	BS	7.54°E, 43.22°N	2001–2007	75	PI: Antoine and Ras
CATCH	CA	North Atlantic	Jan–Feb 1997	26	*Uitz et al.* [2006] and *Peloquin et al.* [2013]
Dyfamed	DY	7.54°E, 43.22°N	1994–1999	32	*Marty et al.* [2002] and *Peloquin et al.* [2013]
Eumeli3	E3	Sub-tropical North Atlantic	Sep–Oct 1991	15	*Claustre and Marty* [1995] and *Peloquin et al.* [2013]
Eumeli4	E4	Sub-tropical North Atlantic	May–Jun 1992	22	*Claustre and Marty* [1995] and *Peloquin et al.* [2013]
Keops	KE	Southern Ocean	Jan–Feb 2005	16	*Uitz et al.* [2009]
HUDSON 2013-008	LA	Labrador Sea	May 2013	22	PI: Claustre and Ras
Malina	MA	Arctic Ocean	Jul–Nov 2009	24	*Huot et al.* [2013]
Minos	MI	Mediterranean Sea	May–Jun 1996	67	*Uitz et al.* [2006] and *Peloquin et al.* [2013]
Olipac	OL	Equatorial Pacific	Nov 1994	44	*Claustre et al.* [1999], *Uitz et al.* [2006], and *Peloquin et al.* [2013]
Pomme1	P1	North Atlantic	Feb–Mar 2001	58	*Claustre et al.* [2005], *Uitz et al.* [2006], and *Peloquin et al.* [2013]
Pomme2	P2	North Atlantic	Mar–May 2001	65	*Claustre et al.* [2005], *Uitz et al.* [2006], and *Peloquin et al.* [2013]
Pomme3	P3	North Atlantic	Sep 2001	78	*Claustre et al.* [2005], *Uitz et al.* [2006], and *Peloquin et al.* [2013]
Prosope	PR	Mediterranean Sea	Sep 1999	63	*Claustre et al.* [2004], *Uitz et al.* [2006], and *Peloquin et al.* [2013]
Strasse	ST	Sub-tropical North Atlantic	Aug–Sep 2012	36	PI: Claustre and Ras
Tara	TA	Global Ocean	2009–2012	82	PI: Bricaud, Boss and Ras
Total				896	

measuring the concentration of a suite of accessory pigments that can be used to estimate the composition of phytoplankton communities. Here we utilized the diagnostic pigment-based approach of *Uitz et al.* [2006], based on *Claustre* [1994] and *Vidussi et al.* [2001], to estimate the biomass associated with three pigment-derived size classes, i.e., microphytoplankton, nanophytoplankton, and picophytoplankton. Seven major pigments are selected as being representative of distinct phytoplankton groups: fucoxanthin, peridinin, 19′-hexanoyloxyfucoxanthin, 19′-butanoyloxyfucoxanthin, alloxanthin, chlorophyll *b* + divinyl chlorophyll *b*, and zeaxanthin (abbreviations in Table 1).

The fractions of chlorophyll *a* concentration associated with each of the three phytoplankton classes (f_{micro}, f_{nano}, and f_{pico}) are derived from the following equations [*Uitz et al.*, 2006]:

$$f_{micro} = {}^{(1.41\ [Fuco]+1.41\ [Perid])}\Big/\sum DP \quad (1)$$

$$f_{nano} = {}^{(1.27\ [Hex-fuco]+0.35\ [But-fuco]+0.60[Allo])}\Big/\sum DP \quad (2)$$

© 2015. American Geophysical Union. All Rights Reserved.

$$f_{pico} = {}^{(1.01\ [TChl\mathit{b}]+0.86\ [Zea])} \Big/ \sum DP \tag{3}$$

with $\sum$ DP representing the sum of the seven diagnostic pigments concentrations:

$$\sum DP = 1.41\ [Fuco]+1.41[Perid]+1.27[Hex-fuco]+0.35[But-fuco]+0.60[Allo]+1.01[TChl\mathit{b}]+0.86[Zea] \tag{4}$$

It is then possible to derive the chlorophyll *a* concentration associated with each of the three phytoplankton classes ([microChl], [nanoChl], and [picoChl]) according to the following equations:

$$[microChl] = f_{micro} * [TChl] \tag{5}$$

$$[nanoChl] = f_{nano} * [TChl] \tag{6}$$

$$[picoChl] = f_{pico} * [TChl] \tag{7}$$

The pigment-based approach of *Uitz et al*. [2006], admittedly, has some limitations. For example, some diagnostic pigments are shared by several phytoplankton groups and some groups may cover a broad size range. Nevertheless, it has proven valuable at regional and global scales for apprehending the composition of phytoplankton assemblages, both in terms of taxonomic composition and size structure. Recently, some modifications to the *Uitz et al*. [2006] approach have been proposed [*Brewin et al*., 2010; *Devred et al*., 2011]. Yet, in a comparative study, *Brewin et al*. [2014] show that these modifications do not bring major changes to the pigment-derived size structure of phytoplankton communities. Hence, in the present study, we calculate the [microChl], [nanoChl], and [picoChl] following *Uitz et al*. [2006].

Finally, a quality control was applied to each HPLC-determined vertical pigment profile as described in *Uitz et al*. [2006], i.e., (1) samples with [TChl] lower than 0.001 mg m^{-3} were rejected, (2) the first sample has to be located between the surface and 10 m depth, (3) the last sample has to be taken at a depth greater or equal to the euphotic depth, Ze, defined as the depth at which the irradiance is reduced to 1% of its surface value, and (4) a minimum of four samples per profile is required. For this quality control procedure, Ze is estimated according to the method of *Morel and Berthon* [1989], by using the [TChl] profile derived from HPLC estimates.

The data set used in this study includes concurrent fluorescence vertical profiles and HPLC-determined [TChl], [microChl], [nanoChl], and [picoChl] at discrete depths. This data set is representative of a large variety of hydrological, biogeochemical and associated trophic conditions observed in the open ocean (Figure 2). First, the HPLC-fluorescence database is characteristic of a broad variety of hydrological conditions. For example, it includes measurements acquired in the North Atlantic during winter when the mixed layer depth may reach 700 m (CATCH cruise) as well as measurements from the Subtropical South Pacific Gyre when the average mixed layer depth in spring is ~30 m (BIOSOPE cruise). Second, the database is representative of most trophic conditions (from oligotrophic to eutrophic waters) observed in the global ocean. The parameter $[TChl]_{Z_0}$ defined as the [TChl] integrated between the surface and the depth Z_0 describes these different trophic conditions. Z_0 is set as the depth at which the fluorescence profile returns to a constant background value (see Figures 3a and 3b and section 2.2 for the calculation details). Indeed, the data set appears to be representative of the global ocean as $[TChl]_{Z_0}$ covers 2 order of magnitude. The most oligotrophic conditions were found during the Malina cruise in the Arctic Ocean with a value of about 10 mg m^{-2}; similarly, the $[TChl]_{Z_0}$ minimum value measured during the BIOSOPE cruise is 18 mg m^{-2}. The most eutrophic conditions are also represented with some profiles acquired in the Labrador Sea during the spring bloom with $[TChl]_{Z_0}$ values of about 900 mg m^{-2}. Because our data set is representative of a broad variety of conditions, we expect that the method developed here will be applicable at a global scale in the open ocean.

The data set presented here was split into two subsets. One subset was used for training the MLP; the second data set was used for validating the MLP. These two independent data sets, respectively, including about 80% (717 profiles) and 20% (179 profiles) of the initial data set, were randomly built up from a random selection of stations. As both data sets should be representative of the open

© 2015. American Geophysical Union. All Rights Reserved.

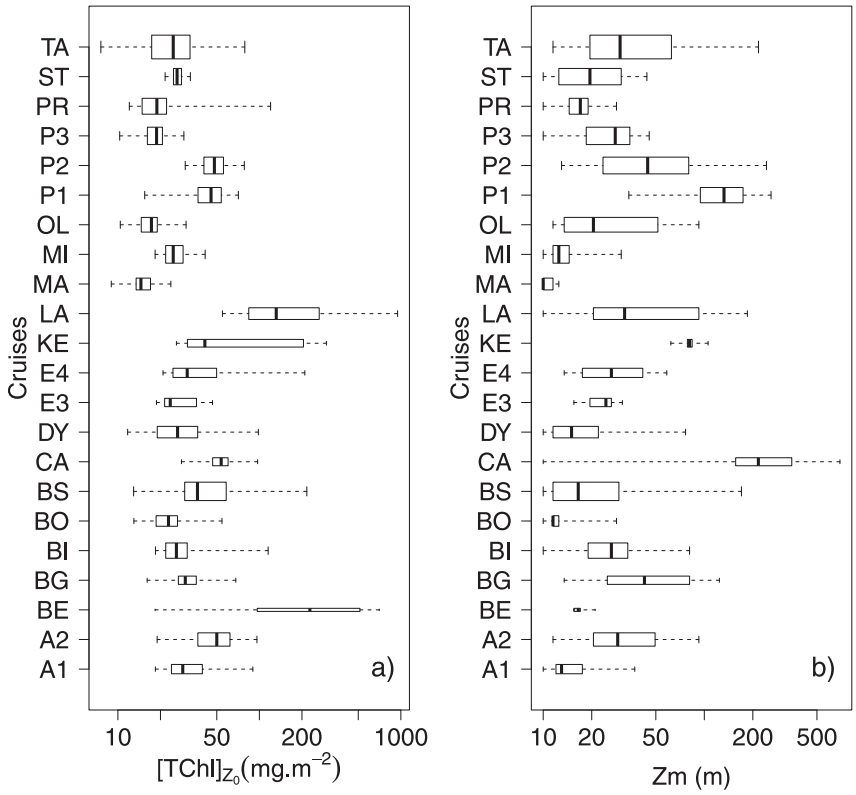


Figure 2. General characteristics of the data set used in this study. Boxplots of (a) the integrated [TChl] from the surface up to the depth Z_0, $[TChl]_{Z0}$ (mg m^{-2}), and (b) the mixed layer depth Z_m (m) for the 22 cruises. Boxplot widths vary depending on the cruise's number of data. The lower and upper whiskers represent the minimum to maximum values; the box represents the upper quartile and the lower quartile with the middle line representing the median of the values.

ocean, the stations were randomly selected from different trophic categories (Table 3), based on the concentration of chlorophyll *a* in the surface layer of the ocean, according to *Uitz et al*. [2006]. The geographic distribution of the sampling stations in the training and validation data sets is shown in Figures S1 and S2, respectively.

2.2. Data Processing

The FLAVOR algorithm has been developed on the basis of two dimensionless quantities. First, it is based on the shape of the fluorescence profile. Hence, all the fluorescence values of the database are normalized within the 0–1 range as follows:

$$\mathrm{fluo}_{\mathrm{norm}} = \frac{\mathrm{fluo} - \min(\mathrm{fluo})}{\max(\mathrm{fluo}) - \min(\mathrm{fluo})} \tag{8}$$

with $\mathrm{fluo}_{\mathrm{norm}}$ the normalized fluorescence values and fluo the fluorescence in relative units. This procedure also enables merging fluorescence profiles from different sources and thus calibrated with different protocols or even uncalibrated.

Second, a dimensionless depth, ζ, is introduced and calculated as the actual depth, z, divided by Z_0:

$$\zeta = z/Z_0 \tag{9}$$

with Z_0 the depth at which the fluorescence profile returns to a constant background value. In order to compute Z_0, the $\mathrm{fluo}_{\mathrm{norm}}$ profile is smoothed using a median filter and the Z_0 depth is calculated as the first depth with a normalized fluorescence value of 0. As shown in Figure 3, the parameter Z_0 displays a large variability among the different fluorescence profiles of our database (17–831 m).

© 2015. American Geophysical Union. All Rights Reserved.

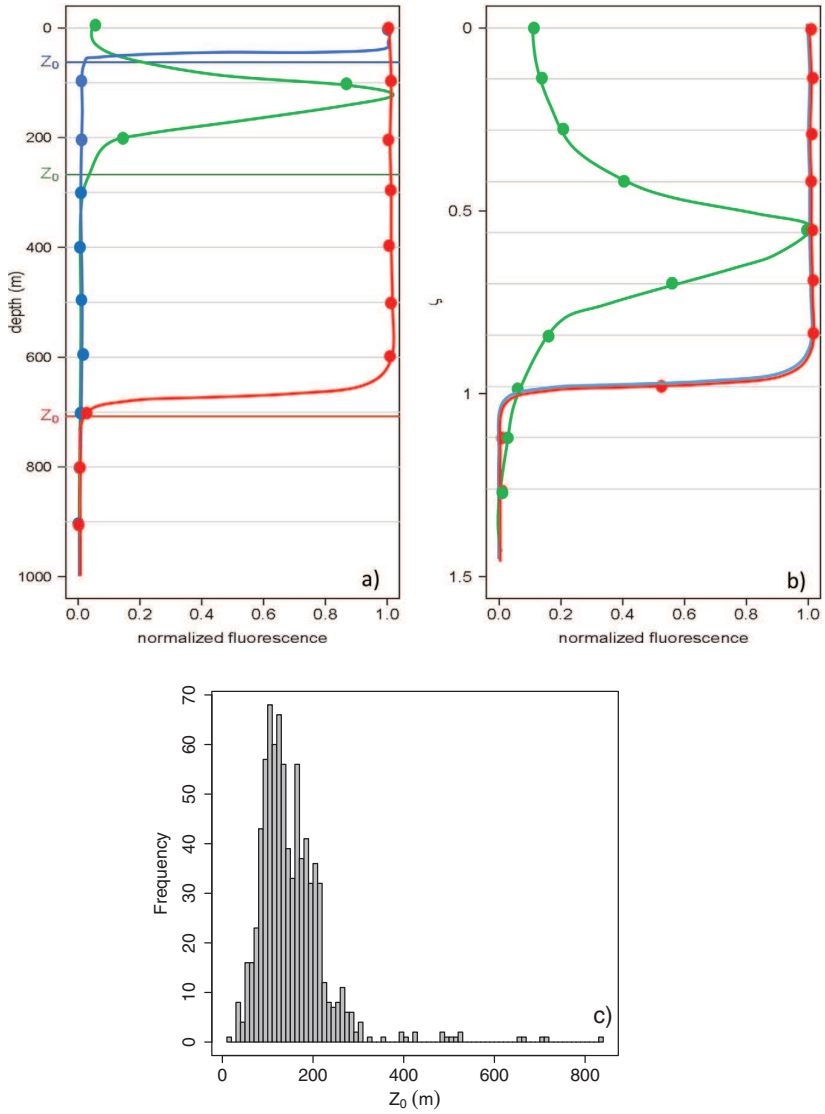


Figure 3. Characteristics of the fluorescence profile shape. (a) Three schematic normalized fluorescence profiles representative of the diversity of the observed situations; red: deeply mixed profile; green: stratified profile; blue: shallow mixed profile. Note that the corresponding Z_0 depths are reported. The lines correspond to the fluorescence profiles while the dots identify the hypothetical inputs of the fluorescence profiles for the MLP (10 inputs). (b) Same as for Figure 3a expect that the geometrical depth is here replaced by the dimensionless depth ζ. (c) Histogram of the Z_0 frequency distribution for the 896 fluorescence profiles of the database.

Scaling the fluorescence profiles with respect to ζ enables to merge all profiles regardless of their vertical shape and range of variation, simultaneously accounting for their variability. This normalization is an essential step because the MLP can only use as input discrete fluorescence values taken at fixed depths. Figure 3 shows three schematic examples of fluorescence profile, either non-normalized (Figure 3a) or depth-normalized (Figure 3b), obtained from contrasted open ocean environments, i.e., a typical profile of deep winter-mixing conditions (red curve), a profile with a Deep Chlorophyll Maximum (DCM) characteristic of stratified oligotrophic systems (green curve), and a profile with a subsurface maximum that can be encountered in mesotrophic or eutrophic environments with a relatively shallow mixed layer (blue curve). Based on the examination of Figure 3, it appears that, if the profile is not scaled with respect to depth, the 10 fixed-depth data

© 2015. American Geophysical Union. All Rights Reserved.

Table 3. Trophic Categories Defined According to *Uitz et al.* [2006] and Number of Sampling Stations in Each Category

Surface [TChl] (mg m^{-3})	Number of Stations
<0.04	102
0.04–0.08	206
0.08–0.12	116
0.12–0.2	117
0.2–0.3	81
0.3–0.4	46
0.4–0.8	128
0.8–2.2	70
2.2–4	18
>4	12

points used as input to the MLP may not account for all the vertical variability of the fluorescence profile.

3. FLAVOR Algorithm Development

In the following, [Chl] refers either to the chlorophyll *a* concentration associated with the total phytoplankton biomass ([TChl]) or with the microphytoplankton, nanophytoplankton, and picophytoplankton ([microChl], [nanoChl], and [picoChl], respectively).

Our proposed FLAVOR algorithm is based on an Artificial Neural Network (ANN). It uses as input data a normalized in situ vertical fluorescence profile along with the corresponding geo-location and date to retrieve vertical profiles of chlorophyll *a* associated with the total phytoplankton biomass ([TChl]) and with three phytoplankton size classes ([microChl], [nanoChl], and [picoChl]). FLAVOR is a two-step calibration method, as shown on the flowchart in Figure 4. First, a discrete (10-point) profile of [Chl] ($[Chl]_{MLP}$) is derived from an in situ fluorescence profile based on an ANN. The second step of the method consists in returning a quasi-continuous profile of [Chl] (i.e., with a vertical resolution identical as that of the in situ fluorescence profile, $[Chl]_{cal}$) based on the discrete ANN-derived [Chl] profile. This second step differs depending on the product to be returned by the ANN ([TChl] or [microChl], [nanoChl] and [picoChl]). Two different ANN-based algorithms are developed for inferring vertical profiles of either (1) [TChl] or (2) simultaneously [microChl], [nanoChl], and [picoChl].

3.1. Principles of Multilayer Perceptron (MLP)

The ANN used in this study is a Multilayered Perceptron (MLP) [*Bishop*, 1995] composed of four layers: one input layer, two hidden layers, and one output layer. In each layer, the neurons, which are elementary transfer functions, are interconnected with the neurons of the preceding and following layers by weights. The transfer of information through the MLP is done between the inputs I and outputs o and can be described as:

$$o = f\left(w^{2,o}\ f\left(w^{1,2}\ f\left(w^{i,1}\ I\right)\right)\right) \tag{10}$$

with f a sigmoid nonlinear function:

$$f(x) = a * \frac{\exp(\alpha.x) - 1}{\exp(\alpha.x) + 1} \tag{11}$$

where a and α are two constants. In equation (10), $w^{i,1}$, $w^{1,2}$, and $w^{2,o}$ are the weight matrices that describe the connections between the input and first hidden layers, the first and second hidden layers, and the second hidden layer and the output layer, respectively. The coefficients of the weight matrix are iteratively readjusted during the training of the MLP in order to minimize a cost function defined as the quadratic difference between the desired and computed outputs. To this end, we used the back-propagation conjugate-gradient technique [*Hornik et al.*, 1989; *Bishop*, 1995], which is an iterative optimization method particularly adapted to MLPs. The training data set (80% of the entire initial data set) is further randomly split into two subdata sets (50% each), the so-called "learning" and "test" data sets. During the training process, these two subsets of data are used to cross validate the MLP and prevent from overlearning [*Bishop*, 1995]. Eventually, the validation data set used to evaluate the performance of the MLP is composed of 20% of the entire initial database.

3.2. Application of the MLP to the Chlorophyll Fluorescence and HPLC Pigment Database

Multiple tests were carried out to identify the optimal combination of input and output parameters which yield the best MLP performance. We selected the final following set of parameters as inputs (see also Figure 4): (1) 10 data points from the normalized fluorescence profile taken at regular intervals between the dimensionless depths 0 and 1.3; (2) the depth Z_0; (3) the dimensionless depth ζ at which [Chl] is to be computed; and (4) the location (latitude and longitude) and acquisition date (day of the year) of the considered fluorescence profile. Several tests were also performed to determine the optimal architecture of the MLP. Two types of architecture were tested: one or two hidden layers with a number of neurons in each layer varying between 1 and 50 and between 1 and

© 2015. American Geophysical Union. All Rights Reserved.

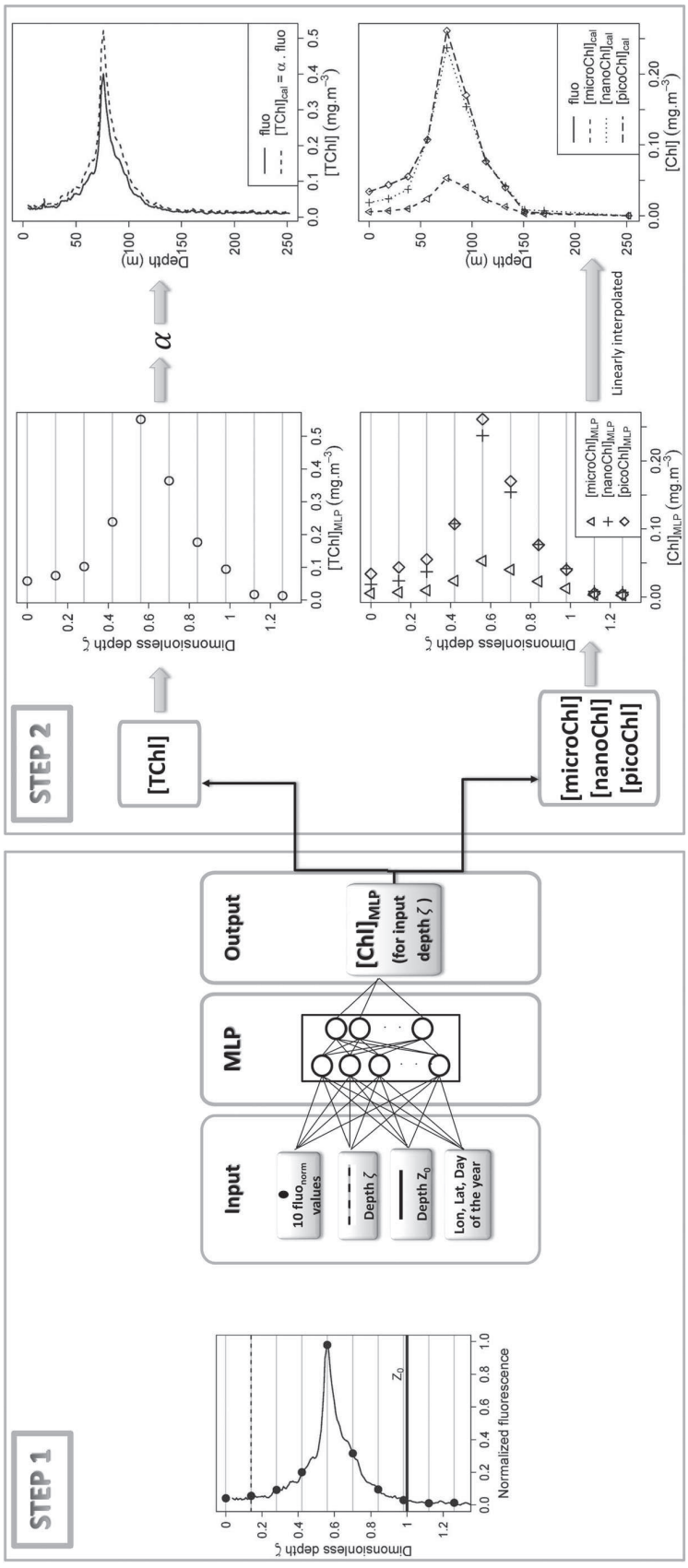


Figure 4. Schematic overview of the two steps involved in the FLAVOR method to retrieve a calibrated [Chl] profile from an uncalibrated fluorescence profile. Step 1: retrieval of a discrete [Chl] profile from the fluorescence profile through the use of a MLP. Step 2: retrieval of a continuous [Chl] profile from the discrete [Chl] profile retrieved by the MLP.

20, respectively (with the number of neurons always higher in the first hidden layer than in the second one in the case of two hidden layers). The optimal architecture was chosen as a compromise between a minimum error of validation and minimum number of neurons on the hidden layers to enhance computing efficiency.

Finally, the optimal architecture was a MLP composed of two hidden layers with seven neurons in the first hidden layer and seven neurons in the second one for the prediction of [TChl], and two hidden layers with nine neurons in the first hidden layer and five neurons in the second one for the simultaneous prediction of [microChl], [nanoChl], and [picoChl].

In addition, different partitions of the training data set have been tested for the validation of the MLP robustness. There were no significant differences in the performances of the MLP.

In order to take advantage of the nonlinearity of the function f (cf. equation (11)) that varies within the range [−1;1], the inputs and outputs of the MLP are transformed to match the [−1;1] domain, with respect to the following equation:

$$x_{i,j} = \frac{2}{3} * \frac{x_{i,j} - mean(x_{i,j})}{\sigma(x_{i,j})} \tag{12}$$

with σ, the standard deviation of the considered input variables x or output variable log([Chl]).

For the longitude and date inputs, we applied a different normalization procedure that accounts for the periodicity of these properties. For example, the month of January (numerically identified as days 1 to 31) and the month of December (identified as days 334 to 365) are relatively similar from a seasonal perspective. In an analogous manner the longitude 0° is equivalent to the longitude 360°. Thus, these variables were transformed in radian units as follows:

$$Lon_{rad} = \frac{Lon * \pi}{180} \tag{13}$$

$$Day_{rad} = \frac{Day * \pi}{182.625} \tag{14}$$

where Lon_{rad} and Day_{rad} are the longitude and day of the year in radian units, respectively, and

© 2015. American Geophysical Union. All Rights Reserved.

the coefficient 182.625 accounts for the number of days per year (365.25) reduced by half.

As final inputs, we used the new variables $\sin(\mathrm{Lon}_{rad})$, $\cos(\mathrm{Lon}_{rad})$, $\sin(\mathrm{Day}_{rad})$, and $\cos(\mathrm{Day}_{rad})$ that vary within the interval $[-1;1]$. For example, $\cos(\mathrm{Day}_{rad})$ is maximum in winter ($\cos(0) = 1$) and minimum in summer ($\cos(\pi) = -1$). Similarly, $\sin(\mathrm{Day}_{rad})$ is maximum in spring ($\sin(\pi/2) = 1$) and minimum in autumn ($\sin(3\pi/2) = -1$). We note that, unlike the longitude and date, the latitude has no periodicity. Therefore, this variable was processed as the x inputs (cf. equation (12)).

3.3. Final Retrieval of Vertical Profiles of Chlorophyll *a* Concentration

The MLP returns 10 discrete normalized values of log([Chl]) as output. In the operational use of the MLP, the output needs to be "denormalized" (i.e., rescaled to physical units) using the inverse formulation of equation (12) with appropriate mean and standard deviation. The resulting discrete profiles of chlorophyll *a* concentration ($[\mathrm{Chl}]_{MLP}$) are then transformed into calibrated vertical profiles ($[\mathrm{Chl}]_{cal}$) with a resolution similar to that of the initial fluorescence profile used as input to the MLP. This procedure is different for the retrieval of [TChl] than for that of [microChl], [nanoChl], and [picoChl].

For the retrieval of the [TChl] vertical profile, we assumed that the in situ fluorescence profile yields the "true" shape of the [TChl] profile. The fluorescence profile used as input to the MLP is scaled to the chlorophyll *a* concentration using the discrete [TChl] values derived from the MLP ($[\mathrm{TChl}]_{MLP}$). In other words, we forced the vertical fluorescence profile to the $[\mathrm{TChl}]_{MLP}$ data points as in *Morel and Maritorena* [2001]. The new profile is then integrated within the layer $0–Z_0$ and used to compute the coefficient α:

$$\alpha = \frac{\int_0^{Z_0} [TChl]_{MLP}(z).dz}{\int_0^{Z_0} fluo(z).dz} \tag{15}$$

In order to obtain a final, calibrated high-vertical resolution [TChl] profile, each data point of the in situ fluorescence profile, fluo(z), is multiplied by the calibration coefficient α as follows:

$$[TChl]_{cal}(z) = \alpha \ . \ fluo(z) \tag{16}$$

The assumption that the fluorescence profile yields the actual vertical distribution of phytoplankton chlorophyll biomass is not applicable at the level of the phytoplankton size group because each group may have its own vertical distribution. Thus, the quasi-continuous final calibrated profiles of chlorophyll *a* concentration associated to the three size classes ($[\mathrm{microChl}]_{cal}$, $[\mathrm{nanoChl}]_{cal}$, and $[\mathrm{picoChl}]_{cal}$) are derived directly from linear interpolation between the 10 discrete concentrations computed by the MLP for each size class.

3.4. Evaluation of the Method Performance

For the retrieval of [TChl] and the simultaneous retrieval of [microChl], [nanoChl], and [picoChl], the validation was done using an independent database, comprising 20% of our entire database (cf. section 2.1). This validation subset is composed of 179 chlorophyll fluorescence profiles with concomitant HPLC chlorophyll *a* reference values collected from 22 oceanographic cruises (Figure S2). The values retrieved from the FLAVOR method were evaluated against in situ HPLC measurements at two different steps: (1) after the first step of the method (performance of the MLP); (2) after the second step of the method (performance of the full calibration method). To assess the first step method performance, the [Chl] derived from the MLP, $[\mathrm{CHL}]_{MLP,}$ is compared to the HPLC linearly interpolated for the 10 input depths (see Figure 4). For the assessment of the second step method performance, the quasi-continuous [Chl] profile, $[\mathrm{Chl}]_{cal}$, is compared to HPLC values for each corresponding depths.

To evaluate the performance of the method, several statistical indices were utilized. We calculated the determination coefficient (R^2) and the slope of the linear regression between the computed calibrated fluorescence values and the in situ reference HPLC values. In addition, the Median Absolute Percent Difference (MAPD) between the reference and predicted values was computed:

© 2015. American Geophysical Union. All Rights Reserved.

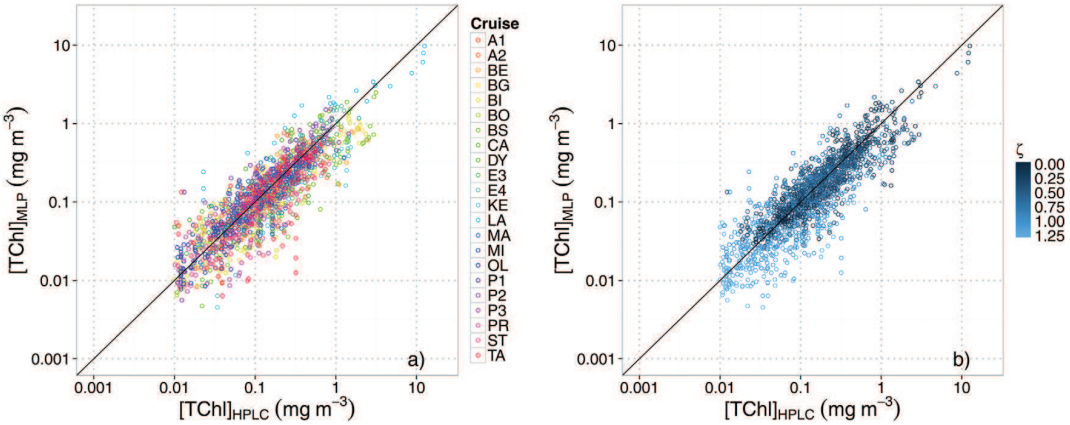


Figure 5. Comparison between [TChl] retrieved by the MLP ($[TChl]_{MLP}$) with HPLC reference ($[TChl]_{HPLC}$). (a) Data identified according to the 22 cruises and (b) data ordered according to the dimensionless depth ζ. The HPLC pigment reference values correspond to linear interpolation of the HPLC measurements for the 10 [TChl] restitution depths of the MLP. The 1:1 line is represented in black in each plot.

$$\mathrm{MAPD} = \frac{(|\mathrm{fluo}_{cal} - [\mathrm{Chl}]|)}{[\mathrm{Chl}]} * 100 \quad (17)$$

As shown by *Campbell* [1995] the chlorophyll *a* concentration follows a lognormal distribution in the open ocean. Therefore, the values were log-transformed prior to the calculation of the statistical indices, except for the MAPD.

4. Results and Discussion

4.1. Retrieval of Chlorophyll *a* Concentration Associated With the Total Phytoplankton Biomass

The accurate retrieval of [TChl] relies both on the performance of the MLP (first step of the method) and on the subsequent application of the calibration coefficient α (second step; cf. Figure 4). The scatterplot of the [TChl] values retrieved from the MLP versus those measured by HPLC (Figure 5) reveals that the data are fairly well distributed around the 1:1 line (see also Table 4 for detailed statistics). This relationship does not show any bias related to the sampling cruise (Figure 5a), therefore suggesting that the proposed method is relatively robust both spatially and temporally. To verify this statement, the relationships between the [TChl] values retrieved from the MLP and measured by HPLC were computed for eight major oceanic basins: Antarctic, Arctic, Mediterranean Sea, Indian, South Pacific, North Pacific, South Atlantic, and North Atlantic Ocean (Figure S3). The determination coefficient, R^2, and the MAPD between the retrieved and reference values for the different basins indicate that the method is robust, with slightly less accurate results for the Arctic basin and the Indian Ocean which are two areas known for data scarcity (see Figure S4). Nevertheless, Figure 5b shows more scatter at low chlorophyll levels which, in general, correspond to the greatest dimensionless depths ζ. This observation supports the results shown in Figure 6 where the determination coefficients and the MAPD between the [TChl] values derived from the MLP and measured by HPLC were computed for six dimensionless depth intervals. The determination coefficient R^2 decreases with increasing ζ, whereas the corresponding MAPD increases with increasing ζ. This supports the idea that the MLP is less robust for retrieving [TChl] values at large depths. Nevertheless, the relationship between $[TChl]_{MLP}$ and $[TChl]_{HPLC}$ for each of the six dimensionless depth intervals is significant ($p_{value} < 0.005$). Furthermore, it should be noticed that the retrieval of $[TChl]_{MLP}$ is

Table 4. Comparison of Values Retrieved by the MLP ($[Chl]_{MLP}$, First Step of Calibration) or Through the Second Step of Calibration ($[Chl]_{cal}$) With Concomitant HPLC Reference Values ($[Chl]_{HPLC}$)[a]

	R^2	a	MAPD (%)
First Step			
$[TChl]_{MLP}$	0.74	0.83	32
$[microChl]_{MLP}$	0.73	0.78	44
$[nanoChl]_{MLP}$	0.60	0.69	35
$[picoChl]_{MLP}$	0.57	0.64	44
Second Step			
$[TChl]_{cal}$	0.68	0.96	40
$[microChl]_{cal}$	0.72	0.75	46
$[nanoChl]_{cal}$	0.64	0.68	35
$[picoChl]_{cal}$	0.58	0.61	40

[a]Determination coefficient (R^2) and slope (a) corresponding to linear regression analyses between retrieved and reference values. For each case, the MAPD (Median Absolute Percent Difference) between retrieved and reference values is also indicated.

© 2015. American Geophysical Union. All Rights Reserved.

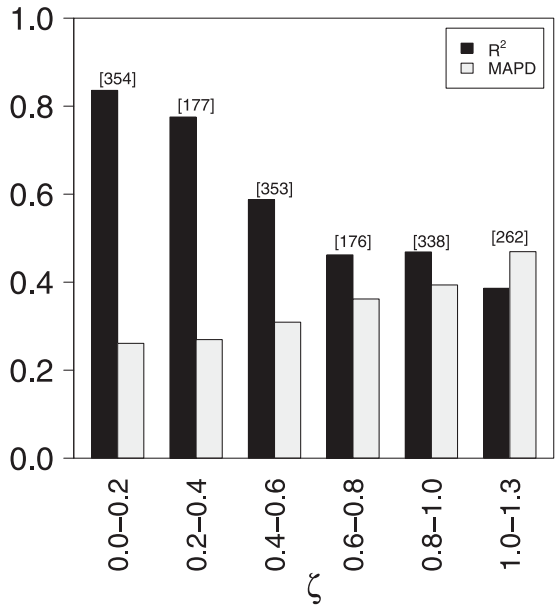


Figure 6. Determination coefficient R^2 and Median Absolute Percent Difference (MAPD) of the linear models between [TChl] retrieved by the MLP ($[TChl]_{MLP}$) and HPLC reference ($[TChl]_{HPLC}$) values computed for several dimensionless depth intervals. The number of points corresponding to each interval is identified in bracket. The MAPD is presented here divided by 100 so values range is 0–1.

less critical for deeper layers than for upper layers where most of the phytoplankton biomass occurs. The calibration coefficient α applied to the in situ fluorescence profile takes into account only the MLP-predicted values comprised between surface and Z_0 (cf. section 3.3).

The second step of our method allows the retrieval of calibrated [TChl] ($[TChl]_{cal}$) profiles with the same vertical resolution as the initial fluorescence profiles. The performance of the method is assessed through the comparison of $[TChl]_{cal}$ with the corresponding in situ HPLC measurements, $[TChl]_{HPLC}$ (Figure 7 and Table 4). The relationship between $[TChl]_{cal}$ and $[TChl]_{HPLC}$ appears more scattered than the relationship between $[TChl]_{MLP}$ and $[TChl]_{HPLC}$ (i.e., first step of the method; Figure 5 and Table 4). This might be due, at least partially, to the reintroduction of the signal noise of the in situ measured fluorescence profile into the MLP-retrieved profile that can be sometimes particularly pronounced. However, compared with the HPLC reference values, the fluorescence profiles calibrated into [TChl] appear globally unbiased.

The impact of the daytime Non-Photochemical Quenching (NPQ) [see, e.g., *Cullen and Lewis*, 1995], which is responsible for a reduction in the chlorophyll fluorescence at high irradiance, deserves further consideration in the evaluation of the proposed method. In the first step of FLAVOR, the MLP learning is based on in situ measurements so that the NPQ is implicitly taken into consideration and thus corrected for a proper restitution of the chlorophyll *a* concentration (i.e., to a given quenched chlorophyll fluorescence profile is associated an unquenched HPLC-determined chlorophyll *a* profile). In the second step of FLAVOR, the initial fluorescence profile shape is used to compute the final $[TChl]_{cal}$ (see equation (16)); thus, the potential bias due to the NPQ is implicitly reintroduced. This is admittedly a weakness of the present method. If density profiles are acquired simultaneously to fluorescence profiles, the NPQ could be corrected following the method of *Xing et al.* [2012]. This method involves substituting the fluorescence values acquired within the mixed layer by the maximum value within this layer. If the concomitant acquisition of density is not available, there is presently no solution to overcome the potential issue of the NPQ.

Finally, the performance of FLAVOR can be compared with that of other methods developed to retrieve the vertical distribution of the total chlorophyll *a* concentration from fluorescence profiles. The method proposed by *Lavigne et al.* [2012] was evaluated using the long-term time series data sets from the BATS, HOT, and Dyfamed stations. The method developed by *Mignot et al.* [2011] and the method presented here were evaluated with data sets representative of the global ocean. Although the performances of these three methods were assessed on the basis of different data sets, the statistical indices of performance of these methods can be compared, at least in an indicative manner. The MAPD is 33%, 31% and 32% for the methods developed by *Mignot et al.* [2011], *Lavigne et al.* [2012] and FLAVOR, respectively. In other words, FLAVOR performs well in comparison to the other methods and presents the additional advantage of being self-sufficient, i.e., it does not require any other external information or data to retrieve the phytoplankton total or class-specific chlorophyll *a* concentration (except the geo-location and date of acquisition which

© 2015. American Geophysical Union. All Rights Reserved.

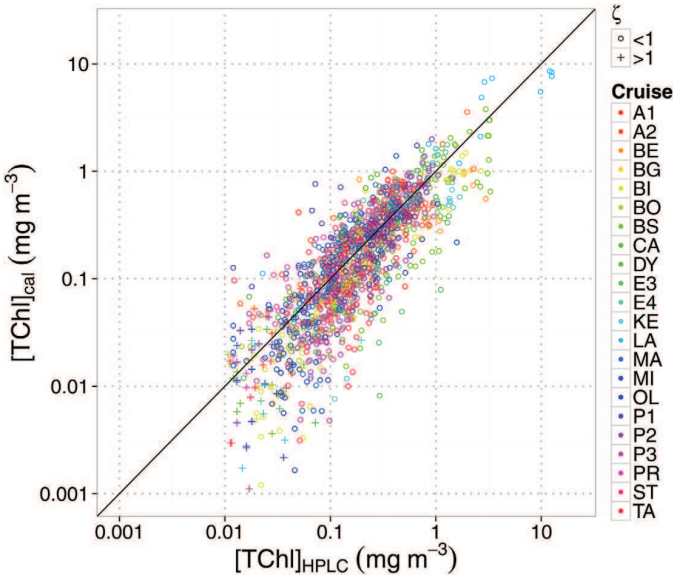


Figure 7. Comparison between [TChl] retrieved by the calibration method ($[TChl]_{cal}$) and reference HPLC values ($[TChl]_{HPLC}$) for the 22 cruises analyzed in this study. The cruises are represented by different colors and according to two intervals of dimensionless depths ($\zeta < 1$: circle; and $\zeta > 1$: cross). The 1:1 line is represented in black.

are information always available with the fluorescence profiles). For instance, the method of *Lavigne et al.* [2012] uses the remotely sensed surface chlorophyll *a* concentration as additional input information, whereas that of *Mignot et al.* [2011] requires a classification of fluorescence profiles into two types of shapes (mixed or stratified).

4.2. Retrieval of Chlorophyll *a* Concentration Associated With the Three Phytoplankton Size Classes

The performance of FLAVOR is here evaluated for the retrieval of chlorophyll *a* profiles associated with the three phytoplankton pigment-based size classes from in situ fluorescence profiles. For the first step of the method, there is a reasonably good agreement between the MLP-retrieved values (i.e., $[microChl]_{MLP}$, $[nanoChl]_{MLP}$, and $[picoChl]_{MLP}$) and the corresponding in situ HPLC determinations (Figure 8 and Table 4); no systematic bias is observed. The method performance is higher for [microChl] that presents the largest range of variation of the three size classes (0.0003–10 mg m^{-3}), than for [picoChl] that varies over a limited range of relatively low concentrations (0.0005–0.3 mg m^{-3}). The method performance is intermediate for [nanoChl] that varies over a limited range of moderate concentrations (0.003–3 mg m^{-3}). In addition, the relationships between the MLP-retrieved and reference in situ values were computed for eight major oceanic basins (see Figure S3 and section 4.1). Figures S5–S7, which display the determination coefficients R^2 and the MAPD between the retrieved and reference values for the different basins, show that the method is relatively robust spatially.

The performance of the calibration method is now evaluated after the second step of the method (see Figure 4) which allows the retrieval of "high-resolution" vertical profiles of $[microChl]_{cal}$, $[nanoChl]_{cal}$, and $[picoChl]_{cal}$. The relationships between these derived quantities and their HPLC-determined counterparts (Figure 9) are quite similar to those observed at the first step of the calibration method (cf. Figure 8). These results are coherent with the statistical analysis (Table 4) that confirms that the second step of the calibration method does not deteriorate the initial, MLP-based retrieval of $[microChl]_{MLP}$, $[nanoChl]_{MLP}$, and $[picoChl]_{MLP}$ (first step of the method). On Figure 9, a few outliers can be identified, which are related to the very low chlorophyll *a* concentrations typical of very deep samples.

When studying the specific $[Chl]_{cal}$ profiles individually, their vertical distributions appear closely similar to the vertical distributions of their corresponding discrete profiles estimated by HPLC (data not shown). This observation is important for validating the present method because the second step of the calibration is here not "forced" as for the [TChl] retrieval. As might be expected, we verified that the retrieval of [TChl] is better with a specific MLP than through the sum of [microChl], [nanoChl], and [picoChl] retrieved simultaneously by the MLP. This likely results from the addition of errors associated with each of the three retrieved specific [Chl], whereas for the retrieval of the [TChl], only one error of prediction is involved. However, the retrieval of [TChl] is coherent with the sum of [microChl], [nanoChl], and [picoChl] retrievals as there is no systematic bias between both quantities when compared with a linear model ($R^2 = 0.81$ and slope = 0.97).

In summary, compared with the HPLC reference values, the fluorescence profiles calibrated simultaneously into [microChl], [nanoChl], and [picoChl] are globally unbiased, with R^2 values of 0.72, 0.64, and 0.58,

© 2015. American Geophysical Union. All Rights Reserved.

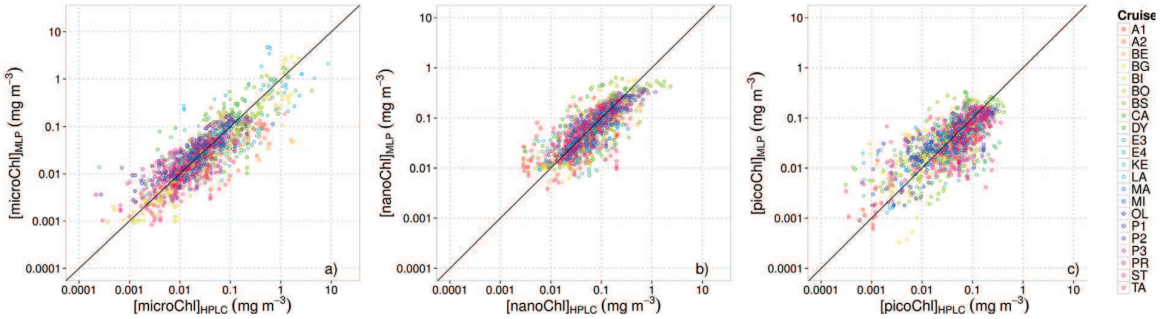


Figure 8. Comparison between [Chl] retrieved by the MLP ($[Chl]_{MLP}$) and measured by HPLC ($[Chl]_{HPLC}$) for the three pigment size classes [Chl]. (a) [microChl], (b)[nanoChl], and (c) [picoChl]. The HPLC pigment reference values correspond to a linear interpolation of the HPLC measurements for the 10 [Chl] restitution depths of the MLP. Each color refers to one of the 22 cruises analyzed in this study. The 1:1 line is represented in black in each plot.

respectively, and a median error of 46%, 35%, and 40%, respectively (Table 4). The slightly lower performances of the retrieval of the vertical distribution of [microChl], [nanoChl] and [picoChl] compared to that of [TChl] is essentially caused by a larger natural variability introduced by the different phytoplankton communities.

4.3. Possible Application: Development of Global and Regional 3-D Climatologies of Chlorophyll *a* Concentration and Associated Phytoplankton Communities

FLAVOR can be considered as a robust method to reconcile fluorescence profiles from extremely diverse sources (with respect to both sensors and platforms) by converting these profiles into a common "reference": the vertical profile of chlorophyll *a* concentration. To our knowledge, it is so far the only method that allows various fluorescence data sets to become interoperable. This potential interoperability not only concerns the "historical" data from diverse origins used in this study, but also the data that will be acquired in the future. This especially concerns the tremendous amount of fluorescence profiles that will likely result from the growing use of autonomous platforms such as bio-optical profiling floats [*Claustre et al*., 2010b]. The FLAVOR algorithm is, obviously, less accurate than dedicated analytical methods (HPLC) for determining the chlorophyll *a* concentration associated with the total phytoplankton biomass or with three major size classes. Nevertheless, this lower accuracy is largely compensated by the potential of our method to process a larger amount of data, covering a much larger space and time domain. As supported by our statistical results, we argue that FLAVOR could be used to produce climatologies of chlorophyll *a* concentration profiles by using as input large amounts of historical and modern fluorescence data.

To illustrate such potential, the FLAVOR method was applied to the BATS (Bermuda Atlantic Time Series Study) site where a large amount of HPLC and fluorescence profiles have been acquired over the 1998–2012 time period. In this data set, there are approximately twenty times more fluorescence profiles available

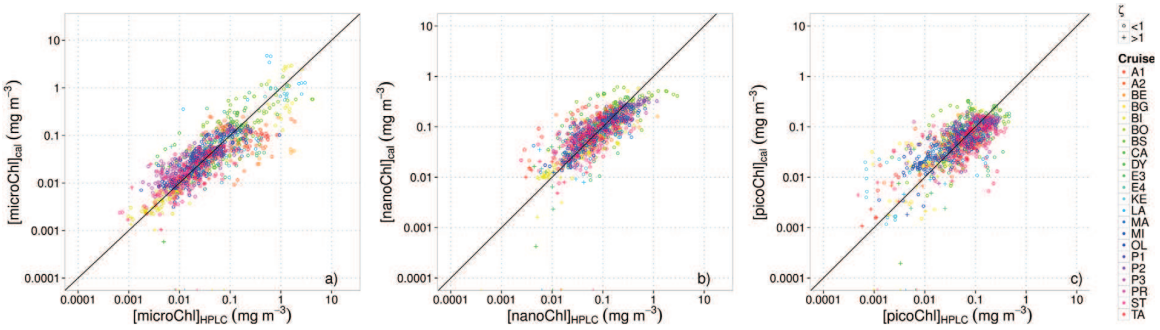


Figure 9. Comparison between [Chl] retrieved by the calibration method ($[Chl]_{cal}$) and measured by HPLC ($[Chl]_{HPLC}$) for the three pigment size classes (a) [microChl], (b) [nanoChl], and (c) [picoChl]. The 22 cruises analyzed in this study are represented by different colors and according to two intervals of dimensionless depths ($\zeta < 1$: circle; and $\zeta > 1$: cross). The 1:1 line is represented in black in each plot.

© 2015. American Geophysical Union. All Rights Reserved.

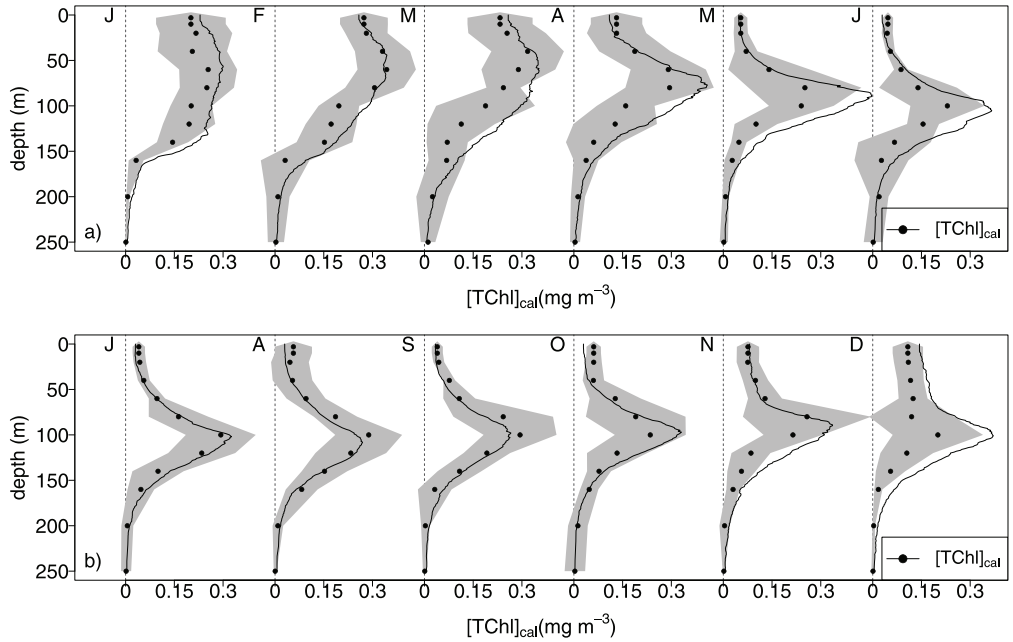


Figure 10. Bermuda Atlantic Time series Study: comparisons of the monthly climatologies of HPLC-based chlorophyll *a* concentrations ($[TChla]_{HPLC}$ in dots) with FLAVOR-retrieved chlorophyll *a* concentrations ($[TChl]_{cal}$ in continuous black line). The gray polygons represent the standard deviation of the HPLC-based climatology. (a) Comparisons presented for the months from January to June. (b) Comparisons presented for the months from July to December.

than HPLC discrete profiles. Hence, in addition to the benefit of providing depth-continuous chlorophyll *a* profiles, our method enables to process all fluorescence profiles thus providing twenty times more profiles than the HPLC method in this region. It is worth to note that this data set was not taken into account for the training of our method. Figure 10 shows monthly climatological fluorescence profiles calibrated into [TChl] compared with the corresponding HPLC measurements. Overall, the monthly pattern in the climatological $[TChl]_{HPLC}$ profiles is well reproduced by the fluorescence-based climatological profiles retrieved from the algorithm. The fluorescence-based climatological profiles indeed generally fall within one standard deviation of the HPLC-based climatological profiles. The observed deviations (i.e., climatological fluorescence profiles outside the envelope of one standard deviation of $[TChl]_{HPLC}$ in Figure 10) can obviously be linked to a FLAVOR failure and/or to the use of climatologies. Nevertheless, the observed deviations might also result from the availability of more fluorescence than HPLC profiles in the BATS database. As a result, the fluorescence records might have captured a broader diversity of situations (e.g., some local events associated to mesoscale activity). Another explanation may be a bias in the HPLC-based climatology caused by sampling at fixed depths, which potentially misses certain vertical features such as a DCM. In any case, the overall agreement between both the fluorescence- and HPLC-based climatological chlorophyll *a* profiles remains satisfactory considering that no BATS data were present in the MLP learning database. This comparison also indicates that our method is not location biased. Thus, the present analysis clearly stands as an additional and independent validation, confirming the robustness of the FLAVOR method. It is also an a posteriori confirmation that the training data set is likely representative of the full range of situations occurring in open ocean waters.

Besides the retrieval of chlorophyll *a* concentration associated with the total phytoplankton biomass ([TChl]), the retrieval of the phytoplankton class-specific chlorophyll *a* concentration ([Chl]) was also evaluated (Figure 11). As expected, the performance of the method for retrieving [Chl] is not as good as for retrieving [TChl]. Nevertheless the fluorescence-based monthly climatologies of [Chl] associated with the three phytoplankton size classes compare relatively well with the corresponding HPLC-based climatologies. Again, given the independence of the BATS data set with respect to MLP training data set, these results appear very promising.

© 2015. American Geophysical Union. All Rights Reserved.

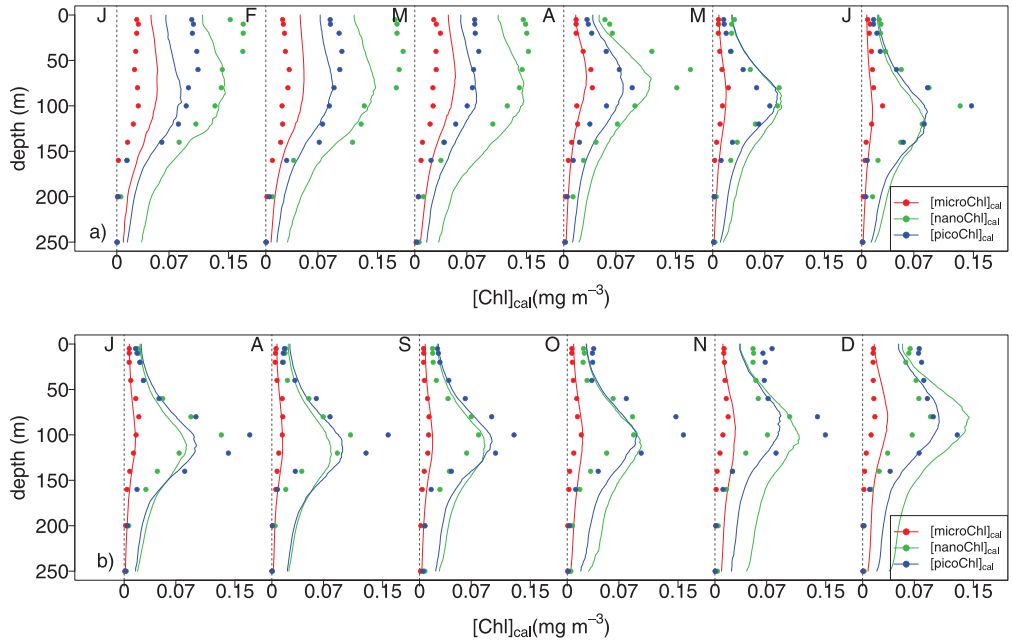


Figure 11. Bermuda Atlantic Time series Study: comparisons of the monthly climatologies of HPLC-based chlorophyll *a* concentrations associated to phytoplankton community size indices ($[Chl]_{HPLC}$ in dots) with FLAVOR-retrieved concentrations ($[Chl]_{CAL}$ in continuous lines). Red: [microChl]; green: [nanoChl]; blue: [picoChl]. (a) Comparisons presented for the months from January to June. (b) Comparisons presented for the months from July to December.

The above application of FLAVOR to the BATS time series can be considered as a preview of the future of climatological fields of the vertical distribution of phytoplankton chlorophyll *a* biomass and associated community composition at a global scale as well as for specific regions. To our knowledge, the current climatologies of chlorophyll *a* concentration available in the World Ocean Atlas (WOA) [*Conkright et al.*, 2002] are based on fixed depths (discrete data) and with no concurrent quantitative information on phytoplankton assemblages. The utilization of large data sets of fluorescence profiles will offer very significant improvements over existing climatologies. Quantitatively it will be possible to have access to an increasing amount of data with enhanced time and space resolution measurements. A parallel can be made with CTD-based temperature and salinity profiles for which, thanks to the development of the Argo program and the million profiles achieved in 2012, robust climatologies are now available.

Apart from a quantitative estimation of phytoplankton biomass through chlorophyll *a* concentration, the possibility of further separating this biomass into three main phytoplankton size groups is also of crucial importance. It is indeed recognized that taking into consideration the nature of the phytoplankton assemblage is key in addressing the marine carbon cycle in a more comprehensive way, for example, using bio-optical [*Uitz et al.*, 2010] or biogeochemical modeling [*Aumont et al.*, 2003; *Le Quere et al.*, 2005; *Hood et al.*, 2006]. FLAVOR will allow the provision of global-scale data required for the initialization and/or validation of such models. The quality and representativeness of the data sets used for the initialization and validation of modeling approaches will improve over time as the density of fluorescence data sets progressively increases.

5. Conclusions and Perspectives

The FLAVOR method allows the assessment of the vertical distribution of phytoplankton chlorophyll *a* biomass and associated phytoplankton communities from in situ chlorophyll fluorescence profile. Based on an Artificial Neural Network, FLAVOR enables the retrieval of the chlorophyll *a* concentration profiles associated (1) with the total phytoplankton biomass and (2), simultaneously, with three pigment-derived size classes, from the sole knowledge of the shape of the fluorescence profile and its acquisition date and geo-location. Consequently, for the first time, the large data set of fluorescence profiles collected

© 2015. American Geophysical Union. All Rights Reserved.

since the 1970s could be harmonized and made interoperable in terms of chlorophyll *a* concentration. Additionally, it is the first method that allows for the retrieval of the composition of phytoplankton communities from a fluorescence profile. However, it is important to note that to apply the method to fluorescence profiles acquired pre-1991, the assumption has to be made that the relationship between the phytoplankton biomass and the community composition with the fluorescence profile is the same as for our data set (post 1991).

Validation results have been presented here regarding the retrieval of total and class-specific chlorophyll *a* concentration versus in situ HPLC reference measurements. Our method appears to be spatially and temporally robust as the relationships between the retrieved and reference in situ values do not show systematic bias with respect to the different oceanic regions and cruises regardless of the product to be retrieved (Figures 5a and 7–9 and Figures S4–S7). Additionally, because FLAVOR was developed using a data set that is representative of most of the hydrologic and trophic conditions prevailing in the open ocean, it can be considered potentially applicable to any situation occurring in the global open ocean. Nevertheless, it should be emphasized that, although FLAVOR is applicable to situations in which it has not been trained (e.g., the BATS time series; see Figures 10 and 11), its use deserves some caution. FLAVOR is not a method for use on a profile-by-profile basis, where a single fluorescence profile would be injected to retrieve accurate profiles of chlorophyll *a* concentration for the entire algal biomass and associated size indices. Instead, the method is intended for use on large data sets for deriving vertical chlorophyll *a* climatologies from which some regional or temporal trends might possibly be extracted. Such data sets could be exploited to improve the open ocean climatologies of chlorophyll *a* concentration.

Hence, thanks to the ongoing and future availability of spatially and temporally well-resolved data sets, it will become possible to develop 3-D and even 4-D global climatologies of chlorophyll *a* concentration and associated community composition in terms of three major phytoplankton size classes. These types of climatologies are not only required for the initialization and validation of biogeochemical models but may also serve as benchmark for documenting possible changes in phytoplankton biomass and distribution in the global ocean.

Acknowledgments

This paper is a contribution to the Remotely Sensed Biogeochemical Cycles in the Ocean (remOcean) project, funded by the European Research Council (grant agreement 246777), to the French Bio-Argo project funded by CNES-TOSCA. The French PROOF and CYBER programs are acknowledged for their support of most of the cruises where fluorescence and HPLC profiles were sampled. Mustapha Ouhssain is acknowledged for his contribution to HPLC analysis. The fluorescence and HPLC profiles used in this study can be obtained on request at sauzede@obs-vlfr.fr and claustre@obs-vlfr.fr. The authors would like to thank all the staff of BATS program for the periodic measurements of oceanographic variables and for the free distribution of data online (http://bats.bios.edu/). Finally, we are grateful to two anonymous reviewers for their valuable comments and suggestions.

References

Aumont, O., E. Maier-Reimer, S. Blain, and P. Monfray (2003), An ecosystem model of the global ocean including Fe, Si, P colimitations, *Global Biogeochem. Cycles*, *17*(2), 1060, doi:10.1029/2001GB001745.

Ben Mustapha, Z., S. Alvain, C. Jamet, H. Loisel, and D. Dessailly (2013), Automatic classification of water-leaving radiance anomalies from global SeaWiFS imagery: Application to the detection of phytoplankton groups in open ocean waters, *Remote Sens. Environ.*, *146*, 97–112, doi:10.1016/j.rse.2013.08.046.

Bishop, C. M. (1995), *Neural Networks for Pattern Recognition*, 482 pp., Oxford Univ. Press, Oxford, U. K.

Brewin, R. J. W., S. Sathyendranath, T. Hirata, S. J. Lavender, R. M. Barciela, and N. J. Hardman-Mountford (2010), A three-component model of phytoplankton size class for the Atlantic Ocean, *Ecol. Modell.*, *221*(11), 1472–1483, doi:10.1016/j.ecolmodel.2010.02.014.

Brewin, R. J. W., S. Sathyendranath, P. K. Lange, and G. Tilstone (2014), Comparison of two methods to derive the size-structure of natural populations of phytoplankton, *Deep Sea Res., Part I*, *85*, 72–79, doi:10.1016/j.dsr.2013.11.007.

Bricaud, A., E. Bosc, and D. Antoine (2002), Algal biomass and sea surface temperature in the Mediterranean Basin, *Remote Sens. Environ.*, *81*(2–3), 163–178, doi:10.1016/S0034-4257(01)00335-2.

Bricaud, A., C. Mejia, D. Blondeau-Patissier, H. Claustre, M. Crepon, and S. Thiria (2007), Retrieval of pigment concentrations and size structure of algal populations from their absorption spectra using multilayered perceptrons, *Appl. Opt.*, *46*(8), 1251–1260.

Buitenhuis, E. T., et al. (2012), Picophytoplankton biomass distribution in the global ocean, *Earth Syst. Sci. Data*, *4*(1), 37–46, doi:10.5194/essd-4-37-2012.

Campbell, J. W. (1995), The lognormal distribution as a model for bio-optical variability in the sea, *J. Geophys. Res.*, *100*(C7), 13,231–13,254.

Claustre, H. (1994), The trophic status of various oceanic provinces as revealed by phytoplankton pigment signatures, *Limnol. Oceanogr.*, *39*(5), 1206–1210.

Claustre, H., and J.-C. Marty (1995), Specific phytoplankton biomasses and their relation to primary production in the tropical North Atlantic, *Deep Sea Res., Part I*, *42*(8), 1475–1493, doi:10.1016/0967-0637(95)00053-9.

Claustre, H., P. Kerhervé, J. C. Marty, L. Prieur, C. Videau, and J.-H. Hecq (1994), Phytoplankton dynamics associated with a geostrophic front: Ecological and biogeochemical implications, *J. Mar. Res.*, *52*(4), 711–742, doi:10.1357/0022240943077000.

Claustre, H., A. Morel, M. Babin, C. Cailliau, D. Marie, J.-C. Marty, D. Tailliez, and D. Vaulot (1999), Variability in particle attenuation and chlorophyll fluorescence in the tropical Pacific: Scales, patterns, and biogeochemical implications, *J. Geophys. Res.*, *104*(C2), 3401–3422, doi:10.1029/98JC01334.

Claustre, H., F. Fell, K. Oubelkheir, L. Prieur, A. Sciandra, B. Gentili, and M. Babin (2000), Continuous monitoring of surface optical properties across a geostrophic front: Biogeochemical inferences, *Limnol. Oceanogr.*, *45*(2), 309–321.

Claustre, H., et al. (2004), An intercomparison of HPLC phytoplankton pigment methods using in situ samples: Application to remote sensing and database activities, *Mar. Chem.*, *85*(1–2), 41–61, doi:10.1016/j.marchem.2003.09.002.

Claustre, H., M. Babin, D. Merien, J. Ras, L. Prieur, S. Dallot, O. Prasil, H. Dousova, and T. Moutin (2005), Toward a taxon-specific parameterization of bio-optical models of primary production: A case study in the North Atlantic, *J. Geophys. Res.*, *110*, C07S12, doi:10.1029/2004JC002634.

© 2015. American Geophysical Union. All Rights Reserved.

Claustre, H., et al. (2010a), Bio-optical profiling floats as new observational tools for biogeochemical and ecosystem studies: Potential synergies with ocean color remote sensing, in *Proceedings of the OceanObs 09: Sustained Ocean Observations and Information for Society Conference*, vol. 2, edited by J. Hall, D. E. Harrison, and D. Stammer, ESA Publ., Venice, Italy.

Claustre, H., et al. (2010b), Guidelines towards an integrated ocean observation system for ecosystems and biogeochemical cycles, in *Proceedings of the OceanObs 09: Sustained Ocean Observations and Information for Society Conference*, vol. 1, edited by J. Hall, D. E. Harrison, and D. Stammer, ESA Publ., Venice, Italy.

Conkright, M. E., R. A. Locarnini, H. E. Garcia, T. D. O'Brien, T. P. Boyer, C. Stephens, and J. I. Antonov (2002), *World Ocean Atlas 2001: Objective Analyses, Data Statistics, and Figures* [CD-ROM], U.S. Dep. of Commer., Natl. Oceanic and Atmos. Admin., Natl. Oceanogr. Data Cent., Ocean Clim. Lab., Silver Spring, Md.

Crombet, Y., K. Leblanc, B. Quéguiner, T. Moutin, P. Rimmelin, J. Ras, H. Claustre, N. Leblond, L. Oriol, and M. Pujo-Pay (2011), Deep silicon maxima in the stratified oligotrophic Mediterranean Sea, *Biogeosciences*, *8*(2), 459–475, doi:10.5194/bg-8-459-2011.

Cullen, J. J., and M. R. Lewis (1995), Biological processes and optical measurements near the sea surface: Some issues relevant to remote sensing, *J. Geophys. Res.*, *100*(C7), 13,255–13,266, doi:10.1029/95JC00454.

Cunningham, A. (1996), Variability of in-vivo chlorophyll fluorescence and its implication for instrument development in bio-optical oceanography, *Sci. Mar.*, *60*(1), 309–315.

Devred, E., S. Sathyendranath, V. Stuart, and T. Platt (2011), A three component classification of phytoplankton absorption spectra: Application to ocean-color data, *Remote Sens. Environ.*, *115*(9), 2255–2266, doi:10.1016/j.rse.2011.04.025.

D'Ortenzio, F., et al. (2010), White Book on Oceanic autonomous Platforms for Biogeochemical Studies: Instrumentation and Measure (PABIM), version 1.3. [Available at: http://www.obs-vlfr.fr/OAO/file/PABIM white book version1.3.pdf.]

Falkowski, P., D. A. Kiefer, O. S. Division, and L. Angeles (1985), Chlorophyll a fluorescence in phytoplankton: Relationship to photosynthesis and biomass, *J. Plankton Res.*, *7*(5), 715–731.

Friedrich, T., and A. Oschlies (2009), Neural network-based estimates of North Atlantic surface pCO 2 from satellite data: A methodological study, *J. Geophys. Res.*, *114*, C03020, doi:10.1029/2007JC004646.

Gordon, H. R., and W. R. McCluney (1975), Estimation of the depth of sunlight penetration in the sea for remote sensing, *Appl. Opt.*, *14*(2), 413–416, doi:10.1364/AO.14.000413.

Gross, L., S. Thiria, R. Frouin, and B. G. Mitchell (2000), Artificial neural networks for modeling the transfer function between marine reflectance and phytoplankton pigment concentration, *J. Geophys. Res.*, *105*(C2), 3483–3495, doi:10.1029/1999JC900278.

Hood, R. R., et al. (2006), Pelagic functional group modeling: Progress, challenges and prospects, *Deep Sea Res., Part II*, *53*(5–7), 459–512, doi:10.1016/j.dsr2.2006.01.025.

Hornik, K., M. Stinchcombe, and H. White (1989), Multilayer feedforward networks are universal approximators, *Neural Networks*, *2*, 359–366.

Huot, Y., M. Babin, and F. Bruyant (2013), Photosynthetic parameters in the Beaufort Sea in relation to the phytoplankton community structure, *Biogeosciences*, *10*(5), 3445–3454, doi:10.5194/bg-10-3445-2013.

Jamet, C., H. Loisel, and D. Dessailly (2012), Retrieval of the spectral diffuse attenuation coefficient K d (λ) in open and coastal ocean waters using a neural network inversion, *J. Geophys. Res.*, *117*, C10023, doi:10.1029/2012JC008076.

Johnson, K. S., L. J. Coletti, H. W. Jannasch, C. M. Sakamoto, D. D. Swift, and S. C. Riser (2013), Long-term nitrate measurements in the ocean using the in situ ultraviolet spectrophotometer: Sensor integration into the APEX profiling float, *J. Atmos. Oceanic Technol.*, *30*(8), 1854–1866, doi:10.1175/JTECH-D-12-00221.1.

Kiefer, D. A. (1973), Chlorophyll a fluorescence in marine centric diatoms: Responses of chloroplasts to light and nutrient stress, *Mar. Biol.*, *23*(1), 39–46, doi:10.1007/BF00394110.

Lavigne, H., F. D'Ortenzio, H. Claustre, and A. Poteau (2012), Towards a merged satellite and in situ fluorescence ocean chlorophyll product, *Biogeosciences*, *9*(6), 2111–2125, doi:10.5194/bg-9-2111-2012.

Le Quere, C., et al. (2005), Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models, *Global Change Biol*, *11*, 2016–2040.

Leblanc, K., et al. (2012), A global diatom database—Abundance, biovolume and biomass in the world ocean, *Earth Syst. Sci. Data*, *4*, 149–165.

Lek, S., and J. F. Guégan (1999), Artificial neural networks as a tool in ecological modelling, an introduction, *Ecol. Modell.*, *120*(2–3), 65–73, doi:10.1016/S0304-3800(99)00092-7.

Lorenzen, C. J. (1966), A method for the continuous measurement of in vivo chlorophyll concentration, *Deep Sea Res. Oceanogr. Abstr.*, *13*(2), 223–227, doi:10.1016/0011-7471(66)91102-8.

Marty, J.-C., J. Chiavérini, M.-D. Pizay, and B. Avril (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 Res., Part II*, *49*(11), 1965–1985, doi:10.1016/S0967-0645(02)00022-X.

Marzban, C. (2009), Basic statistics and basic AI: Neural networks, in *Artificial Intelligence Methods in the Environmental Sciences*, edited by S. E. Haupt, A. Pasini, and C. Marzban, pp. 15–47, Springer, Netherlands.

McClain, C. R. (2009), A decade of satellite ocean color observations, *Annu. Rev. Mar. Sci.*, *1*, 19–42, doi:10.1146/annurev.marine.010908.163650.

Mignot, A., H. Claustre, F. D'Ortenzio, X. Xing, A. Poteau, and J. Ras (2011), From the shape of the vertical profile of in vivo fluorescence to chlorophyll-a concentration, *Biogeosciences*, *8*(2), 3697–3737, doi:10.5194/bgd-8-3697-2011.

Morel, A., and J.-F. Berthon (1989), Surface pigments, algal biomass profiles, and potential production of the euphotic layer: Relationships reinvestigated in view of remote-sensing applications, *Limnol. Oceanogr.*, *34*(8), 1545–1562.

Morel, A., and S. Maritorena (2001), Bio-optical properties of oceanic waters: A reappraisal, *J. Geophys. Res.*, *106*(C4), 7163–7180, doi:10.1029/2000JC000319.

Morel, A., B. Gentili, M. Chami, and J. Ras (2006), Bio-optical properties of high chlorophyll Case 1 waters and of yellow-substance-dominated Case 2 waters, *Deep Sea Res., Part I*, *53*(9), 1439–1459, doi:10.1016/j.dsr.2006.07.007.

Palacz, A. P., M. A. St. John, R. J. W. Brewin, T. Hirata, and W. W. Gregg (2013), Distribution of phytoplankton functional types in high-nitrate, low-chlorophyll waters in a new diagnostic ecological indicator model, *Biogeosciences*, *10*(11), 7553–7574, doi:10.5194/bg-10-7553-2013.

Peloquin, J., et al. (2013), The MAREDAT global database of high performance liquid chromatography marine pigment measurements, *Earth Syst. Sci. Data*, *5*(1), 109–123, doi:10.5194/essd-5-109-2013.

Raitsos, D. E., S. J. Lavender, C. D. Maravelias, J. Haralabous, A. J. Richardson, and P. C. Reid (2008), Identifying four phytoplankton functional types from space: An ecological approach, *Limnol. Oceanogr.*, *53*(2), 605–613.

© 2015. American Geophysical Union. All Rights Reserved.

Ras, J., H. Claustre, and J. Uitz (2008), Spatial variability of phytoplankton pigment distributions in the Subtropical South Pacific Ocean: Comparison between in situ and predicted data, *Biogeosciences*, *5*(2), 353–369.

Siegel, D. A., et al. (2013), Regional to global assessments of phytoplankton dynamics from the SeaWiFS mission, *Remote Sens. Environ.*, *135*, 77–91, doi:10.1016/j.rse.2013.03.025.

Telszewski, M., et al. (2009), Estimating the monthly pCO2 distribution in the north Atlantic using a self-organizing neural network, *Biogeosciences*, *6*, 1405–1421.

Uitz, J., H. Claustre, A. Morel, and S. B. Hooker (2006), Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll, *J. Geophys. Res.*, *111*, C08005, doi:10.1029/2005JC003207.

Uitz, J., H. Claustre, F. B. Griffiths, J. Ras, N. Garcia, and V. Sandroni (2009), A phytoplankton class-specific primary production model applied to the Kerguelen Islands region (Southern Ocean), *Deep Sea Res., Part I*, *56*(4), 541–560, doi:10.1016/j.dsr.2008.11.006.

Uitz, J., H. Claustre, B. Gentili, and D. Stramski (2010), Phytoplankton class-specific primary production in the world's oceans: Seasonal and interannual variability from satellite observations, *Global Biogeochem. Cycles*, *24*, GB3016, doi:10.1029/2009GB003680.

Vidussi, F., H. Claustre, J. Bustillos-Guzmàn, C. Cailliau, and J.-C. Marty (1996), Determination of chlorophylls and carotenoids of marine phytoplankton: Separation of chlorophyll a from divinylchlorophyll a and zeaxanthin from lutein, *J. Plankton Res.*, *18*(12), 2377–2382, doi: 10.1093/plankt/18.12.2377.

Vidussi, F., H. Claustre, B. B. Manca, A. Luchetta, and J.-C. Marty (2001), Phytoplankton pigment distribution in relation to upper thermocline circulation in the eastern Mediterranean Sea during winter, *J. Geophys. Res.*, *106*(C9), 19,939–19,956, doi:10.1029/1999JC000308.

Xing, X., H. Claustre, S. Blain, F. D'Ortenzio, D. Antoine, J. Ras, and C. Guinet (2012), Quenching correction for in vivo chlorophyll fluorescence acquired by autonomous platforms: A case study with instrumented elephant seals in the Kerguelen region (Southern Ocean), *Limnol. Oceanogr. Methods*, *10*, 483–495, doi:10.4319/lom.2012.10.483.

© 2015. American Geophysical Union. All Rights Reserved.

2.2.3 Conclusion et perspectives de l'étude

La méthode FLAVOR, basée sur un réseau de neurones, permet de modéliser de manière relativement robuste la distribution verticale de la $[Chl]$ ainsi que de la composition de l'assemblage phytoplanctonique à partir de la seule forme d'un profil vertical *in situ* de fluorescence. Des résultats de **validation** très satisfaisant comparant la $[Chl]$ estimée par la méthode avec la $[Chl]$ de référence estimée par HPLC ont été présentés et discutés dans l'article présenté ci-dessus. La performance du réseau de neurones pourra très certainement être améliorée en utilisant les nouvelles données à venir, données correspondantes à des mesures concomitantes de fluorescence et de $[Chl]$ estimée par HPLC.

Pour la première fois et grâce à cette méthode, l'ensemble des bases de données de profils de fluorescence acquis depuis le début des années 70 va pourvoir être **harmonisé** et rendu **interopérable** en termes de $[Chl]$. De plus et à notre connaissance, c'est aussi la première fois qu'une méthode permet d'estimer la distribution verticale des **classes de taille phytoplanctoniques** à partir d'un profil de fluorescence. En utilisant FLAVOR, une base de données globale de fluorescence permettrait de générer, pour l'océan global, des **climatologies 3D** de la $[Chl]$ et des communautés phytoplanctoniques en termes de trois classes de taille. Comme nous l'avons évoqué en Introduction de ce chapitre (voir Sect. 2.1), ce type de climatologies est non seulement très important pour **initialiser et valider les modèles biogéochimiques** mais aussi pour servir de **référence** pour documenter un possible changement dans la distribution ainsi que dans l'évolution temporelle de la biomasse phytoplanctonique dans l'océan global.

2.3 Création d'une base de données de la distribution verticale de la composition des communautés phytoplanctoniques pour l'océan global

2.3.1 Résumé de l'étude

La **fluorescence** de la chlorophylle est une des propriétés biogéochimiques les plus mesurées dans l'océan mondial. Des **milliers de profils**, **hétérogènes** quant à leur acquisition, sont déjà disponibles

dans les bases de données historiques. L'intégration progressive de **capteurs de fluorescence** sur les plateformes autonomes de type flotteurs-profileurs ou planeurs sous-marins met à disposition une quantité **croissante** de profils de fluorescence, généralement utilisés dans le cadre d'analyses locales ou régionales. Cette importante source d'information que constitue les bases de données « historiques » et celles voyant désormais le jour, n'a toujours pas été intégrée et analysée dans le cadre **d'approches globales**. C'est en grande partie l'objet de l'étude présentée ici.

Au total, 268 127 profils de fluorescence provenant de **différentes bases de données** (aussi bien publiées que non publiées) ont été récupérés et regroupés pour former une nouvelle **base de données unique**. Après l'application d'un contrôle de qualité approprié (détaillé dans l'article ci-dessous), environ **49 000 profils de fluorescence** « qualifiés » ont servi de base à cette étude. Ils ont été convertis en profils de biomasse phytoplanctonique (i.e. concentration en chlorophylle *a*) ainsi qu'en profils d'indices de taille de la composition des communautés (i.e. micro-, nano- et pico-phytoplancton). La calibration de la fluorescence en $[Chl]$ et en **indice de tailles** a été faite en utilisant la méthode présentée précédemment (i.e. FLAVOR), qui a été spécifiquement développée pour harmoniser et « réconcilier » les profils de fluorescence provenant de sources diverses et hétérogènes, c'est-à-dire dans la majorité des cas, des profils de fluorescence non comparables directement.

Les données de $[Chl]$ et de la composition des communautés phytoplanctoniques associée constituant cette base de données s'étendent sur cinq décennies de 1958 à 2015. Cet ensemble de profils inclue des observations provenant de tous les bassins océaniques majeurs et de toutes les saisons (voir Figures 4 et 5 du papier ci-dessous). Les profondeurs d'acquisition sont comprises entre la surface et une profondeur maximum médiane de 700 m.

Grâce à cette base de données, des **cartes globales** de concentration en chlorophylle *a* et de la composition de l'assemblage phytoplanctonique associée ont été générées et sont présentées dans l'article qui suit (voir Figures 6 et 7 de l'article ci-dessous). Les **premières climatologies 4D** (mensuelles) sont également présentées pour trois des provinces écologiques définies par Longhurst (2010) (voir Figure 10 de l'article ci-dessous) ; Elles illustrent le **potentiel** de telles méthodes, notamment dans le cadre d'études de biogéographie et de phénologie. En effet, ces climatologies permettent d'avoir la dimension z correspondant à la profondeur, de manière **fine**, ce qui permet de mettre en avant certains phénomènes jusqu'ici « noyés » dans la discrétisation des données. En effet, dans les

climatologies calculées à partir des profils de fluorescence, nous montrons qu'il est possible de restituer par exemple le maximum profond de concentration en chlorophylle *a* (DCM pour « Deep Chlorophyll Maximum ») omniprésent pendant l'année dans les grands gyres subtropicaux (voir exemple sur la Figure 10-d dans le papier). Les données originales (i.e. profils de fluorescence bruts) ainsi que les profils calibrés en biomasse phytoplanctonique et communauté phytoplanctonique associée sont disponibles en libre accès à PANGAEA (Data Publisher for Earth and Environmental Science).

2.3.2 Article publié dans *Earth System Science Data* : Vertical distribution of chlorophyll *a* concentration and phytoplankton community composition from *in situ* fluorescence profiles : a first database for the global ocean

Earth Syst. Sci. Data, 7, 261–273, 2015
www.earth-syst-sci-data.net/7/261/2015/
doi:10.5194/essd-7-261-2015
© Author(s) 2015. CC Attribution 3.0 License.

Open Access Earth System Science Data

Vertical distribution of chlorophyll *a* concentration and phytoplankton community composition from in situ fluorescence profiles: a first database for the global ocean

R. Sauzède[1,2], H. Lavigne[3], H. Claustre[1,2], J. Uitz[1,2], C. Schmechtig[1,2], F. D'Ortenzio[1,2], C. Guinet[4], and S. Pesant[5,6]

[1]Laboratoire d'Océanographie de Villefranche, CNRS, UMR7093, Villefranche-Sur-Mer, France
[2]Université Pierre et Marie Curie-Paris 6, UMR7093, Laboratoire d'océanographie de Villefranche, Villefranche-Sur-Mer, France
[3]Istituto Nazionale di Oceanografia e di Geofisica Sperimentale, Sgonico (OGS), Italy
[4]Centre d'Etudes Biologiques de Chizé, CNRS, Villiers en Bois, France
[5]MARUM, Center for Marine Environmental Sciences, Universität Bremen, Bremen, Germany
[6]PANGAEA, Data Publisher for Earth and Environmental Science, Bremen, Germany

Correspondence to: R. Sauzède (sauzede@obs-vlfr.fr)

Received: 29 March 2015 – Published in Earth Syst. Sci. Data Discuss.: 21 April 2015
Revised: 31 August 2015 – Accepted: 18 September 2015 – Published: 5 October 2015

Abstract. In vivo chlorophyll *a* fluorescence is a proxy of chlorophyll *a* concentration, and is one of the most frequently measured biogeochemical properties in the ocean. Thousands of profiles are available from historical databases and the integration of fluorescence sensors to autonomous platforms has led to a significant increase of chlorophyll fluorescence profile acquisition. To our knowledge, this important source of environmental data has not yet been included in global analyses. A total of 268 127 chlorophyll fluorescence profiles from several databases as well as published and unpublished individual sources were compiled. Following a robust quality control procedure detailed in the present paper, about 49 000 chlorophyll fluorescence profiles were converted into phytoplankton biomass (i.e., chlorophyll *a* concentration) and size-based community composition (i.e., microphytoplankton, nanophytoplankton and picophytoplankton), using a method specifically developed to harmonize fluorescence profiles from diverse sources. The data span over 5 decades from 1958 to 2015, including observations from all major oceanic basins and all seasons, and depths ranging from the surface to a median maximum sampling depth of around 700 m. Global maps of chlorophyll *a* concentration and phytoplankton community composition are presented here for the first time. Monthly climatologies were computed for three of Longhurst's ecological provinces in order to exemplify the potential use of the data product. Original data sets (raw fluorescence profiles) as well as calibrated profiles of phytoplankton biomass and community composition are available on open access at PANGAEA, Data Publisher for Earth and Environmental Science.

Raw fluorescence profiles: http://doi.pangaea.de/10.1594/PANGAEA.844212 and
Phytoplankton biomass and community composition: http://doi.pangaea.de/10.1594/PANGAEA.844485

Published by Copernicus Publications.

1 Introduction

Phytoplankton biomass is generally recognized to play a key role in the global carbon cycle, stressing the need for a better understanding of its spatio-temporal distribution and variability in the global ocean. Chlorophyll *a* concentration is widely used as a proxy to estimate phytoplankton biomass. The geographic and temporal distribution of this proxy is already well documented at a global scale thanks to synoptic remote sensing observations by ocean-color radiometry (OCR, McClain, 2009; Siegel et al., 2013). Nevertheless, OCR observations are restricted to the ocean surface layer, "sensing" only one-fifth of the so-called euphotic layer where phytoplankton photosynthesis is realized and which can sometimes extend to well below 100 m (Gordon and McCluney, 1975; Morel and Berthon, 1989). It is therefore essential to better resolve the global distribution of phytoplankton biomass in the vertical.

The vertical distribution of chlorophyll *a* can be estimated with greatest accuracy from the analysis of water samples by high-performance liquid chromatography (HPLC, Claustre et al., 2004; Peloquin et al., 2013). However, these in situ measurements are relatively scarce because their acquisition requires ship-based sampling and their analysis is costly. Moreover, because these measurements are made on water samples, the vertical resolution is generally weak (e.g., around one measurement every 10 m). The measurement of in vivo chlorophyll *a* fluorescence is widely used as a proxy for chlorophyll *a* concentration (Lorenzen, 1966). Besides dissolved oxygen concentration, fluorescence is the most measured biogeochemical property in the global ocean. The advantages of this method are as follows: (1) it can be easily measured in situ using reliable sensors; (2) the vertical resolution is high, yielding several values per meter; and (3) data are available in digital format immediately after their acquisition. The integration of fluorescence sensors on autonomous platforms (e.g., profiling floats, animals, gliders) has recently led to a sudden rise in the acquisition of in vivo chlorophyll *a* fluorescence data (Claustre et al., 2010a). However, the relationship between chlorophyll *a* fluorescence and phytoplankton biomass is highly variable and depends on several factors, including phytoplankton physiological state and community composition (Cunningham, 1996; Falkowski et al., 1985; Kiefer, 1973). The conversion of in situ chlorophyll *a* fluorescence measurements into phytoplankton biomass must therefore be done with great care.

FLAVOR (Fluorescence to Algal communities Vertical distribution in the Oceanic Realm) is a method developed to transform and combine large numbers of fluorescence profiles from various sampling sensors and platforms (Sauzède et al., 2015a). This neural network-based method generates vertical distributions of (1) chlorophyll *a* concentration and (2) phytoplankton community size indices (i.e., microphytoplankton, nanophytoplankton and picophytoplankton) based on the shape of in situ fluorescence profiles (i.e., normalized profiles) and the day and location of acquisition. In addition to chlorophyll *a* concentration, community composition is an essential variable that determines the possible impact of phytoplankton on oceanic carbon fluxes and climate change scenarios (e.g., Le Quere et al., 2005). Global data compilations of phytoplankton community composition from discrete water samples have recently been published in ESSD (Peloquin et al., 2013) but data remain rather sparse. It could be an invaluable source of information to have a database of phytoplankton community size indices with the same spatio-temporal resolution as the fluorescence data sets. It has now become possible using the FLAVOR method to transform and combine all available in situ fluorescence data into a single-reference database that comprises essential information on chlorophyll *a* concentration and phytoplankton community size indices vertical distributions.

Presently, the widely used climatology of the global vertical distribution of chlorophyll *a* concentration is published in the World Ocean Atlas 2001 (Conkright et al., 2002). The latter climatology is based on estimates from analyzed water samples available in the World Ocean Database (WOD, Levitus et al., 2013) and the World Data Center (WDC, http://gcmd.gsfc.nasa.gov/). This climatology, based on seven discrete depths (0-10-20-30-50-75-100 m), is mainly limited by the lack of in situ estimations of chlorophyll *a* concentration, which leads to a strong spatial interpolation of data. Moreover, the discrete depths used to compute the climatology fail to finely reproduce the vertical distribution of the phytoplankton biomass, especially in areas characterized by very deep (> 100 m) deep chlorophyll maxima (DCM) such as the core of subtropical oligotrophic gyres. Using FLAVOR, the potential of the high vertical (around one data point per meter) and spatial resolution of chlorophyll fluorescence measurements would improve the 3-D climatologies of chlorophyll *a* concentration significantly. Moreover, climatologies of phytoplankton community size indices could be created with a similar spatio-temporal resolution.

This paper presents a global compilation of chlorophyll fluorescence profiles obtained from online databases and from published and unpublished individual sources. These were converted into a global compilation of phytoplankton biomass (i.e., chlorophyll *a* concentration) and community composition using the FLAVOR method. Prior to the application of FLAVOR, a 10-step quality control procedure was specifically developed. The remaining profiles were then analyzed. As examples of application, we present the first maps of global mean chlorophyll *a* concentration for several oceanic layers as well as global maps of phytoplankton community size indices. To further assess the quality of the resulting database, the climatological chlorophyll *a* concentration computed here for the surface layer is compared to the climatological remotely sensed chlorophyll *a* concentration available from Modis Aqua. Moreover, monthly 3-D climatologies of chlorophyll *a* concentration and associated phytoplankton community size indices are analyzed for

Table 1. Summary of the contributions of the chlorophyll fluorescence profiles in the database presented in this study.

Data source/institute/investigator	Period	Number of fluorescence profiles	Percentage of data in the database	Website if available or contact for requests
National Oceanographic Data Center (NODC)	Jun 1958–Mar 2014	30 977	63.7 %	http://www.nodc.noaa.gov/
Oceanographic Autonomous Observations (OAO)	May 2008–Jan 2015	6092	12.5 %	http://www.oao.obs-vlfr.fr/
Laboratoire d'Océanographie de Villefranche (LOV) cruises	May 1991–Jan 2012	3320	6.8 %	claustre@obs-vlfr.fr, sauzede@obs-vlfr.fr
Japan Oceanographic Data Center (JODC)	Jan 1998–Jul 2004	2262	4.6 %	http://www.jodc.go.jp/
PANGAEA	Nov 1980–Apr 2009	2294	4.7 %	http://www.pangaea.de/
C. Guinet (data acquired by elephant seals, Guinet et al., 2013)	Dec 2007–Jan 2011	1908	3.9 %	christophe.guinet@cebc.cnrs.fr,
British Oceanographic Data Center (BODC)	Sep 1996–Nov 2008	1219	2.5 %	http://www.bodc.ac.uk/
Systèmes d'Informations Scientifiques pour la MER (SISMER)	Sep 1999–May 2008	237	0.5 %	http://www.ifremer.fr/sismer/
Australian Antarctic Data Center (AADC)	Jan 2001–Feb 2006	234	0.5 %	http://data.aad.gov.au/
Southern Ocean Iron RElease Experiment (SOIREE)	Feb 1999	57	0.1 %	http://www.uea.ac.uk/~e610/soiree/index.html

several ecological provinces defined by Longhurst (2010). Overall, the data set presented here can be readily exploited to deepen our understanding of the spatio-temporal distribution and variability of phytoplankton biomass and associated community composition in the global ocean. It is obviously a first step towards a database that will regularly be improved thanks to the ongoing intensification of chlorophyll *a* fluorescence profile acquisition by Bio-Argo profiling floats, gliders and mammals equipped with instruments.

2 Data and methods

2.1 Origins of in situ chlorophyll fluorescence measurements

The database presented in this study is available from PANGAEA, Data Publisher for Earth and Environmental Science in two formats: (1) the database containing all compiled raw fluorescence profiles (the raw database, http://doi.pangaea.de/10.1594/PANGAEA.844212, Sauzède et al., 2015b) and (2) the database containing the fluorescence profiles which are calibrated into chlorophyll *a* concentration and associated phytoplankton community size indices (the calibrated database, http://doi.pangaea.de/10.1594/PANGAEA.844485, Sauzède et al., 2015c). The data of in situ vertical fluorescence profiles compiled for creating the raw database were obtained from several available online databases as well as published and unpublished individual sources. The duplicates and single-surface values, which are not vertical profiles, were automatically removed (not integrated in the raw database). Finally, the raw database contains 268 127 fluorescence profiles. Following a robust quality control procedure detailed hereafter (Sect. 2.2), about 49 000 chlorophyll fluorescence profiles were converted into phytoplankton biomass (i.e., chlorophyll *a* concentration) and size-based community composition (i.e., microphytoplankton, nanophytoplankton and picophytoplankton). The origin of this calibrated database is summarized in Table 1. The majority of the data come from the National Oceanographic Data Center (NODC) and the fluorescence profiles acquired by Bio-Argo floats are available on the Oceanographic Autonomous Observations (OAO) web platform (63.7 and

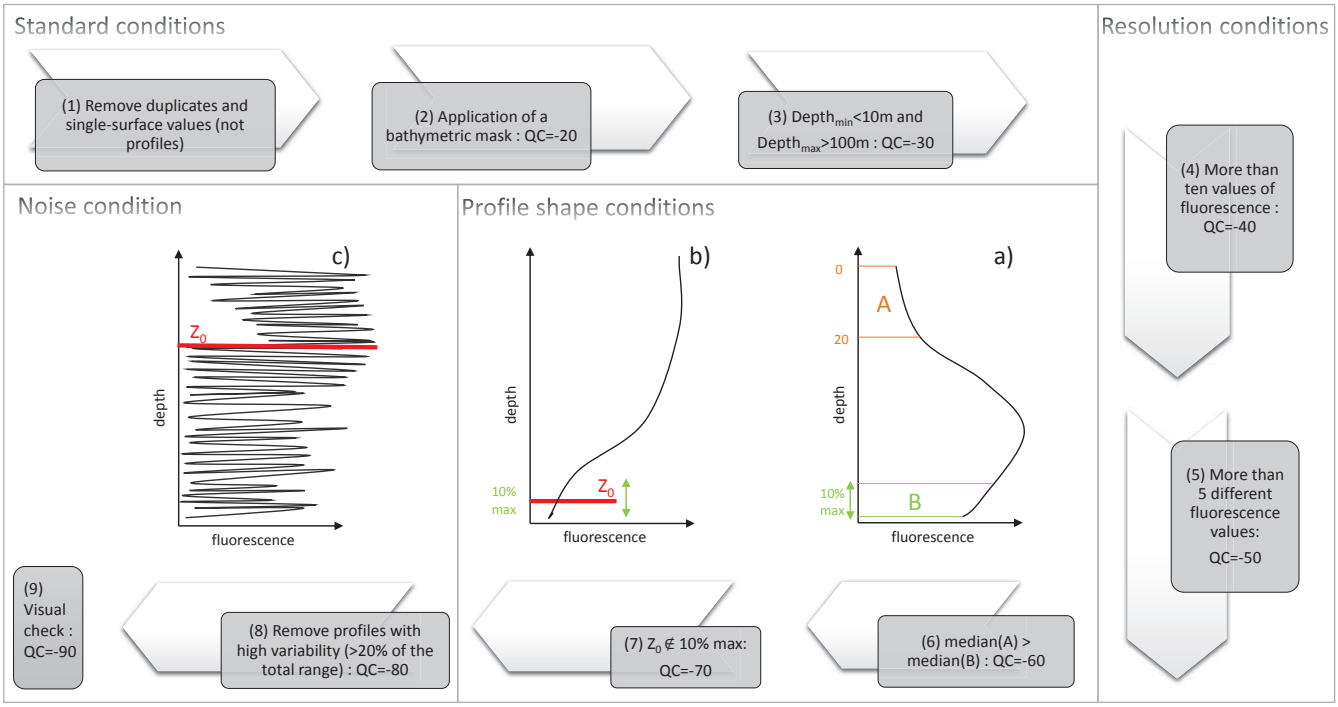


Figure 1. Schematic overview of the quality control procedure specifically developed for the database presented in this study. The fluorescence profiles represented in the **(a)**, **(b)** and **(c)** panels are examples of profiles which are rejected by the quality control steps (6), (7) and (8) respectively.

12.5 % respectively, see percentages of data in the database depending on their origin in Table 1).

Different modes of acquisition were used to collect the data presented in this study: (1) the CTD (conductivity, temperature and depth) profiles are acquired using a fluorometer mounted on a CTD rosette; (2) the OSD (Ocean Station Data) profiles are derived from water samples analyzed by fluorometry and are defined as "low" resolution profiles (Boyer et al., 2009); (3) the UOR (Undulating Oceanographic Recorder) profiles are acquired by a "fish" equipped with fluorometer and towed by a research vessel; (4) AP (Autonomous Platforms) profiles are acquired by Bio-Argo profiling floats or elephant seals equipped with a fluorometer (Claustre et al., 2010b; Guinet et al., 2013). Table 2 lists the number of profiles in the calibrated database according to these four modes of acquisition.

It is worth noting that the data acquired from gliders were not included in the database. Although glider data are extremely numerous, they are restricted to a very small spatio-temporal window. As a consequence, a database including glider data would likely be spatially and temporally biased, in contradiction with our first aim of building a global climatological database.

Table 2. Summary of the chlorophyll fluorescence profiles in the database presented in the study depending on the different modes of data acquisition.

Acquisition	Number of fluorescence profiles	Percentage of data in the database
CTD	27 433	56.4 %
OSD	10 831	22.3 %
UOR	2952	6 %
AP	7384	15.2 %
Total	48 600	

2.2 Quality control

In order to use the FLAVOR method (see details in Sect. 2.3), a specific and adapted data quality control procedure was developed and applied to each in situ chlorophyll fluorescence profile. This procedure was schematically implemented according to four main steps of data control (Fig. 1), each step being developed for discarding most, if not all, spurious fluorescence profiles that would deteriorate the quality of the database. Firstly, several basic tests were applied: (1) duplicates and single-surface values, which are not vertical profiles, were removed (these profiles were removed from the beginning of the process so they are not included in the so-called raw database); (2) coastal profiles were removed using

Table 3. Summary of the number of fluorescence profiles rejected at each step of quality control.

QC step number (see Fig. 1)	Number of fluorescence profiles deleted	% of data deleted
2	162 609	74 %
3	31 904	14.5 %
4	15 396	7 %
5	286	0.1 %
6	3569	1.6 %
7	2891	1.3 %
8	1597	0.7 %
9 – Visual check	244	0.1 %
Chauvenet's criterion and range criterion after calibration (see Sect. 2.3)	1031	0.5 %

a bathymetric mask of 500 m depth; (3) the uppermost measurement has to be located within the 0–10 m layer, while the deepest measurement has to be at or below 100 m. Secondly, tests on the profile vertical resolution are applied: (4) a minimum of 10 values per profile is required (i.e., condition on the vertical resolution acquisition); (5) a minimum of five non-equal values per profile are required (i.e., condition on the sensor resolution). Then, several tests are applied on the fluorescence profile shape. These conditions are based on the parameter used for the development of the FLAVOR method, Z_0, which is the depth at which the fluorescence profile returns to a constant background value (see details in Sect. 2.3 and examples in Fig. 1b and c). (6) The median of the fluorescence values from the surface down to 20 m has to be greater than the median of the values of the last 10 % of the deepest samples of the profile (see Fig. 1a); (7) the depth Z_0 has not to be within the last 10 % of the deepest samples of the profile (see Fig. 1b). Finally, a test on the noise of the profiles was developed and applied: (8) profiles with aberrant data caused by electronic noise are removed (i.e., variability greater than 20 % of the total profile range, see Fig. 1c). To finish, a visual check allowed all the remaining fluorescence profiles to be verified. The number of raw fluorescence profiles rejected at each step of the quality control procedure is presented in Table 3. Around 80 % of the raw fluorescence profiles were thus removed by this procedure. This step is an essential prerequisite for the development of a "clean" database of vertical distributions of phytoplankton biomass and community composition in the global ocean. The quality control procedure removed 77, 71, 28 and 25 % of the OSD, UOR, AP and CTD profiles, respectively, with profiles removed by the test on the bathymetry not taken into account.

2.3 Conversion of chlorophyll fluorescence into chlorophyll *a* concentration and phytoplankton community composition

In order to assess the vertical distribution of the total chlorophyll *a* concentration (hereafter, [TChl]) and the chlorophyll *a* concentration associated to each phytoplankton size index (hereafter, [microChl], [nanoChl] and [picoChl] for microphytoplankton, nanophytoplankton and picophytoplankton respectively), the FLAVOR method (Sauzède et al., 2015a) is applied to each chlorophyll fluorescence profile, satisfying the quality control procedure (see Sect. 2.2). In summary, FLAVOR is a neural network-based method which uses (1) the shape of the chlorophyll fluorescence profile (10 values from the normalized profile with values range between 0 and 1); (2) the depth Z_0, which is the depth at which the fluorescence profile returns to a constant background value (see examples of Z_0 depths represented by the horizontal red line for two profiles on Fig. 1b and c); and (3) the location (latitude and longitude) and the day of acquisition of the fluorescence profile as inputs. The outputs of FLAVOR are the vertical distributions of (1) [TChl] and (2) [microChl], [nanoChl] and [picoChl] with the same vertical resolution as the input raw fluorescence profile. FLAVOR is composed of two different neural networks: the first one was adapted to retrieve the vertical distribution of [TChl] and the second one to retrieve the vertical distributions of [microChl], [nanoChl] and [picoChl] simultaneously. Both neural networks were adapted and validated using a large database including 896 concomitant in situ vertical profiles of HPLC pigments and chlorophyll fluorescence. These profiles were collected as part of 22 oceanographic cruises representative of the global ocean in terms of trophic and oceanographic conditions, making the method applicable to most oceanic waters. The diagnostic pigment-based approach of Uitz et al. (2006), based on Claustre (1994) and Vidussi et al. (2001), was utilized to estimate the biomass associated with the three pigment-derived size classes for each profile. Finally, the data set of concurrent fluorescence profiles and HPLC-determined [TChl], [microChl], [nanoChl] and [picoChl] at discrete depths was used to establish the neural network-based relationships between the fluorescence profile shape and the vertical distributions of [TChl] and phytoplankton community. The schematic overview of the FLAVOR method is shown on Fig. 4 in Sauzède et al. (2015a). The global absolute errors of FLAVOR retrievals are 40, 46, 35 and 40 % for the [TChl], [microChl], [nanoChl] and [picoChl], respectively (Sauzède et al., 2015a).

Admittedly, the FLAVOR method has some limitations. The dependence of chlorophyll fluorescence on the light environment is probably intrinsically accounted for in the algorithm thanks to the geolocation and date of acquisition used as inputs for the training. However, one of the potential concerns with FLAVOR is that the impact of the daytime non-photochemical quenching (NPQ; see, e.g., Cullen

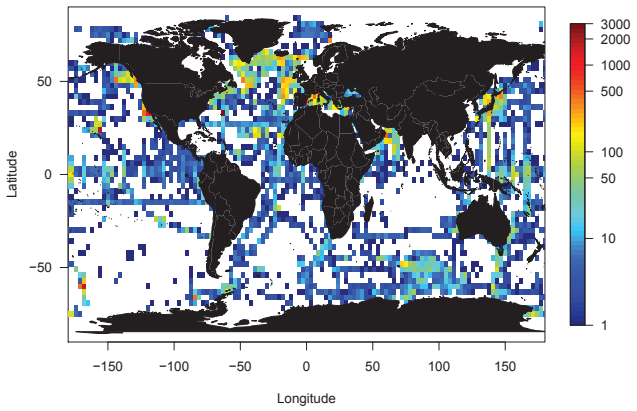


Figure 2. Geographic distribution of the 48 600 chlorophyll fluorescence profiles in the database that passed through all the steps of the quality control procedure. The color scale indicates the number of fluorescence profiles in boxes of 3° per 3°.

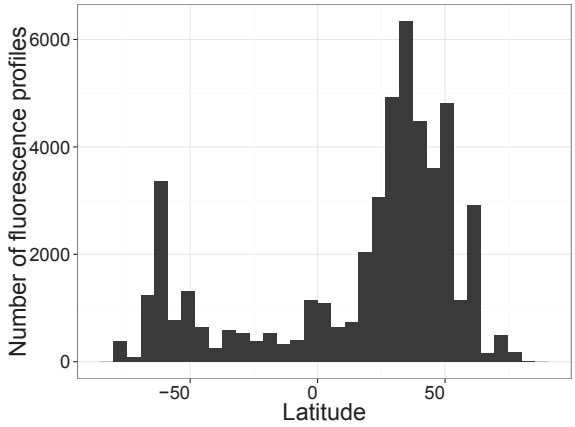


Figure 3. Frequency distribution of the 48 600 profiles of chlorophyll *a* concentration and associated phytoplankton community composition in the database as a function of latitude.

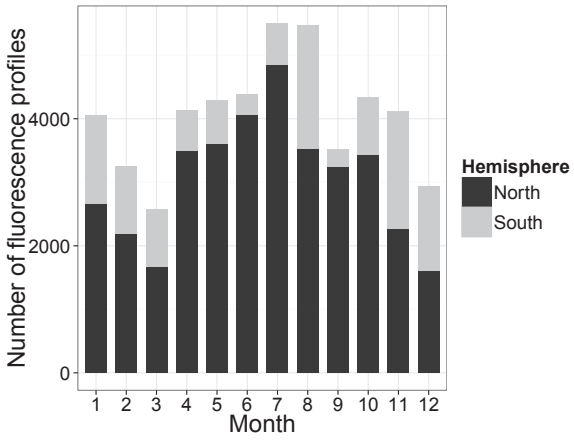


Figure 4. Temporal distribution of the 48 600 profiles of chlorophyll *a* concentration and associated phytoplankton community composition in the database as a function of months with black and gray colors, indicating the hemispheres of data acquisition.

and Lewis, 1995), responsible for a decrease in chlorophyll fluorescence values at high irradiance, is not accounted for by the method. The NPQ uncorrected fluorescence profile shape is indeed used to retrieve the vertical distribution of phytoplankton biomass (see details in Sauzède et al., 2015a). Note that, if density profiles are available together with fluorescence profiles, NPQ can be corrected using the method of Xing et al. (2012). This method involves substituting the fluorescence values acquired within the mixed layer by the maximum value within this layer.

It has been previously mentioned that FLAVOR is not adapted for the retrieval of chlorophyll *a* concentration on a fluorescence profile-by-profile basis (Sauzède et al., 2015a). Rather, FLAVOR and, hence, the resulting database, are relevant for large-scale investigations, e.g., development of climatologies of the vertical distribution of chlorophyll *a*, from which regional anomalies or temporal trends might be evidenced. In fact, the method was validated using a global database and it is not excluded that the retrievals from FLAVOR might be regionally biased. For instance, Sauzède et al. (2015a) have shown that FLAVOR retrievals for the Southern, Arctic and Indian oceans are slightly less accurate than for the other basins. This is likely because the method is not constrained enough in these specific areas which are known for data scarcity. Additional details about the performance of the method for various oceanic basins are given in Sauzède et al. (2015a), in Figs. S3, S5–S7. Finally, it is worth recalling here that the relationships between the phytoplankton biomass or community composition profiles and the fluorescence profiles are assumed to be identical for profiles acquired before 1991 (not involved in the training data set because of lack of HPLC data) and after 1991 (only used for the training process). In the context of possible use of this database for supporting analysis in looking for trends or a shift in chlorophyll *a* time series, this assumption will have to be taken into consideration.

An additional step of quality control is further applied once the FLAVOR method has been operated. It is based on Chauvenet's criterion which is used to identify statistical outliers in the retrieved biomass data (Buitenhuis et al., 2013; Glover et al., 2011; O'Brien et al., 2013). The criterion was applied to the surface data of each profile (median of values from the surface down to 20 m). As Chauvenet's criterion is based on the assumption that the data follow a normal distribution, the analysis was performed on the log-normalized [TChl] surface values. Such a criterion removes aberrant data partially caused by the failure of the FLAVOR method (see number of profiles removed by Chauvenet's criterion in Table 3).

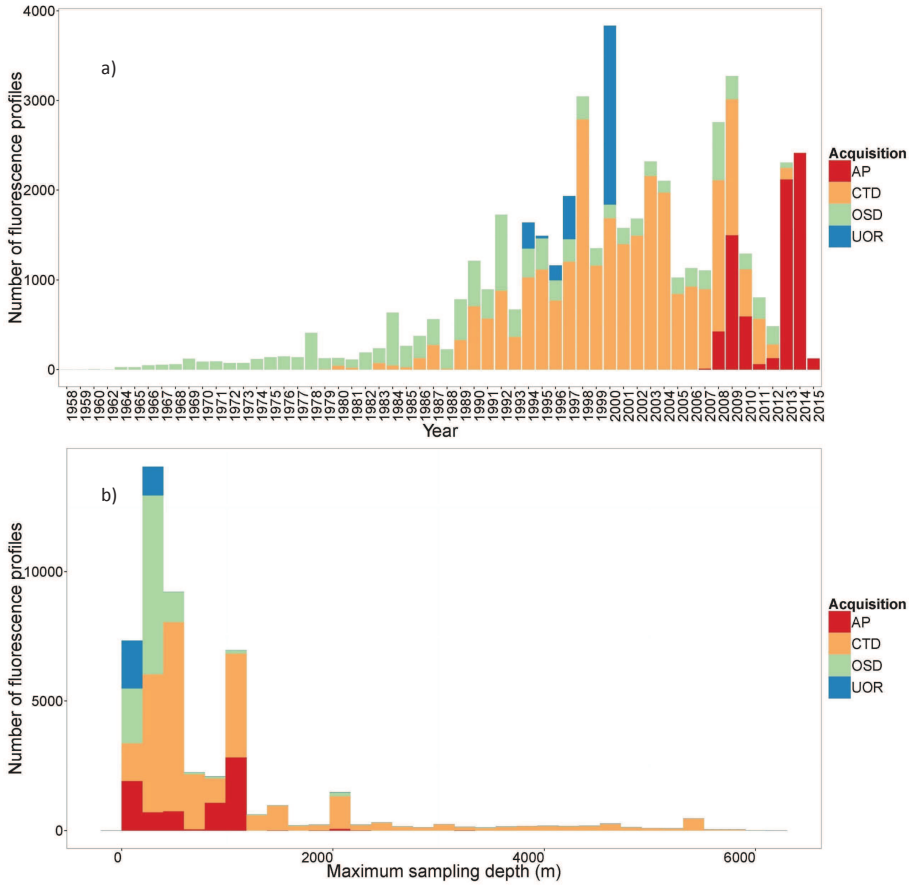


Figure 5. Frequency distribution of the 48 600 profiles of chlorophyll *a* concentration and associated phytoplankton community composition in the database as a function of: **(a)** years of acquisition and **(b)** the maximum depth of acquisition. Colors refer to the different modes of data acquisition.

3 Results and discussion

3.1 Spatial and temporal coverage of the database

The 48 600 chlorophyll fluorescence profiles which successfully passed all the steps of quality control were transformed into total chlorophyll *a* concentration and associated phytoplankton community size indices (i.e., microphytoplankton, nanophytoplankton and picophytoplankton) using FLAVOR (see details in Sect. 2.3). The resulting database covers all ocean basins with more profiles in the Northern Hemisphere (75 %) than in the Southern Hemisphere (25 %, see Figs. 2 and 3). However, the Southern Hemisphere remains relatively well represented with the profiles acquired by autonomous platforms and especially by elephant seals equipped with a fluorometer. Few data were acquired in the Indian Ocean and in the tropical South Atlantic and South Pacific (see Fig. 2). The highest numbers of fluorescence profiles are found at the BATS (the Bermuda Atlantic Time-series Study) and HOT (the Hawaii Ocean Time-series) time-series stations, which are located at 31.67° N–64.17° W and 22.75° N–158.00° W, respectively, and where data acquisition started in 1988. On the annual scale, the data acquisition appears evenly distributed, with a slight underrepresentation of autumn months (April to June) in the Southern Hemisphere (Fig. 4). The temporal distribution of fluorescence profiles in the database covers 56 years from 1958 to the present (Fig. 5a) and most of the observations were collected after the late 1980s. There are fewer observations from 2010 to 2012 because all data generally acquired by ship-based platforms have not been archived yet in the online databases. A significant increase in data density observed between 2013 and 2015 (in 2015, 124 profiles were acquired in half a month) mainly results from data acquired by Bio-Argo profiling floats. Around one-sixth of this global database has been sampled in only 2 years by the Bio-Argo platforms. This illustrates the potential of this new type of acquisition which is expected to dramatically increase the number of collected fluorescence profiles in the future.

Vertically, the database includes values of total chlorophyll *a* concentration and associated phytoplankton community composition from the surface down to a mean sampling

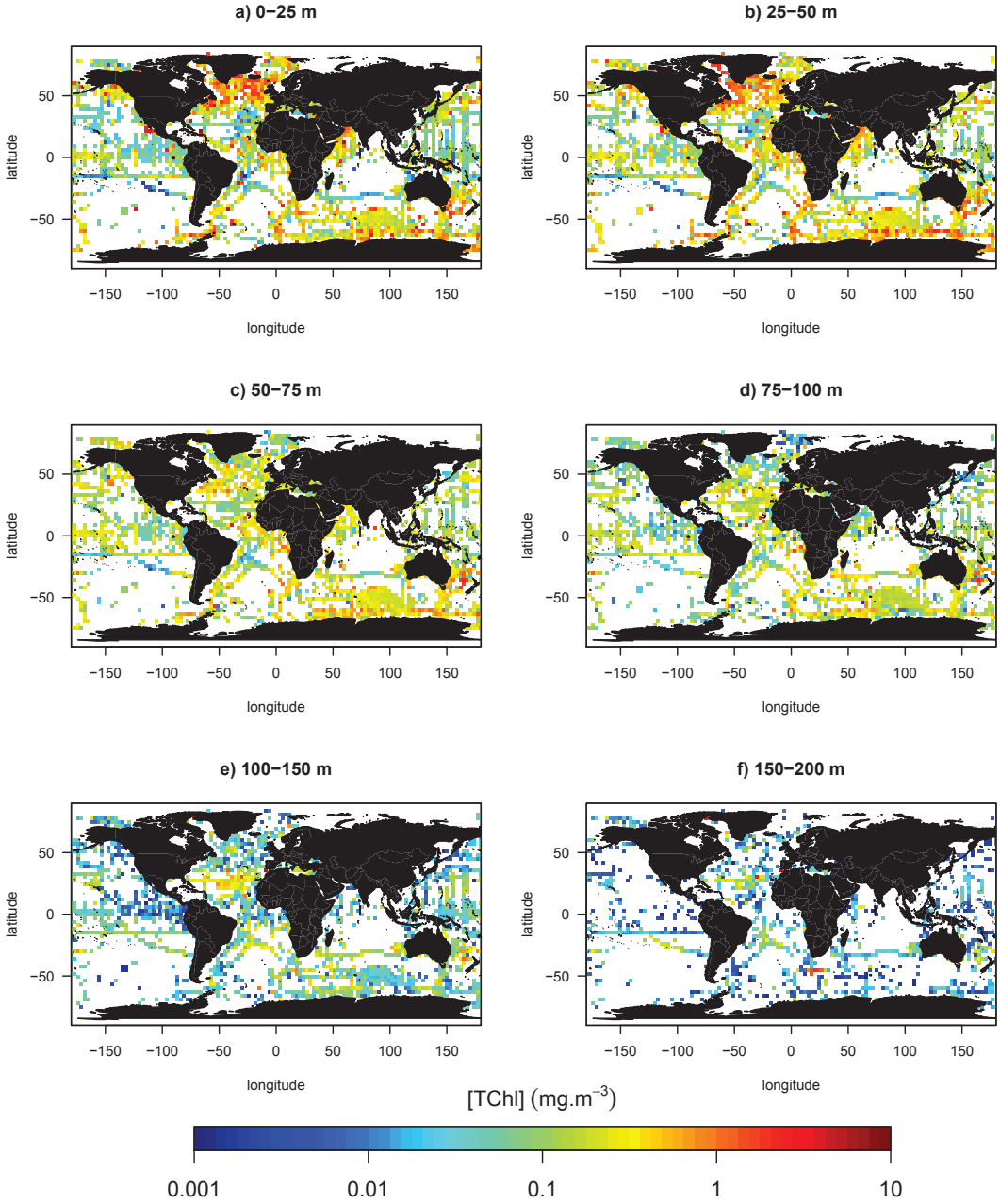


Figure 6. Median total chlorophyll *a* concentration (mg m^{-3}) scaled to a 3° spatial resolution for six vertical layers: **(a)** 0–25 m, **(b)** 25–50 m, **(c)** 50–75 m, **(d)** 75–100 m, **(e)** 100–150 m and **(f)** 150–200 m.

depth of 743 m (with a maximum sampling depth ranging from 100 to 6000 m; Fig. 5b).

3.2 Vertical distribution of the chlorophyll biomass

We present the database with respect to the vertical distribution of the total chlorophyll *a* concentration ([TChl]). Figure 6 displays the median [TChl] gridded within squares of 3° latitude by 3° longitude and over six vertical layers (0–25, 25–50, 50–75, 75–100, 100–150 and 150–200 m). In the surface layer (0–25 m, see Fig. 6a), the [TChl] median is the highest in the North Atlantic and the lowest in the South Pacific subtropical gyre. The median [TChl] decreases with depth for all the data, except for data acquired in South Pacific and Atlantic subtropical gyres where the median [TChl] increases with depth. This increase is associated with the so-called deep chlorophyll maximum (DCM) that is typical of these oligotrophic regions (e.g., Cullen, 1982; Mignot et al., 2011, 2014).

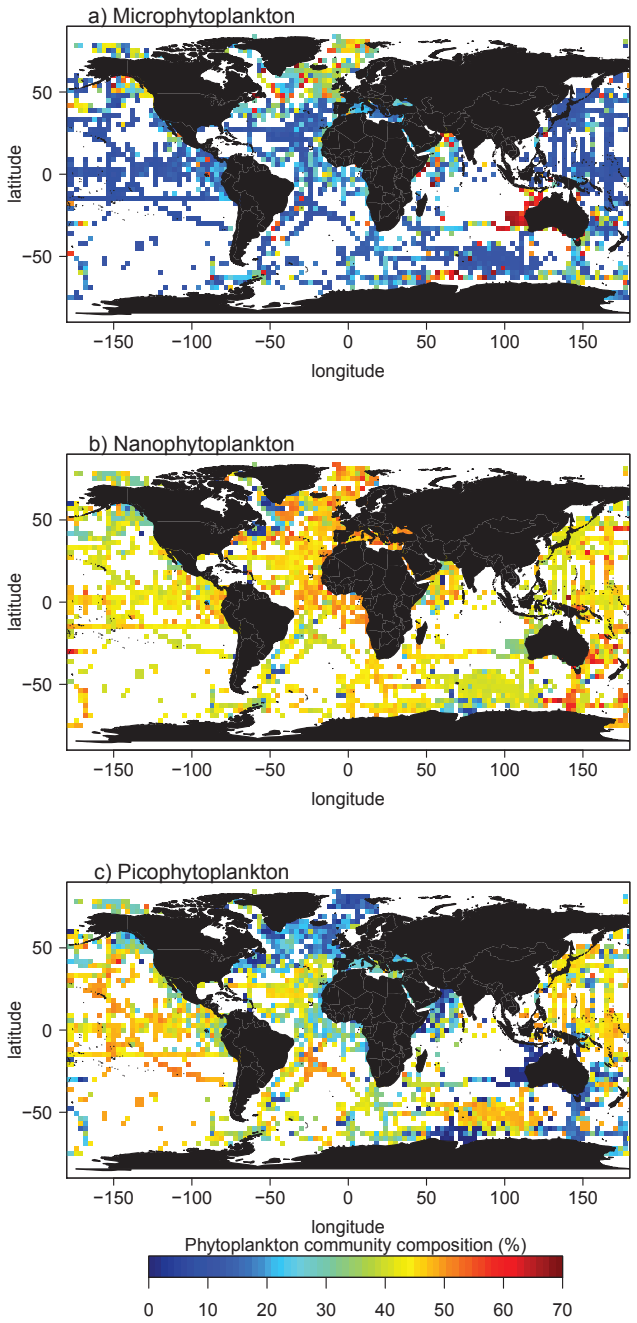


Figure 7. Mean relative contribution to the total chlorophyll *a* biomass (%) for the three phytoplankton size-based groups gridded and scaled to a 3° resolution within the 0–1.5 Z_e layer: **(a)** microphytoplankton, **(b)** nanophytoplankton and **(c)** picophytoplankton.

The global distribution of the phytoplankton community composition, given in terms of fraction of chlorophyll *a* concentration associated to micro-, nano- and picophytoplankton, is presented for the 0–1.5 Z_e layer (Fig. 7a, b and c respectively). Here Z_e, the euphotic depth is defined as the depth at which the irradiance is reduced to 1 % of its surface value. It was estimated according to the method of Morel and Berthon (1989), using the [TChl] profiles derived from FLAVOR. Figure 7 reveals general geographic patterns which are consistent with the knowledge about the ecological domains and biogeochemical provinces (Longhurst, 2010). On average microphytoplankton are dominant in the subarctic zone, with a relative contribution to the chlorophyll biomass reaching more than 70 % in these areas (Fig. 7a). Picophytoplankton are dominant in the subtropical gyres (South and North Pacific as well as South and North Atlantic), with a contribution reaching 45–55 % (Fig. 7c). Nanophytoplankton appear to be ubiquitous with a relatively stable contribution to biomass of 40–50 % (Fig. 7b).

To further assess the quality, range and representation of the FLAVOR-retrieved [TChl] database presented in this study, the retrieved surface [TChl] is compared to the remotely sensed [TChl]. In this context, the climatological [TChl] mean was extracted at a 9 km spatial resolution from NASA Modis Aqua archive for the time period covering 2002 to 2014. The extracted satellite [TChl] data were re-gridded to a 3° × 3° spatial resolution. Similarly the FLAVOR-retrieved [TChl] values for the upper layer of the database (i.e., mean value calculated between the surface and 20 m) for the same period were re-gridded to 3° × 3° squares. Figures 8 and 9 show that climatological averaged [TChl] from Modis Aqua and from the present database are generally consistent (Fig. 8a and b). The log-transformed ratio of the Modis Aqua to the database [TChl] estimates reveals a rather good agreement with a median value of −0.16 and a standard deviation of 0.58 (see histogram in Fig. 8c). Figure 9 displays the geographic distribution of the log-transformed ratio between the Modis Aqua and the database estimates of climatological surface [TChl]. The ratio shows no specific spatial bias. However, as it is mentioned in Sect. 2.3, FLAVOR retrievals for the Southern, Arctic and Indian oceans are slightly less accurate than for the other basins; it is therefore possible that the estimation errors are greater in these areas. Moreover, this observation has to be nuanced considering the difficulties in retrieving accurate ocean color satellite [TChl] in these high-latitude environments (Gregg and Casey, 2004; Guinet et al., 2013; Johnson et al., 2013; Peloquin et al., 2013; Siegel et al., 2005; Szeto et al., 2011).

3.3 Example of application: climatological time series of the vertical distribution of chlorophyll *a* concentration and phytoplankton community composition

As an example of application, monthly climatologies were computed for three ecological provinces defined by Longhurst (2010) and well represented in the current data set (Fig. 10a): (1) the North Atlantic Subtropical Gyral Province West (NASW, Fig. 10b), (2) the Atlantic Subarctic Province (SARC, Fig. 10c) and (3) the North Pacific Subtropical Gyre Province (NPTG, Fig. 10d). Overall the time series of the vertical distribution in [TChl] are consistent

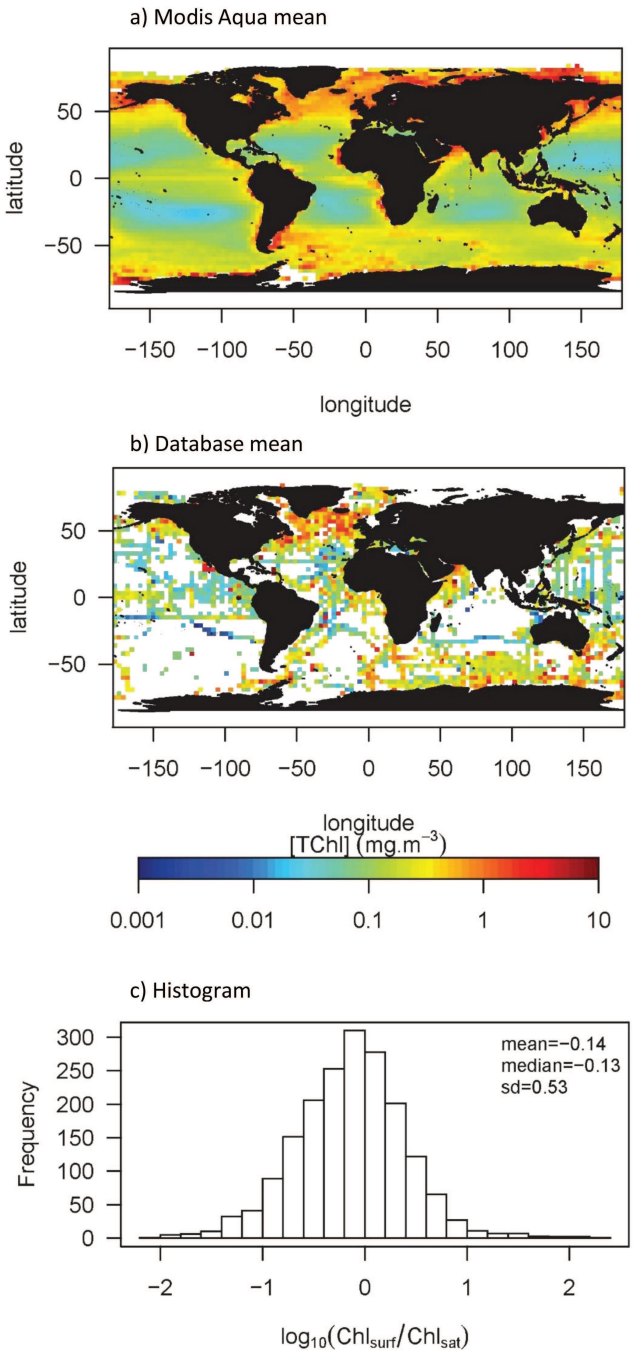


Figure 8. **(a)** Climatological mean (2002–2014) chlorophyll *a* concentration (mg m^{-3}) from Modis Aqua (scaled to a 3° resolution); **(b)** climatological mean (1958–2015) surface chlorophyll *a* concentration (mg m^{-3}) from the present database (averaged over the upper 20 m and scaled to a 3° resolution); **(c)** histogram of the log10 ratio of the chlorophyll *a* concentration from the database to the chlorophyll *a* concentration from Modis Aqua. The mean, median and standard deviation of the ratio are indicated in the figure. The color scale applies to panels **(a)** and **(b)**.

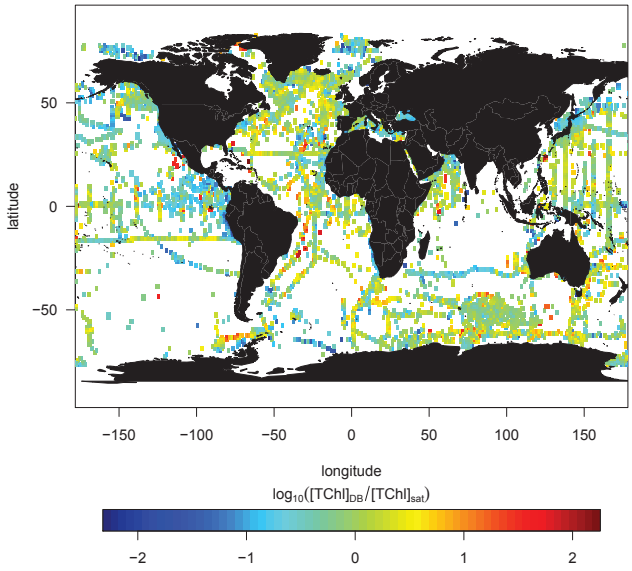


Figure 9. Geographic distribution of the log10-transformed ratio of the climatological mean surface [TChl] of the database over the upper 20 m of the water column ([TChl]$_{DB}$) and the climatological mean satellite [TChl] from Modis Aqua ([TChl]$_{sat}$). Both [TChl] data were scaled to a 3° resolution.

with expectations as detailed by Longhurst (2010). For the NASW province (Fig. 10b), [TChl] is relatively homogeneous from the surface to around 140 m from January to March; then the stratification of the water column leads to the establishment of a deep chlorophyll maximum (DCM) from April to November. Over the year, [TChl] varies in a restricted range of values (0.35–0.55 mg m^{-3}). The dominant phytoplankton groups are the nano- and the picophytoplankton with relative chlorophyll contribution reaching 40–45 % for both size-based groups. The contribution of microphytoplankton remains low throughout the year (10 %). For the SARC province, the phytoplankton bloom starts in May (as indicated by a significant increase in [TChl], Fig. 10c). The bloom continues for 4 to 5 months with [TChl] within the 1.5–2 mg m^{-3} range (with maximum values in July). The microphytoplankton contribution increases during the bloom and reaches a maximum (60 %) in August, whereas the nanophytoplankton relative contribution decreases from April to August. The contribution of picophytoplankton increases slightly all along the year to reach a maximum of about 40 % in December. For the NPTG province (Fig. 10d), a DCM (0.15–0.25 mg m^{-3}) is established at a depth of 100–125 m and persists all year long. This DCM deepens in summer, consistently with a deeper light penetration in the water column at this period. The [TChl] at DCM reaches a maximum value in June and July. The dominant phytoplankton groups are the nano- and the picophytoplankton with relative contribution reaching 45–50 % for both size-based groups and slight opposite temporal evolutions. The

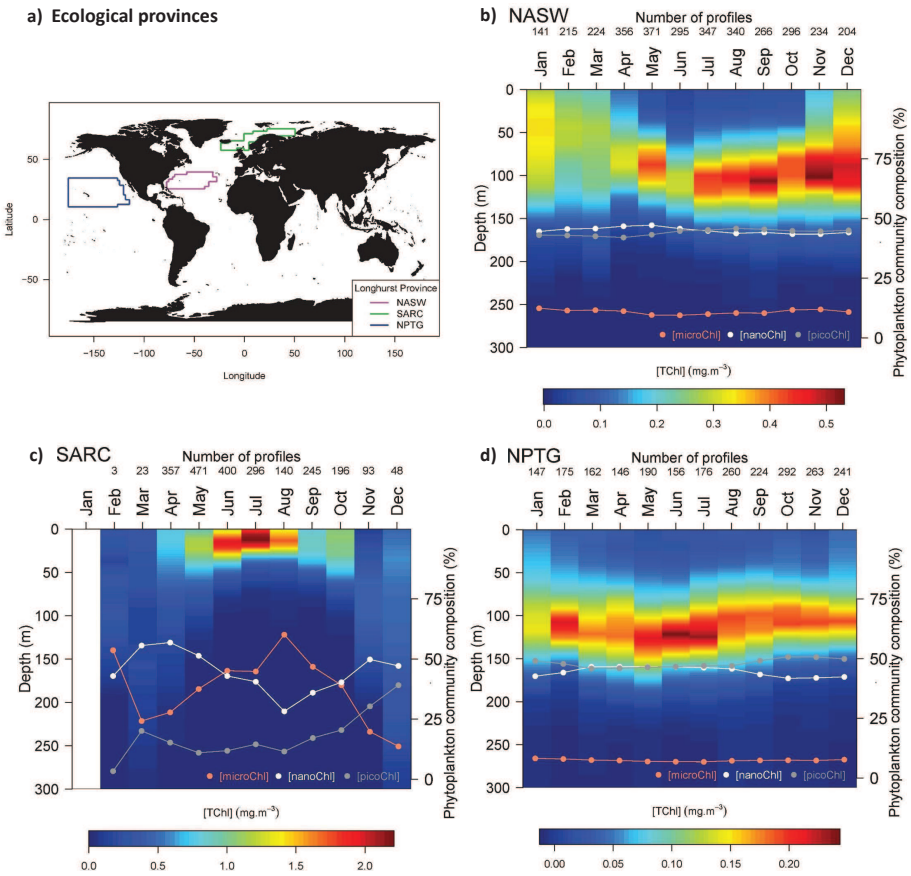


Figure 10. Monthly climatologies of the vertical distribution of the total chlorophyll *a* concentration and associated phytoplankton size-based groups for three ecological provinces defined by Longhurst (2010). **(a)** Geographic distribution of the considered provinces: North Atlantic Subtropical Gyral Province West (NASW), Atlantic Subarctic Province (SARC) and North Pacific Subtropical Gyre Province (NPTG). Climatologies obtained for the **(b)** NASW, **(c)** SARC and **(d)** NPTG. The color scale indicates the total chlorophyll *a* concentration ($mg\,m^{-3}$); the data points superimposed onto the colored monthly vertical profiles show the percentages of integrated chlorophyll *a* concentration associated with the micro-, nano- and picophytoplankton within the water column.

contribution of microphytoplankton remains low throughout the year ($< 10\,\%$).

4 Conclusions and recommendations for use

The phytoplankton biomass (i.e., chlorophyll *a* concentration) and phytoplankton community size indices were derived from chlorophyll fluorescence profiles using a dedicated calibration method (FLAVOR, Sauzède et al., 2015a). For the first time, in situ chlorophyll fluorescence profiles from various data centers have been collected and synthesized in a global data set to create unified and interoperable products related to chlorophyll *a* concentration and phytoplankton communities. This work can thus be considered as a first step towards the development of a 3-D climatological representation of chlorophyll *a* concentration and phytoplankton community composition. As mentioned before, we recall here that this database should not be used on a profile-by-profile basis. Instead, this database has rather to be used to derive climatologies from which regional or temporal trends might possibly be extracted. To date, and because of the lack of in situ vertical data, the identification of such trends has been based exclusively on surface remotely sensed data (Beaulieu et al., 2013; Boyce et al., 2010; Gregg, 2005; Gregg et al., 2002). Obviously, the present data set offers a potential refinement to improve open-ocean climatologies of chlorophyll *a* with respect to the vertical dimension.

Finally, this database has to be considered as a reference that has the potential to evolve. It is now clear that numerous fluorescence profiles will be acquired through robotic observations (e.g., Claustre et al., 2010b; Johnson et al., 2009). In fact, about one-sixth of the profiles of the present database have been sampled by Bio-Argo profiling floats in only 2 years. Therefore the database proposed here represents a first step towards a global single-reference database reconciling the oldest data sets of chlorophyll fluorescence with the future ones, mostly acquired remotely by autonomous platforms.

Acknowledgements. This paper is a contribution to the Remotely Sensed Biogeochemical Cycles in the Ocean (remOcean) project, funded by the European Research Council (grant agreement 246777), to the French Bio-Argo project funded by CNES-TOSCA and to the French "Equipement d'avenir" NAOS project (ANR J11R107-F). The French PROOF and CYBER programs are acknowledged for their support of cruises where in situ chlorophyll fluorescence profiles were acquired. We are grateful to all the project PIs who contributed data, as well as to the anonymous staff who took part in the data acquisition during the cruises.

Edited by: F. Schmitt

References

Beaulieu, C., Henson, S. A., Sarmiento, Jorge L., Dunne, J. P., Doney, S. C., Rykaczewski, R. R., and Bopp, L.: Factors challenging our ability to detect long-term trends in ocean chlorophyll, Biogeosciences, 10, 2711–2724, doi:10.5194/bg-10-2711-2013, 2013.

Boyce, D. G., Lewis, M. R., and Worm, B.: Global phytoplankton decline over the past century, Nature, 466, 591–596, doi:10.1038/nature09268, 2010.

Boyer, T. P., Antonov, J. I., Baranova, O. K., Garcia, H. E., Johnson, D. R., Locarnini, R. A., Mishonov, A. V, O'Brien, T. D., Seidov, D., Smolyar, I. V., and Zweng, M. M.: World Ocean Database 2009, edited by: Levitus, S., 2009.

Buitenhuis, E. T., Vogt, M., Moriarty, R., Bednaršek, N., Doney, S. C., Leblanc, K., Le Quéré, C., Luo, Y.-W., O'Brien, C., O'Brien, T., Peloquin, J., Schiebel, R., and Swan, C.: MAREDAT: towards a world atlas of MARine Ecosystem DATa, Earth Syst. Sci. Data, 5, 227–239, doi:10.5194/essd-5-227-2013, 2013.

Claustre, H.: The trophic status of various oceanic provinces as revealed by phytoplankton pigment signatures, Limnol. Oceanogr., 39, 1206–1210, 1994.

Claustre, H., Hooker, S. B., Van Heukelem, L., Berthon, J.-F., Barlow, R., Ras, J., Sessions, H., Targa, C., Thomas, C. S., van der Linde, D., and Marty, J.-C.: An intercomparison of HPLC phytoplankton pigment methods using in situ samples: application to remote sensing and database activities, Mar. Chem., 85, 41–61, doi:10.1016/j.marchem.2003.09.002, 2004.

Claustre, H., Antoine, D., Boehme, L., Boss, E., D'Ortenzio, F., D'Andon, O. F., Guinet, C., Gruber, N., Handegard, N. O., Hood, M., Johnson, K., Körtzinger, A., Lampitt, R., LeTraon, P. Y., Lequéré, C., Lewis, M., Perry, M. J., Platt, T., Roemmich, D., Testor, P., Sathyendranth, S., Send, U., and Yoder, J.: Guidelines towards an integrated ocean observation system for ecosystems and biogeochemical cycles, in: Proceedings of the OceanObs 09: Sustained Ocean Observations and Information for Society Conference (Vol.1), edited by: Hall, J., Harrison, D. E., and Stammer, D., ESA Publ., Venice, Italy, 2010a.

Claustre, H., Bishop, J., Boss, E., Bernard, S., Johnson, K., Lotiker, A., Ulloa, O., Perry, M. J., Uitz, J., Curie, M., Villefranche, D., Lazaret, C., Division, E. S., Berkeley, L., Road, O. C., Observation, E., Africa, S., Fermi, V., De Brest, C., Valley, O., and Jolla, L.: Bio-optical profiling floats as new observational tools for biogeochemical and ecosystem studies: potential synergies with ocean color remote sensing, in: Proceedings of the OceanObs 09: Sustained Ocean Observations and Information for Society Conference (Vol.2), edited by: Hall, J., Harrison, D. E., and Stammer, D., ESA Publ., Venice, Italy, 2010b.

Conkright, M. E., Locarnini, R. A., Garcia, H. E., O'Brien, T. D., Boyer, T. P., Stephens, C., and Antonov, J. I.: World Ocean Atlas 2001: Objective analyses, data statistics, and figures: CD-ROM documentation, US Department of Commerce, National Oceanic and Atmospheric Administration, National Oceanographic Data Center, Ocean Climate Laboratory, 2002.

Cullen, J. J.: The deep chlorophyll maximum: comparing vertical profiles of chlorophyll a, Can. J. Fish. Aquat. Sci., available at: http://agris.fao.org/agris-search/search.do?recordID=US201302185941 (last access: 23 February 2015), 1982.

Cullen, J. J. and Lewis, M. R.: Biological processes and optical measurements near the sea surface: Some issues relevant to remote sensing, J. Geophys. Res., 100, 13255, doi:10.1029/95JC00454, 1995.

Cunningham, A.: Variability of in-vivo chlorophyll fluorescence and its implication for instrument development in bio-optical oceanography, Sci. Mar., 60, 309–315, 1996.

Falkowski, P., Kiefer, D. A., Division, O. S., and Angeles, L.: Chlorophyll a fluorescence in phytoplankton: relationship to photosynthesis and biomass, J. Plankton Res., 7, 715–731, 1985.

Glover, D. M., Jenkins, W. J., and Doney, S. C.: Modeling Methods for Marine Science, Cambridge University Press, Cambridge, 2011.

Gordon, H. R. and McCluney, W. R.: Estimation of the depth of sunlight penetration in the sea for remote sensing, Appl. Optics, 14, 413–416, doi:10.1364/AO.14.000413, 1975.

Gregg, W. W.: Recent trends in global ocean chlorophyll, Geophys. Res. Lett., 32, L03606, doi:10.1029/2004GL021808, 2005.

Gregg, W. W. and Casey, N. W.: Global and regional evaluation of the SeaWiFS chlorophyll data set, Remote Sens. Environ., 93, 463–479, doi:10.1016/j.rse.2003.12.012, 2004.

Gregg, W. W., Conkright, M. E., Atlantic, N., Chlorophyll, B. O., Indian, S., Pacific, S., and Atlantic, S.: Decadal changes in global ocean chlorophyll, Geophys. Res. Lett., 29, 1–4, 2002.

Guinet, C., Xing, X., Walker, E., Monestiez, P., Marchand, S., Picard, B., Jaud, T., Authier, M., Cotté, C., Dragon, A. C., Diamond, E., Antoine, D., Lovell, P., Blain, S., D'Ortenzio, F., and Claustre, H.: Calibration procedures and first dataset of Southern Ocean chlorophyll *a* profiles collected by elephant seals equipped with a newly developed CTD-fluorescence tags, Earth Syst. Sci. Data, 5, 15–29, doi:10.5194/essd-5-15-2013, 2013.

Johnson, K., Berelson, W., Boss, E., Chase, Z., Claustre, H., Emerson, S., Gruber, N., Kortzinger, A., Perry, M., and Riser, S.: Observing Biogeochemical Cycles at Global Scales With Profiling Floats and Gliders Prospects for a Global Array, Oceanography, 22, 216–225, 2009.

Johnson, R., Strutton, P. G., Wright, S. W., McMinn, A., and Meiners, K. M.: Three improved satellite chlorophyll algorithms for the Southern Ocean, J. Geophys. Res.-Ocean., 118, 3694–3703, doi:10.1002/jgrc.20270, 2013.

Kiefer, D. A.: Chlorophyll a fluorescence in marine centric diatoms: Responses of chloroplasts to light and nutrient stress, Mar. Biol., 23, 39–46, doi:10.1007/BF00394110, 1973.

Le Quere, C., Harrison, S., Prentice, I., Buitenhuis, E., Aumont, O., Bopp, L., Claustre, H., de Cunha, L., Geider, R., Giraud, X., Klaas, C., Kohfield, K., Legendre, L., Manizza, M., Platt, T., Rivkin, R., Sathyendranath, S., Uitz, J., Watson, A., and Wolf-Gladrow, D.: Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models, Glob. Chang. Biol., available at: https://ueaeprints.uea.ac.uk/27956/ (last access: 19 March 2014), 2005.

Levitus, S., Antonov, J. I., Baranova, O. K., Boyer, T. P., Coleman, C. L., Garcia, H. E., Grodsky, A. I., Johnson, D. R., Locarnini, R. A., Mishonov, A. V., Reagan, J. R., Sazama, C. L., Seidov, D., Smolyar, I., Yarosh, E. S., and Zweng, M. M.: The World Ocean Database, Data Sci. J., 12, WDS229–WDS234, 2013.

Longhurst, A. R.: Ecological Geography of the Sea, available at: http://books.google.com/books?hl=fr&lr=&id=QdJZezzrCfQC&pgis=1 (last access: 22 January 2015), 2010.

Lorenzen, C. J.: A method for the continuous measurement of in vivo chlorophyll concentration, Deep Sea Res. Oceanogr. Abstr., 13, 223–227, doi:10.1016/0011-7471(66)91102-8, 1966.

McClain, C. R.: A decade of satellite ocean color observations, Ann. Rev. Mar. Sci., 1, 19–42, doi:10.1146/annurev.marine.010908.163650, 2009.

Mignot, A., Claustre, H., D'Ortenzio, F., Xing, X., Poteau, A., and Ras, J.: From the shape of the vertical profile of in vivo fluorescence to Chlorophyll-*a* concentration, Biogeosciences, 8, 2391–2406, doi:10.5194/bg-8-2391-2011, 2011.

Mignot, A., Claustre, H., Uitz, J., Poteau, A., D'Ortenzio, F., and Xing, X.: Understanding the seasonal dynamics of phytoplankton biomass and the deep chlorophyll maximum in oligotrophic environments: A Bio-Argo float investigation, Global Biogeochem. Cycles, 28, 856–876, doi:10.1002/2013GB004781, 2014.

Morel, A. and Berthon, J.-F.: Surface Pigments, Algal Biomass Profiles, and Potential Production of the Euphotic Layer: Relationships Reinvestigated in View of Remote-Sensing Applications, Limnol. Oceanogr., 34, 1545–1562, 1989.

O'Brien, C. J., Peloquin, J. A., Vogt, M., Heinle, M., Gruber, N., Ajani, P., Andruleit, H., Arístegui, J., Beaufort, L., Estrada, M., Karentz, D., Kopczynska, E., Lee, R., Poulton, A. J., Pritchard, T., and Widdicombe, C.: Global marine plankton functional type biomass distributions: coccolithophores, Earth Syst. Sci. Data, 5, 259–276, doi:10.5194/essd-5-259-2013, 2013.

Peloquin, J., Swan, C., Gruber, N., Vogt, M., Claustre, H., Ras, J., Uitz, J., Barlow, R., Behrenfeld, M., Bidigare, R., Dierssen, H., Ditullio, G., Fernandez, E., Gallienne, C., Gibb, S., Goericke, R., Harding, L., Head, E., Holligan, P., Hooker, S., Karl, D., Landry, M., Letelier, R., Llewellyn, C. A., Lomas, M., Lucas, M., Mannino, A., Marty, J.-C., Mitchell, B. G., Muller-Karger, F., Nelson, N., O'Brien, C., Prezelin, B., Repeta, D., Jr. Smith, W. O., Smythe-Wright, D., Stumpf, R., Subramaniam, A., Suzuki, K., Trees, C., Vernet, M., Wasmund, N., and Wright, S.: The MAREDAT global database of high performance liquid chromatography marine pigment measurements, Earth Syst. Sci. Data, 5, 109–123, doi:10.5194/essd-5-109-2013, 2013.

Sauzède, R., Claustre, H., Jamet, C., Uitz, J., Ras, J., Mignot, A., and D'Ortenzio, F.: Retrieving the vertical distribution of chlorophyll a concentration and phytoplankton community composition from in situ fluorescence profiles: A method based on a neural network with potential for global-scale applications, J. Geophys. Res.-Ocean., 120, 451–470, doi:10.1002/2014JC010355, 2015a.

Sauzède, R., Lavigne, H., Claustre, H., Uitz, J., Schmechtig, C., D'Ortenzio, F., Guinet, C., and Pesant, S.: Global compilation of chlorophyll fluorescence profiles from several data bases, and from published and unpublished individual sources, available at: http://doi.pangaea.de/10.1594/PANGAEA.844212, last access: 19 March 2015b.

Sauzède, R., Lavigne, H., Claustre, H., Uitz, J., Schmechtig, C., D'Ortenzio, F., Guinet, C., and Pesant, S.: Global data product of chlorophyll a concentration and phytoplankton community composition (microphytoplankton, nanophytoplankton and picophytoplankton) computed from in situ fluorescence profiles, available at: http://doi.pangaea.de/10.1594/PANGAEA.844485, last access: 19 March 2015c.

Siegel, D. A., Maritorena, S., Nelson, N. B., and Behrenfeld, M. J.: Independence and interdependencies among global ocean color properties: Reassessing the bio-optical assumption, J. Geophys. Res., 110, C07011, doi:10.1029/2004JC002527, 2005.

Siegel, D. A., Behrenfeld, M. J., Maritorena, S., McClain, C. R., Antoine, D., Bailey, S. W., Bontempi, P. S., Boss, E. S., Dierssen, H. M., Doney, S. C., Eplee, R. E., Evans, R. H., Feldman, G. C., Fields, E., Franz, B. A., Kuring, N. A., Mengelt, C., Nelson, N. B., Patt, F. S., Robinson, W. D., Sarmiento, J. L., Swan, C. M., Werdell, P. J., Westberry, T. K., Wilding, J. G., and Yoder, J. A.: Regional to global assessments of phytoplankton dynamics from the SeaWiFS mission, Remote Sens. Environ., 135, 77–91, doi:10.1016/j.rse.2013.03.025, 2013.

Szeto, M., Werdell, P. J., Moore, T. S., and Campbell, J. W.: Are the world's oceans optically different?, J. Geophys. Res., 116, C00H04, doi:10.1029/2011JC007230, 2011.

Uitz, J., Claustre, H., Morel, A., and Hooker, S. B.: Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll, J. Geophys. Res., 111, C08005, doi:10.1029/2005JC003207, 2006.

Vidussi, F., Claustre, H., Manca, B. B., Luchetta, A., and Marty, J.-C.: Phytoplankton pigment distribution in relation to upper thermocline circulation in the eastern Mediterranean Sea during winter, J. Geophys. Res., 106, 19939, doi:10.1029/1999JC000308, 2001.

Xing, X., Claustre, H., Blain, S., D'Ortenzio, F., Antoine, D., Ras, J., and Guinet, C.: Quenching correction for in vivo chlorophyll fluorescence acquired by autonomous platforms: a case study with instrumented elephant seals in the Kerguelen region (Southern Ocean), Limnol. Oceanogr. Methods, 10, 483–495, doi:10.4319/lom.2012.10.483, 2012.

2.3.3 Résultats complémentaires

Les **premières climatologies en 3D** à l'échelle globale obtenues grâce à la base de données présentée dans l'article ci-dessus (base de données qui sera par la suite nommée BD-FLAVOR) peuvent être comparées aux climatologies provenant du World Ocean Atlas (WOA).

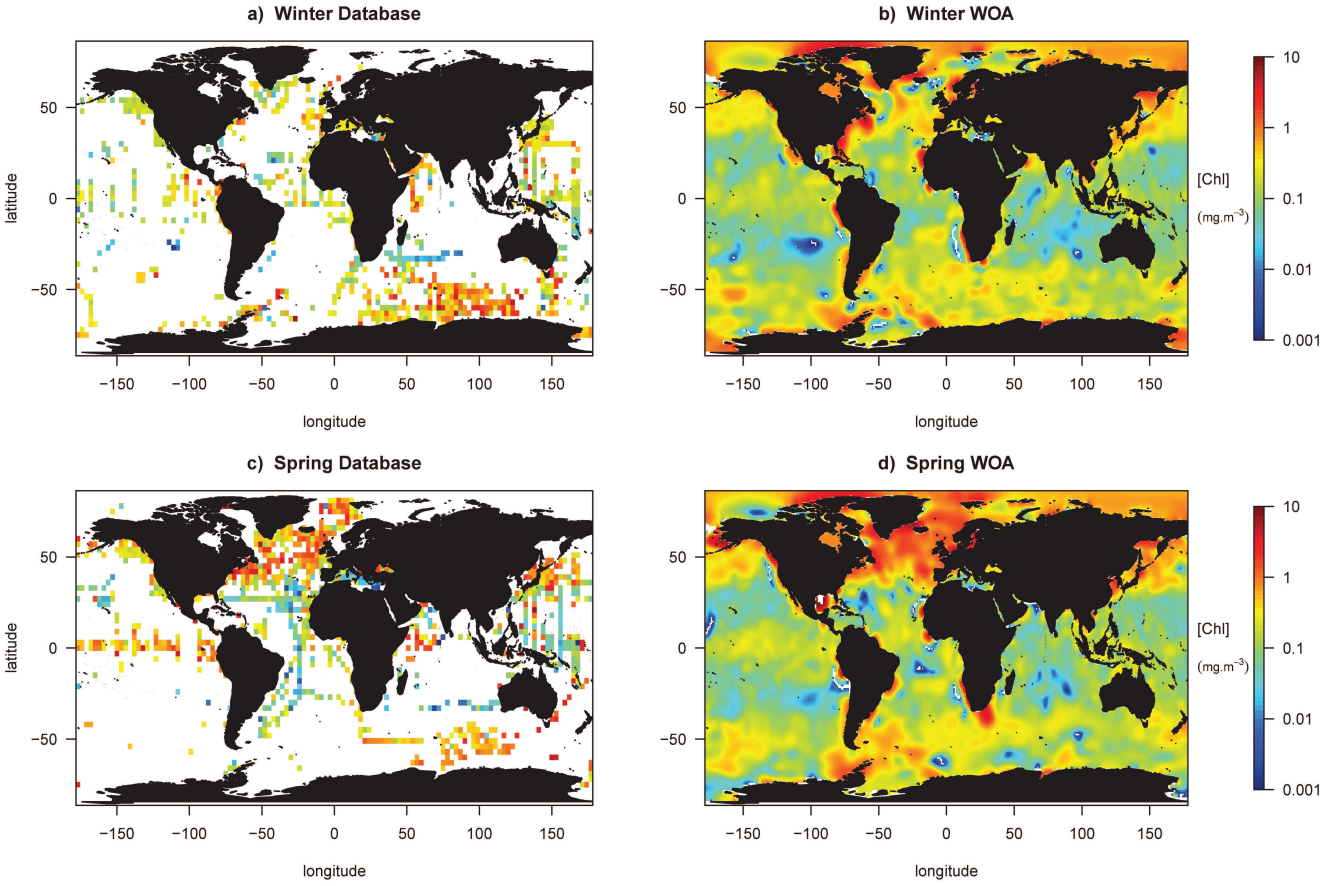


FIGURE 2.2 – *Comparaison des données climatologiques saisonnières de surface provenant de BD-FLAVOR (i.e. base de données calibrée) moyennées dans des mailles de 3˚x3˚avec les climatologies saisonnières provenant du WOA pour : a) et b) l'hiver (mois de janvier, février et mars) et c) et d) le printemps (mois d'avril, mai et juin).*

Tout d'abord, les **climatologies saisonnières de surface** provenant de BD-FLAVOR sont comparées aux climatologies saisonnières de surface du WOA sur la Figure 2.2 pour l'hiver (i.e mois de janvier, février et mars) et le printemps (i.e. mois d'avril, mai et juin) et sur la Figure 2.3 pour l'été (i.e. mois de juillet, aout et septembre) et l'automne(i.e. mois d'octobre, novembre et décembre). Les valeurs de $[Chl]$ provenant de la base de données calibrée sont moyennées selon des mailles de 3˚x3˚. Le biais d'échantillonnage présent entre l'hémisphère Sud et Nord pour la base de données

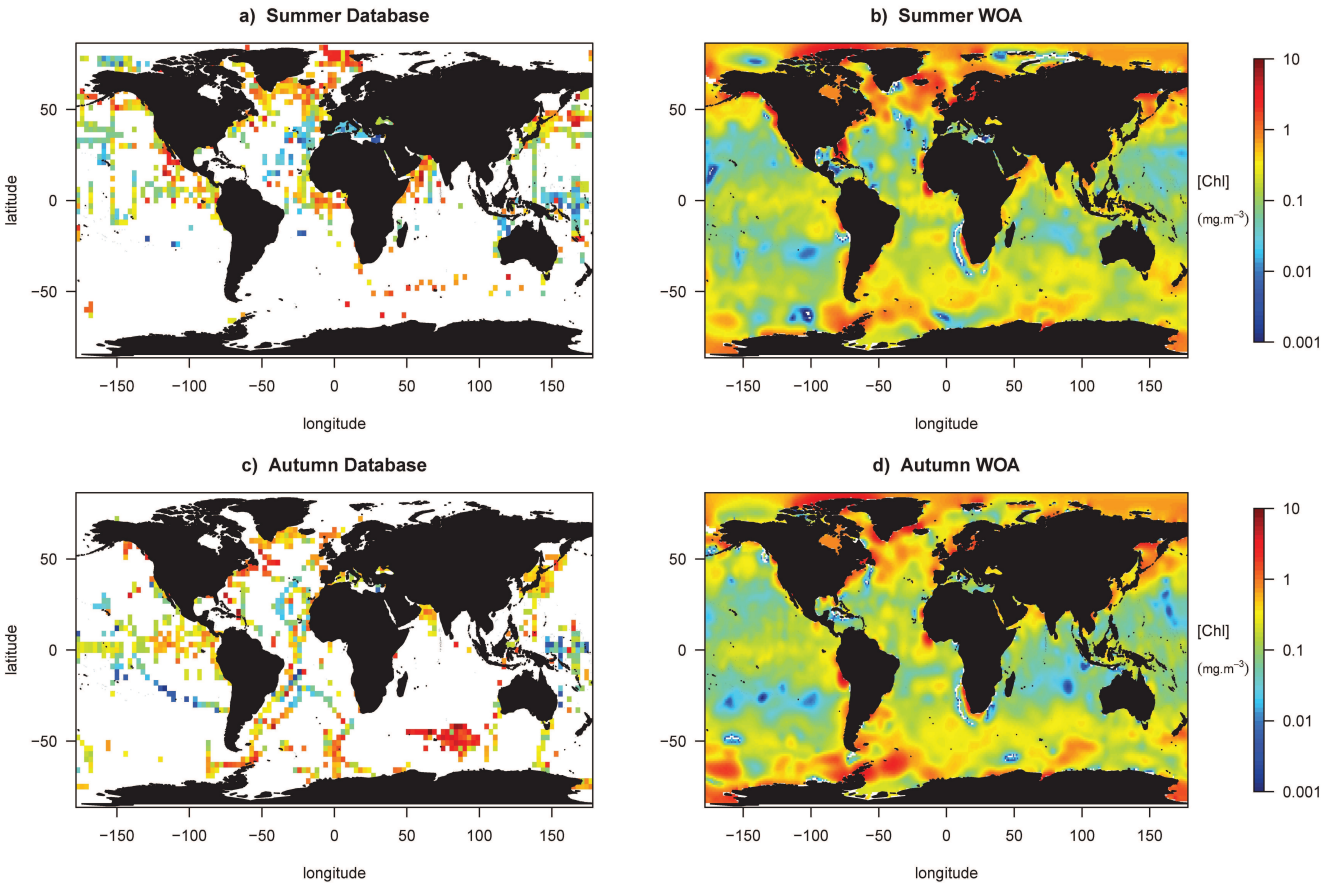


FIGURE 2.3 – *Comparaison des données climatologiques saisonnières de surface provenant de BD-FLAVOR moyennées dans des mailles de 3˚x3˚avec les climatologies saisonnières provenant du WOA pour : a) et b) l'été (mois de juillet, aout et septembre) et c) et d) l'automne (mois d'octobre, novembre et décembre).*

calibrée est particulièrement visible durant l'été (voir Figure 2.3-a). Nous nous attendons à ce que les données provenant des plateformes autonomes aident à corriger ce biais dans les années à venir. La gamme des valeurs de $[Chl]$ de surface climatologiques retrouvées pour chaque saison à partir de la base de données calibrée est équivalente à celle du WOA. Similairement aux résultats provenant du WOA, l'**Atlantique Nord** est caractérisé par de fortes valeurs de $[Chl]$ pour toutes les saisons. Nous pouvons tout de même noter une légère diminution de la $[Chl]$ en surface pendant l'hiver. La **zone équatoriale** est caractérisée par des valeurs intermédiaires de $[Chl]$ (i.e. environ $0.1\ mg.m^{-3}$) tout au long de l'année. Cependant, une légère augmentation de la $[Chl]$ durant le printemps est visible dans cette zone. Les **gyres subtropicaux** sont caractérisés tout au long de l'année par de très faibles valeurs de $[Chl]$ en surface (pouvant atteindre $0.01\ mg.m^{-3}$).

De manière similaire, les **climatologies en profondeur** du WOA ont été comparées aux résultats

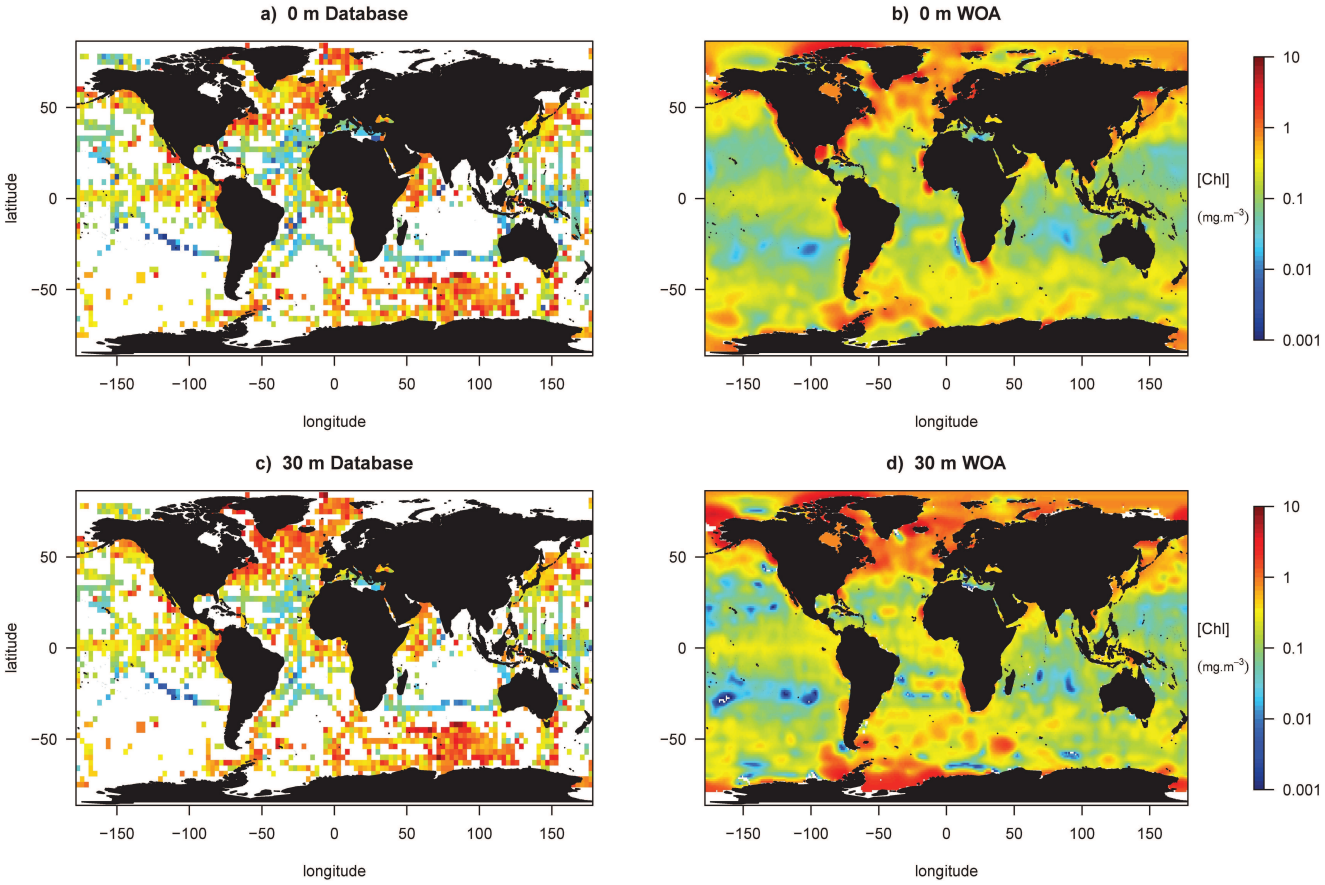


FIGURE 2.4 – *Comparaison des données climatologiques en profondeur provenant de BD-FLAVOR moyennées dans des mailles de 3˚x3˚avec les climatologies provenant du WOA pour : a) et b) la surface et c) et d) 30 m de profondeur.*

provenant de la base de données globale calibrée en $[Chl]$. Les comparaisons entre les climatologies issues de BD-FLAVOR pour 4 des profondeurs standards (i.e. surface, 30 m, 75 m et 100 m) et les climatologies du WOA sont présentées sur les Figures 2.4 et 2.5. Les valeurs climatologiques en profondeur provenant de BD-FLAVOR pour **l'Océan Austral** semblent être plus élevées que celles estimées par le WOA. Il est possible que notre méthode de calibration surestime les valeurs dans cette zone. Cependant, si nous comparons sur une même grille (1˚x1˚) la distribution des données BD-FLAVOR aux données utilisées par le WOA pour établir leurs climatologies, nous pouvons voir que les données du WOA sont moins nombreuses dans cette zone (voir Figures 2.1 et 2.6). À part pour l'Océan Austral, la gamme de $[Chl]$ climatologique estimée à chaque profondeur à partir de BD-FLAVOR est relativement similaire à celle obtenue par le WOA. De manière opposée, nous retrouvons bien de fortes valeurs de $[Chl]$ pour les **hautes latitudes** et des faibles valeurs de $[Chl]$ dans les **gyres**

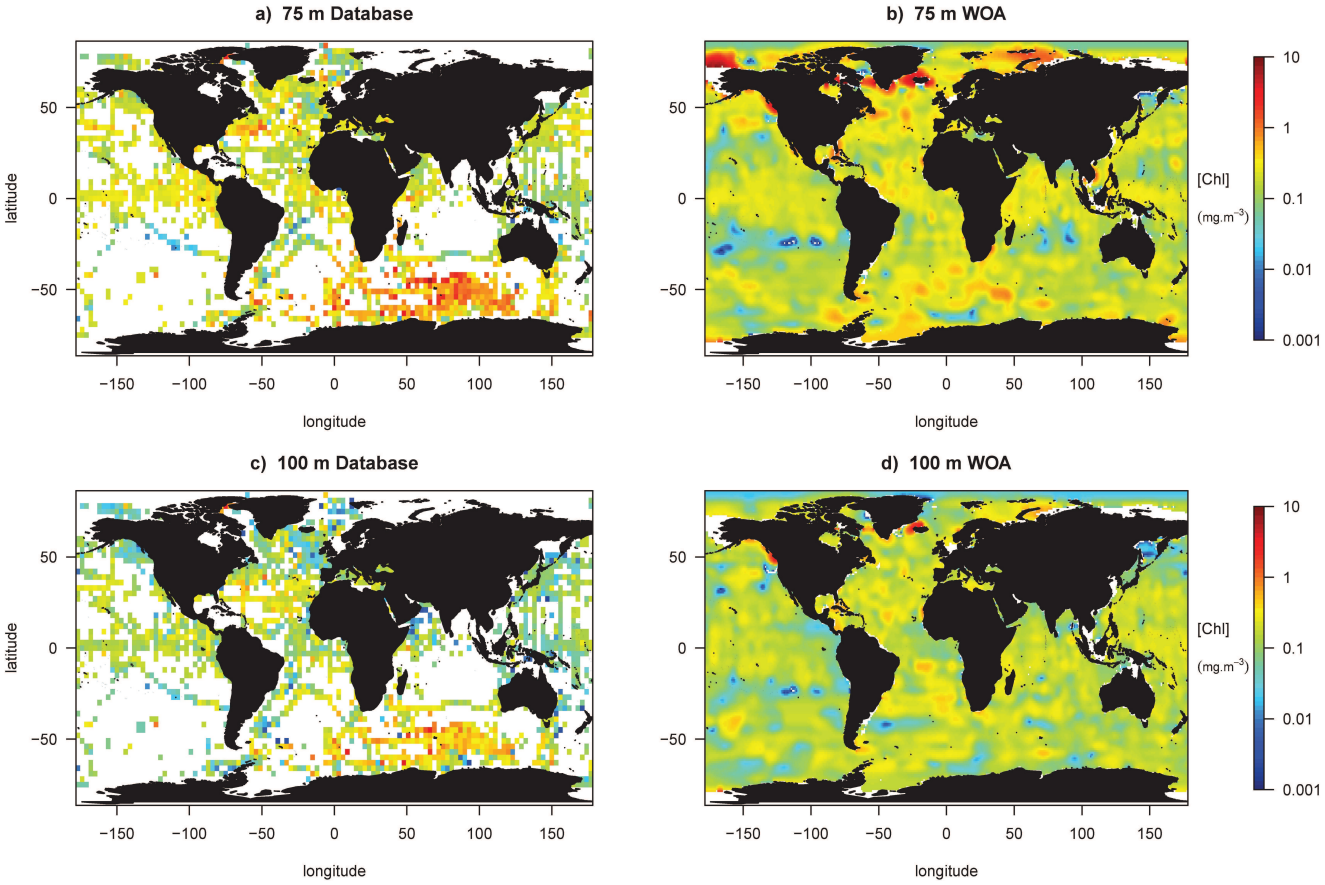


FIGURE 2.5 – *Comparaison des données climatologiques en profondeur provenant de BD-FLAVOR moyennées dans des mailles de 3˚x3˚avec les climatologies provenant du WOA pour : a) et b) 75 m de profondeur et c) et d) 100 m de profondeur.*

subtropicaux. D'après les climatologies du WOA et les résultats obtenus dans cette étude, les valeurs de $[Chl]$ ont globalement tendance à diminuer de la surface à 100 m.

Une telle base de données peut également être utilisée dans le but d'améliorer certaines **relations globales** basées sur la distribution verticale de la biomasse phytoplanctonique. Par exemple, une relation similaire à celle de Mignot *et al.* (2011) reliant la profondeur du DCM avec la concentration en chlorophylle *a* de surface des profils de $[Chl]$ stratifiés (i.e. profils pouvant être modélisés par une gaussienne) a été établie à partir de BD-FLAVOR (voir Figure 2.7). Notre relation semble mieux contraindre les valeurs extrêmes de profondeurs du DCM (e.g. DCM à 5-10m ou DCM >200 m). En effet, la relation de Mignot *et al.* (2011) repose sur un **nombre limité de données *in situ*** (~ 225 profils stratifiés) et l'utilisation d'un grand nombre de profils comme ceux utilisés ici (i.e. 22 739 profils stratifiés dans BD-FLAVOR) pourrait permettre d'affiner ce type de relations.

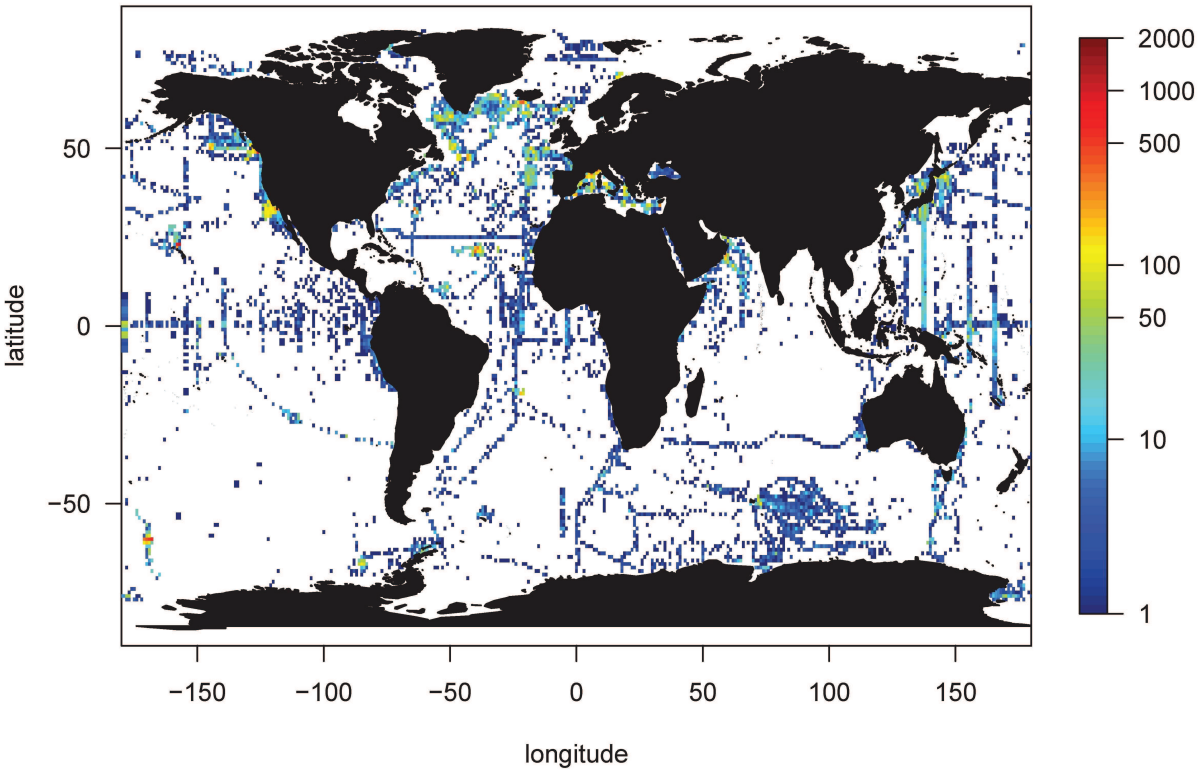


FIGURE 2.6 – *Distribution des profils de fluorescence calibrés en* $[Chl]$ *dans la base de données globale calibrée (BD-FLAVOR). La barre de couleurs correspond au nombre de profils agrégés dans des pixels de 1˚x1˚.*

2.3.4 Exemple d'utilisation de la base de données : une étude spécifique à la Méditerranée

Les profils de fluorescence de notre base de données sont également disponibles sous leurs formes « brutes » (i.e. non calibrés, mais ayant néanmoins passé le contrôle de qualité). Un jeu de données spécifique à la **Mer Méditerranée** a été assemblé à partir notamment des profils de fluorescence « bruts » de la base de données BD-FLAVOR (hébergée par PANGAEA). Ce sous ensemble a été créé dans le but de mieux connaître et comprendre la **distribution géographique** et **l'évolution temporelle** de la $[Chl]$ dans le bassin méditerranéen (Lavigne *et al.*, 2015). Les profils de fluorescence « bruts » de cette base de données ont été calibrés en $[Chl]$ en utilisant la méthode de Lavigne *et al.* (2012). Cette base de données de $[Chl]$ a pu mettre à jour notre connaissance de la distribution verticale de la $[Chl]$ en Méditerranée. Etant donné la rareté des données *in situ*, les études précédentes étaient essentiellement basées sur des mesures satellites de surface. La présente étude, réalisée à l'échelle locale et régionale, constitue un bon exemple des applications potentielles de la base de

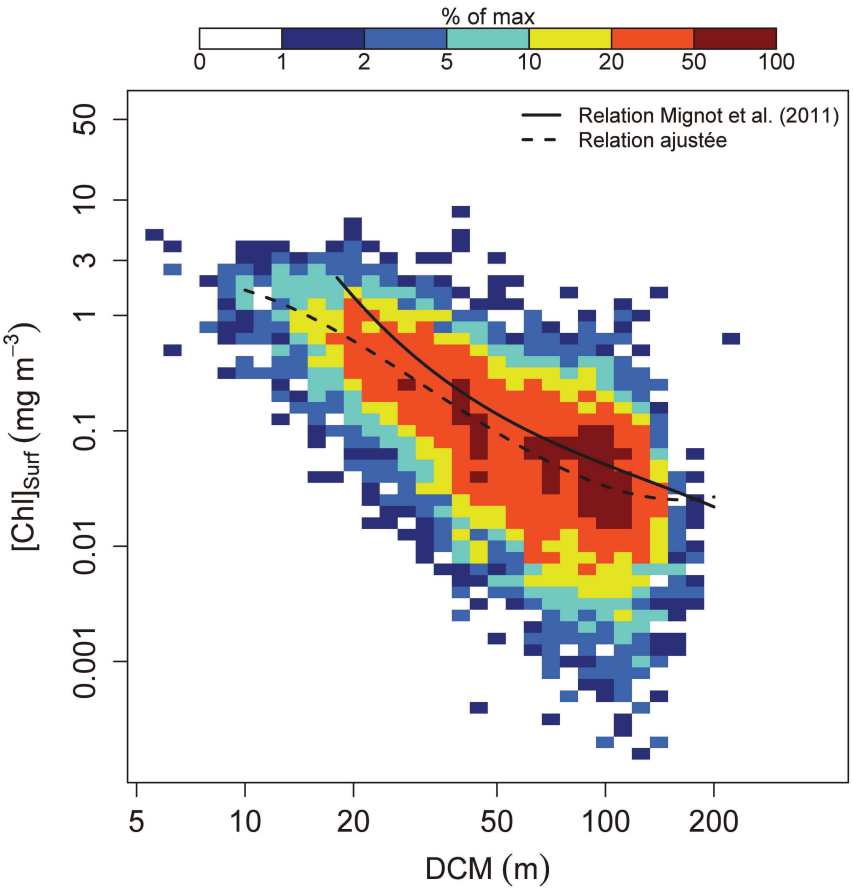


FIGURE 2.7 – *Relation entre la profondeur du DCM et la concentration en chlorophylle a de surface, [Chl]surf, pour les profils stratifiés de la base de données globale calibrée en* $[Chl]$ *(BD-FLAVOR). Les profils stratifiés sont définis ici comme les profils pouvant être modélisés par une gaussienne. La couleur correspond au pourcentage de points dans chaque pixel par rapport au maximum présent. La relation polynomiale d'ordre 3 établie par Mignot* et al. *(2011) est en trait plein tandis que la relation polynomiale ajustée aux données de BD-FLAVOR est en pointillés.*

données globale présentée dans l'article de la Sect. 2.3.2 (i.e. BD-FLAVOR).

Dans cette étude, les distributions verticales de $[Chl]$ ont été analysées au travers de **climatologies mensuelles** pour quatre sites clés de la Mer Méditerranée : le Nord-Ouest, le Sud-Ouest, la mer Ionienne et la mer Levantine (voir Figure 2.8). Ces climatologies confirment les résultats antérieurs concernant la gamme de valeur de $[Chl]$ observée au sein des principaux **régimes trophiques** coexistants en Méditerranée. Cette étude met également en avant des résultats inédits quant à la variabilité saisonnière de la forme du profil vertical de $[Chl]$. Pour cela, une analyse basée sur la reconnaissance de la forme générale du profil de fluorescence a été développée (voir Annexe A pour les détails).

Les **évolutions géographiques et saisonnières** du DCM sont également présentées et discutées pour les quatre régions clés de la Méditerranée (voir l'exemple de l'évolution saisonnière sur la Figure

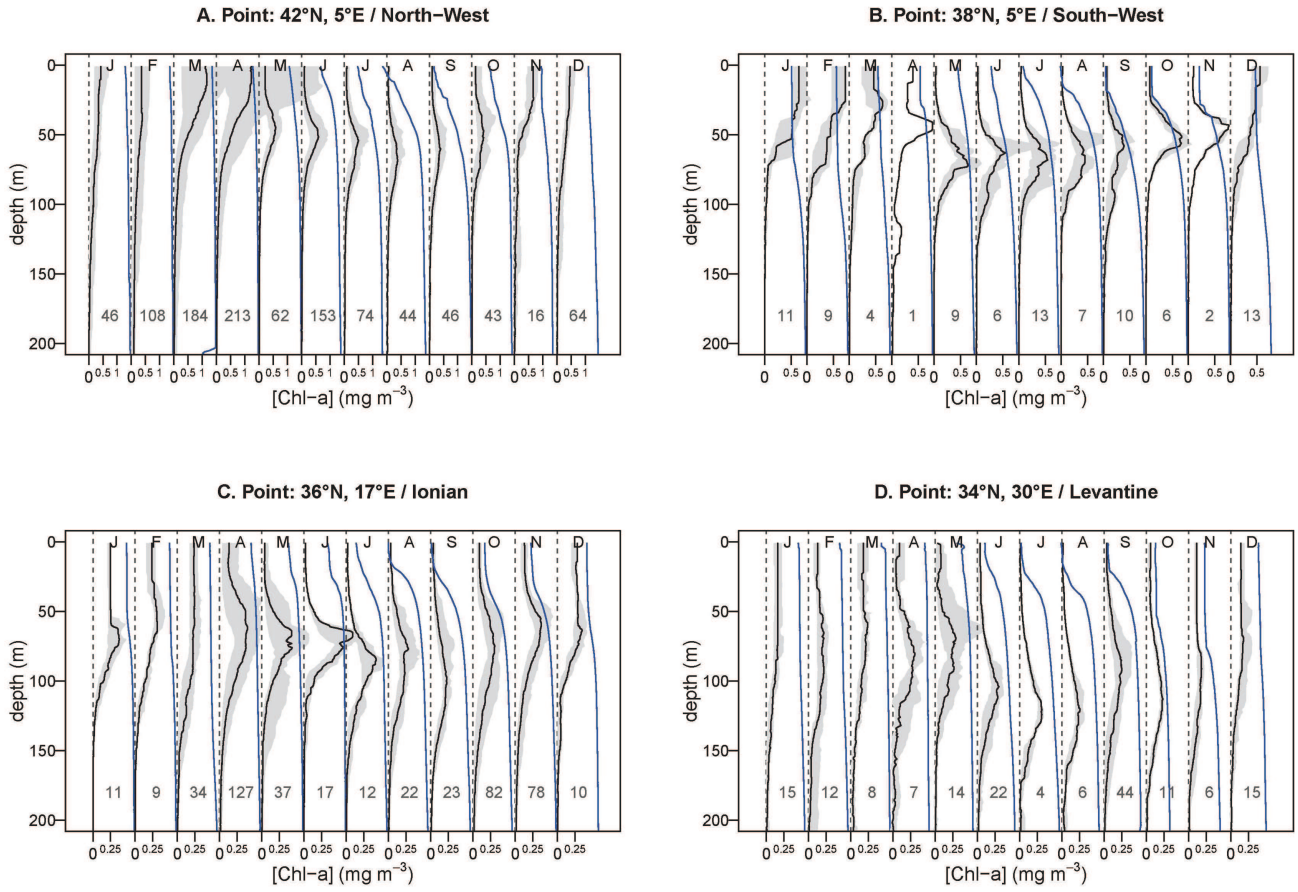


FIGURE 2.8 – *Climatologie des profils verticaux de* $[Chl]$ *pour quatre zones clés de la Méditerranée. La valeur médiane pour chaque mois est représentée par le trait noir. La zone grisée indique la gamme entre les quantiles 0.1 et 0.9. Les nombres en dessous de chaque profil climatologique correspondent aux nombre de profils utilisés pour générer ce profil.*

2.9). Une telle étude n'a pu être menée qu'à partir d'une base de données de $[Chl]$ basée sur des **profils de fluorescence** puisqu'un profil discret n'a pas la **résolution** suffisante pour en extraire des paramètres de forme comme le DCM de manière précise.

2.3.5 Conclusion et perspectives de l'étude

Nous avons montré que les profils de fluorescence jusqu'alors inexploités dans leur ensemble, avec les **outils adéquats**, pouvaient être rassemblés et étudiés dans leur globalité pour en extraire un certain nombre d'informations essentielle. Une des possibles utilisations de cette base de données serait de générer une **climatologie globale en 3D puis à terme en 4D** de la $[Chl]$ avec une **résolution verticale** de 15 à 20 fois meilleure que celle du WOA. De plus, en utilisant la méthode FLAVOR, de

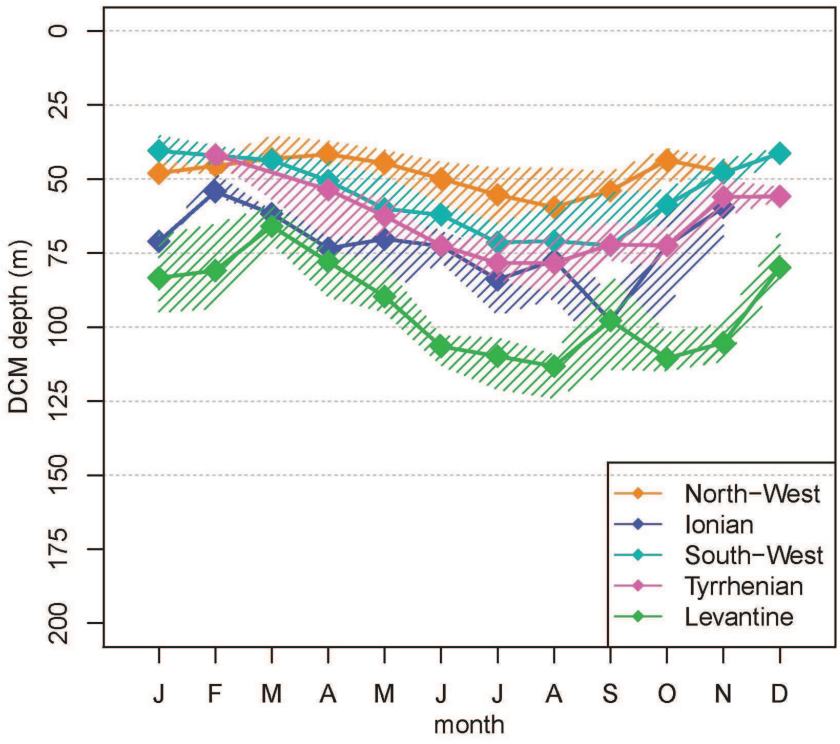


FIGURE 2.9 – *Evolution saisonnière de la profondeur du DCM. Les profondeurs du DCM ont été calculées seulement sur les profils « stratifiés » (détails dans la Sect. 2.3 de l'Annexe A). Les losanges représentent la médiane mensuelle et la zone hachurée indique les quartiles.*

nouveaux produits comme des climatologies de **la composition des communautés phytoplanctoniques** associée à la biomasse phytoplanctonique totale pourraient être également être produites.

Dans la synthèse « Ecological geography of the sea », Longhurst (2010) a réalisé **bio-régionalisation** de l'océan global basée sur la caractérisation des cycles saisonniers de la biomasse phytoplanctonique dérivés des observations satellitaires couplées aux données *in situ* (e.g. évolution de la couche de mélange, grazing). Les premières climatologies en 4D que nous avons présentées dans cette étude sont mensuelles et basées sur un **maillage spatial irrégulier** basé sur le concept des **provinces écologiques** définies par Longhurst (2010). Cependant, même si ces provinces peuvent constituer une forme de référence pour un certain nombre d'études, elles n'en ont pas moins été définies que de manière relativement arbitraire. Leurs contours rectilignes ne sont probablement pas indicatifs de frontières « réelles ». Par exemple, la Mer Méditerranée ne correspond qu'à une seule province écologique selon Longhurst (2010) alors qu'il est évident que plusieurs régimes trophiques coexistent dans ce bassin (D'Ortenzio et Ribera d'Alcalà, 2008). De plus, le **manque de flexibilité temporelle de ces provinces** constitue une limite à notre approche. Une des perspectives intéressante de cette étude

pourrait être de définir des bio-régions plus adaptées à la fois spatialement mais aussi temporellement. Ce type de régionalisation pourrait être entrepris à partir de méthodes **géo-statistiques** comme l'analyse de groupement, l'auto-corrélation spatiale et temporelle etc. A terme, afin d'obtenir une climatologie 4D de l'océan global, des méthodes **d'interpolation** et **d'extrapolation** des données doivent être choisies ou développées. Cependant, ces méthodes sont complexes à mettre en œuvre avec des jeux de données dont la distribution est fortement anisotrope comme en océanographie.

2.4 Conclusion générale

Dans ce chapitre, nous avons développé une méthode basée sur un **réseau de neurones** et permettant de reproduire la **distribution verticale de la biomasse phytoplanctonique** totale (i.e. $[Chl]$) ainsi que de **la composition des communautés phytoplanctoniques** à partir de la seule forme du profil de fluorescence. Cette méthode nous a permis de lier les bases de données « historiques » de profils de fluorescence aux nouvelles bases de données en les calibrant de manière **homogène** en $[Chl]$, les rendant ainsi interopérables. La base de données qui en résulte ouvre de nouvelles applications potentielles comme redéfinir certaines **relations globales** grâce à son nombre considérable de données (voir exemple dans la Sect. 2.3.3) ou la création d'une **climatologie globale en 4D** de la biomasse phytoplanctonique et des communautés phytoplanctoniques associées. De telles climatologies 4D très bien résolues sur la verticale (grâce à la résolution verticale fine des profils de fluorescence) constituent une source d'information essentielle pour un certain nombre d'applications. Par exemple, elles pourraient servir de références à partir desquelles des anomalies ou des **grandes tendances** d'origine climatique pourraient être mises en évidence. Par ailleurs, elles pourraient également être utilisées dans les **modèles biogéochimiques** prenant en compte en entrée la composition des communautés phytoplanctoniques afin d'initialiser les modèles ou de les valider (e.g. Aumont *et al.*, 2003).

Cependant, la base de données présentée dans cette étude reste encore relativement **incomplète**. Des zones comme l'Océan Indien ou le gyre subtropical du Pacifique Sud sont encore très mal échantillonnées. Ce manque de données *in situ* requiert une forte **interpolation** des données pour la création de produits globaux (e.g. climatologies) dans plusieurs zones majeures de l'océan ouvert. Pour extrapoler les données, des méthodes de restitution de la distribution verticale de la biomasse phytoplanctonique en fonction de la $[Chl]$ de surface pourraient par exemple être utilisées (e.g. Uitz *et al.*,

2006). Cependant, ces méthodes ont été développées avec peu de données *in situ* (méthodes basées sur des données « bouteilles » discrètes) et pourraient surement être améliorées grâce aux estimateurs bio-optiques de la biomasse phytoplanctonique (e.g. profils de fluorescence) acquis à haute fréquence grâce aux plateformes autonomes. Utiliser ces « nouvelles » **données bio-optiques** afin d'obtenir une **vue globale tridimensionnelle de la biomasse phytoplanctonique** dans l'océan global fait l'objet du chapitre 3.

CHAPITRE 3

Coupler des données satellites "couleur de l'eau" et des données *in situ* hydrologiques Argo dans le but de paramétriser la distribution verticale de plusieurs paramètres bio-optiques dans l'océan global

Sommaire

3.1 Introduction 86

3.2 Paramétrisation de la distribution verticale du coefficient de rétrodiffusion particulaire 89

3.2.1 Résumé de l'étude 89

3.2.2 Article soumis à *Journal of Geophysical Research* : A neural network-based method for merging ocean color and Argo data to extend surface bio-optical properties to depth : particulate backscattering coefficient retrieval 90

3.2.3 Conclusions et perspectives de l'étude 108

3.3 Paramétrisation de la distribution verticale de la concentration en chlorophylle *a* 109

3.3.1 Introduction 109

3.3.2 Databases and data processing 111

3.3.3 SOCA-CHL algorithm development 113

3.3.4 Results and Discussion 114

3.3.5 Conclusions and perspectives 117

3.4 Conclusion générale 118

3.1 Introduction

Les observations satellites **couleur de l'océan** permettent d'estimer plusieurs indicateurs bio-optiques de la biomasse phytoplanctonique (e.g. concentration en chlorophylle *a*, $[Chl]$, ou coefficient de rétrodiffusion particulaire, b_{bp}) à la surface des océans de manière **quasi-synoptique**. Cette acquisition à haute fréquence (à la fois spatiale et temporelle) de différentes propriétés bio-optiques ont fait de la couleur de l'eau un outil **essentiel** de quantification de la biomasse phytoplanctonique océanique à l'échelle globale (McClain, 2009). Cependant, les estimations provenant de la télédetection ne peuvent décrire que la couche de **surface** de la colonne d'eau (Gordon et McCluney, 1975). Cette couche superficielle ne représente qu'1/$5^{\text{ème}}$ de la couche productive de l'océan (Morel et Berthon, 1989). Or, cette information reste **insuffisante** pour beaucoup d'études biogéochimiques (impliquant la production primaire ou l'export de carbone par exemple). C'est pourquoi, un certain nombre d'études ont tenté d'étendre l'information satellitale de surface sur la dimension verticale. L'objectif premier de ces études était essentiellement de prendre en compte la partie **la colonne d'eau** où peut croitre le phytoplancton.

Tout d'abord, Morel et Berthon (1989) ont validé l'hypothèse de l'existence d'une relation entre la concentration en $[Chl]$ dans la couche de **surface**, pouvant être estimée par satellite, et le contenu en $[Chl]$ dans la couche euphotique mais aussi avec la **distribution verticale** de la $[Chl]$. Ces travaux ont abouti à un modèle empirique permettant de relier la $[Chl]$ de surface, pouvant être estimée à partir de la couleur de l'océan, à la distribution de la $[Chl]$ dans la colonne d'eau (i.e. en vertical). Cette méthode, parmi d'autres (e.g. Lewis *et al.*, 1983, Platt *et al.*, 1988, Sathyendranath *et al.*, 1989, Richardson *et al.*, 2003), a constitué une première étape dans l'étude de la **distribution verticale** de la biomasse phytoplanctonique à partir des **données satellites couleur de l'eau**. Néanmoins, ces méthodes ne donnaient aucune information quant à la **composition** de cette biomasse, pourtant déterminante dans les processus du cycle océanique du carbone. Ainsi, afin de paramétriser la distribution verticale de la composition des communautés phytoplanctoniques en plus de la biomasse phytoplanctonique totale, Uitz *et al.* (2006) ont établi des relations empiriques entre la $[Chl]$ de surface et la **composition des communautés phytoplanctoniques** dans la colonne d'eau. Une autre étude plus récente a tenté d'étendre sur la verticale une variable biogéochimique essentielle dans l'étude des flux

de carbone à l'échelle globale et pouvant être estimée par satellite : le carbone organique particulaire (POC). En effet, similairement à Uitz *et al.* (2006), Duforêt-Gaurier *et al.* (2010) ont paramétrisé la distribution verticale du POC à partir de son estimation par satellite. Cependant, ces différentes **relations globales** ont été établies en utilisant relativement **peu de données *in situ*** puisqu'elles sont basées sur des mesures discrètes de pigments analysées par fluorimétrie (Morel et Berthon, 1989), HPLC (Uitz *et al.*, 2006) ou des profils *in situ* discrets de POC (Duforêt-Gaurier *et al.*, 2010). En effet, les mesures discrètes « traditionnelles » fournissent une information très détaillée des propriétés biogéochimiques de la colonne d'eau mais avec une résolution **verticale, spatiale et temporelle** relativement **limitée** (voir Sect. 1.2.2 du chapitre 1).

Les récentes avancées dans l'implémentation de **capteurs bio-optiques** sur les plateformes autonomes (i.e. flotteurs Bio-Argo) ont permis d'étudier plusieurs variables bio-optiques sur des échelles de temps et d'espace bien plus **larges** que ce qu'offrait les méthodes classiques basées sur des mesures discrètes. Le développement de tels outils a conduit la communauté optique à développer de nouvelles relations à partir des indicateurs bio-optiques afin d'en estimer des **paramètres biogéochimiques clés**. Nous présenterons ici les développements méthodologiques entrepris pour deux propriétés bio-optiques : la fluorescence et le coefficient de rétrodiffusion. La **fluorescence** est un estimateur largement utilisé de la **concentration en chlorophylle *a*** $[Chl]$ (voir chapitre 1 et 2, Lorenzen, 1966) et le **coefficient de rétrodiffusion** (b_{bp}), caractéristique des particules submicroniques (Morel et Ahn, 1991, Loisel *et al.*, 2007), est un estimateur reconnu du **POC** (Balch *et al.*, 2001, Boss *et al.*, 2004, Cetinić *et al.*, 2012a, Loisel *et al.*, 2001, 2002, Stramski *et al.*, 1999, 2008). D'autres études ont également montré que le b_{bp} pouvait également être utilisé comme un proxy du **carbone phytoplanctonique** (Behrenfeld *et al.*, 2005, Graff *et al.*, 2015, Martinez-Vicente *et al.*, 2013). La fluorescence et le b_{bp} sont donc deux **indicateurs bio-optiques clé** à mesurer pour estimer la **biomasse phytoplanctonique** dans l'océan global. L'avantage de ces deux propriétés est qu'elles sont facilement mesurables grâce à des capteurs bio-optiques effectuant des mesures **non-intrusives** à **haute résolution** le long de la colonne d'eau, permettant ainsi l'acquisition d'un profil **quasi-continu** sur la verticale. Grâce aux plateformes autonomes (flotteurs profileurs Bio-Argo), des bases de données très denses de profils bio-optiques (e.g. fluorescence et b_{bp}) sont en train de se constituer (voir exemple pour la fluorescence sur la Figure 1.8 du chapitre 1). Il devient donc envisageable de **renforcer** le

type d'approches présentées précédemment qui tentent de paramétriser la distribution verticale de plusieurs grandeurs biogéochimiques en utilisant l'information de surface provenant des satellites en exploitant dans leur globalité le nombre de **profils bio-optiques conséquent** accessibles à partir des **flotteurs Bio-Argo**.

L'étude présentée dans ce chapitre tente donc de tirer profit de la densité des profils bio-optiques des bases de données Bio-Argo afin d'en dériver une vue tridimensionnelle des **estimateurs bio-optiques de la biomasse phytoplanctonique** dans l'océan global. La distribution verticale des paramètres bio-optiques comme la fluorescence ou le b_{bp} est en quelque sorte guidée par la disponibilité en nutriments et en lumière, qui est elle-même influencée par les forçages physiques dans la colonne d'eau. Ainsi, nous nous attendons qu'en combinant **l'information satellitale** aux propriétés verticales **hydrologiques** de la colonne d'eau, il soit possible d'estimer la **distribution verticale des propriétés bio-optiques** avec plus de précision. Depuis le lancement du programme **Argo**, les profils de température et de salinité sont mesurés de manière quasi-continue avec une **haute résolution spatiale et temporelle** dans l'océan global (tous les ~ 10 jours tous les ~ 3°, Roemmich *et al.*, 2009a). Actuellement, plus de 3800 flotteurs Argo sillonnent les océans et ce réseau permet d'avoir une vue unique des propriétés hydrologiques à haute résolution dans la couche océanique de la surface à 2000 m de profondeur. Ce type de données (i.e. Argo) semble donc tout à fait adapté à être couplé aux données satellitaires dans le but de paramétriser la distribution verticale des propriétés bio-optiques dans l'océan global.

Pour mettre en place une nouvelle paramétrisation de la distribution verticale des propriétés bio-optiques, nous avons choisi d'utiliser les **réseaux de neurones** car ce sont des méthodes très **puissantes** qui permettent d'approximer n'importe quelle fonction continue et dérivable (voir détails dans la Sect. 1.4 du chapitre 1, Hornik *et al.*, 1989).

Dans ce chapitre nous présenterons deux réseaux de neurones permettant d'estimer la distribution verticale : 1) du **coefficient de rétrodiffusion particulaire** (b_{bp}) et 2) de la **concentration en chlorophylle *a*** ($[Chl]$). Ces méthodes seront nommées par la suite **SOCA-BBP** et **SOCA-CHL** respectivement pour Satellite Ocean Color merged with Argo data to infer the vertical distribution of -Particulate Backscattering coefficient et -Chlorophyll *a* concentration. Ces méthodes SOCA utilisent des données **concomitantes** et géo-localisées (matchup) de produits satellitaires couleur de l'eau avec

des profils verticaux hydrologiques Argo pour estimer les distributions verticales de b_{bp} et de la $[Chl]$. Les entrées nécessaires à ces paramétrisations peuvent être séparées plus particulièrement en trois composantes principales : 1) une composante de **surface** composée des produits satellitaires de b_{bp} et de $[Chl]$ pour SOCA-BBP et de $[Chl]$ et de PAR pour SOCA-CHL (estimés par MODIS-Aqua), 2) une composante **verticale** correspondant à des propriétés physiques de la colonne d'eau dérivées des profils verticaux T/S Argo et 3) une composante **temporelle** définie comme le jour du matchup considéré. Les réseaux de neurones sont basés sur une phase d'**apprentissage** et une phase de **validation** qui reposent sur une **base de données globale** de profils de température, salinité, b_{bp} et fluorescence collectés simultanément par les flotteurs **Bio-Argo** (~5 000 profils) co-localisés avec des observations satellitaires couleur de l'eau. La base de données utilisée est représentative des différentes conditions hydrologiques et trophiques pouvant être rencontrées dans l'océan global, ce qui permet de penser que les méthodes SOCA-BBP et SOCA-CHL ici présentées seront applicable à la plupart des eaux océaniques.

3.2 Paramétrisation de la distribution verticale du coefficient de rétrodiffusion particulaire

3.2.1 Résumé de l'étude

Cette étude examine le potentiel de coupler les observations **couleur de l'océan** avec des mesures verticales de température et salinité acquises par des flotteurs profileurs **Argo** dans le but de modéliser la **distribution verticale du coefficient de rétrodiffusion** particulaire (b_{bp}), estimateur reconnu du **POC** et du **carbone phytoplanctonique** dans l'océan.

La méthode développée dans le cadre de ce travail est basée sur un **réseau de neurones** et sera ensuite nommée **SOCA-BBP** en référence à « Satellite Ocean-Color merged with Argo data to infer the vertical distribution of b_{bp} ». Il est possible de résumer les entrées du réseau de neurones en trois composantes principale : 1) les **estimations satellites** (i.e. b_{bp} et concentration en chlorophylle *a* télédetectés) correspondantes dans le temps et dans l'espace à un certain 2) profil vertical des propriétés **physiques** océaniques mesuré par un flotteur profileur Argo et enfin 3) le **jour** du matchup considéré.

Ce réseau de neurones est entrainé puis validé à partir d'une **base de données globale** constituée de 4 725 profils simultanés de température, salinité et b_{bp} mesurés par des flotteurs profileurs Bio-Argo, concomitant aux produits satellitaires (matchups calculés avec les composites MODIS aqua moyennés sur 8jours/9km).

Cette base de données est représentative de l'océan ouvert en termes de conditions trophiques et hydrologiques, ce qui permet de penser que la méthode sera utilisable dans la plupart des régions océaniques.

SOCA-BBP est validée à partir de données **indépendantes** qui n'ont pas été intégrées à l'apprentissage du réseau de neurones (provenant de 20% de la base de données initiale choisis aléatoirement). En se basant sur ces 20% de données, la distribution verticale estimée à partir de SOCA-BBP est relativement précise avec une erreur globale de 21% (voir Figure 5-a de l'article soumis ci-dessous). De plus, **d'autres validations** appuient ces résultats et sont basées sur des jeux de données « totalement » indépendants : 1) acquis par 4 flotteurs Bio-Argo choisies dans différents bassins majeurs océaniques qui n'ont pas été intégrés dans la base de données d'entrainement et de validation et 2) acquis pendant la campagne océanographique AMT de 2009 (Atlantic Meridional Transect). Ces 2 jeux de données indiquent également que les résultats provenant de notre méthode sont très **cohérents** en comparaison aux données observées *in situ* (voir Figures 5-b, 7 et 8 de l'article).

Dans le but d'illustrer le potentiel d'une telle méthode, les profils **climatologiques** de température et salinité provenant des bases de données Argo (2004-2008) ont été fusionnés avec les produits satellites couleur de l'eau **climatologiques** dans le but de produire une climatologie en trois dimensions de la distribution verticale du b_{bp} pour l'océan global. L'évaluation de la distribution en **surface** pour les mois de juin et décembre de ces profils climatologiques de b_{bp} est en **accord** avec notre connaissance actuelle pour l'océan global (voir Figure 9 de l'article soumis ci-dessous).

3.2.2 Article soumis à *Journal of Geophysical Research* : A neural network-based method for merging ocean color and Argo data to extend surface bio-optical properties to depth : particulate backscattering coefficient retrieval

Submitted to Journal of Geophysical Research: Oceans

A neural network-based method for merging ocean color and Argo data to extend surface bio-optical properties to depth : particulate backscattering coefficient retrieval

R. Sauzède,[1] H. Claustre,[1] J. Uitz,[1] C. Jamet,[2] G. Dall'Olmo,[3,4] F. D'Ortenzio,[1]
B. Gentili,[1] A. Poteau,[1] and C. Schmechtig[1]

[1] Sorbonne Universités, UPMC Univ Paris 06, CNRS, Observatoire Océanologique de Villefranche (OOV), Laboratoire dOcéanographie de Villefranche (LOV), 181 Chemin du Lazaret, 06 230 Villefranche-sur-Mer, France.

[2] Laboratoire dOceanologie et de Geosciences, UMR8187, ULCO/CNRS, Wimereux, France.

[3] Plymouth Marine Laboratory, Plymouth, UK.

[4] National Centre for Earth Observation, Plymouth, UK.

Key words : particulate backscattering coefficient, ocean color, Argo profiling float, Bio-Argo profiling float, global ocean, neural network.

Key Points.

A neural network is developed for retrieving the vertical distribution of the backscattering coefficient

The neural network requires as input variables Argo T/S profiles and ocean color remote sensing products

The validation of the method is highly satisfactory which authorizes its application to the global open ocean

Abstract. The present study proposes a novel method that merges satellite ocean-color bio-optical products with Argo temperature-salinity profiles to infer the vertical distribution of the particulate backscattering coefficient (b_{bp}). This neural network-based method (SOCA-BBP for Satellite Ocean-Color merged with Argo data to infer the vertical distribution of the Particulate Backscattering coefficient) uses three main input components : (1) satellite-based surface estimates of b_{bp} and chlorophyll a concentration matched-up in space and time with (2) depth-resolved physical properties derived from temperature-salinity profiles measured by Argo profiling floats and (3) the day of the year of the considered satellite-Argo matchup. The neural network is trained and validated using a database including 4725 simultaneous profiles of temperature-salinity and bio-optical properties collected by Bio-Argo floats, with concomitant satellite-derived products. The Bio-Argo profiles are representative of the global open-ocean in terms of oceanographic conditions, making the proposed method applicable to most open-ocean environments. SOCA-BBP is validated using 20% of the entire database (global error of 21%). We present additional validation results based on two other independent datasets acquired (1) by four Bio-Argo floats deployed in major oceanic basins, not represented in the database used to train the method ; and (2) during an AMT (Atlantic Meridional Transect) field cruise in 2009. These validation tests based on two fully independent datasets indicate the robustness of the predicted vertical distribution of b_{bp}. To illustrate the potential of the method, we merged monthly climatological Argo profiles with ocean color products to produce a depth-resolved climatology of b_{bp} for the global ocean.

1. Introduction

The ocean plays an important role in the regulation of the climate of our planet by influencing the amount of carbon dioxide (CO_2) in the atmosphere. An important part of this regulation takes place through the so-called biological carbon pump, which results from the sinking and sequestration to the deep oceans of part of the stock of Particulate Organic Carbon (POC) produced by phytoplankton photosynthesis [Falkowski et al., 1998 ; Volk and Hoffert, 1985]. Despite their importance to the global carbon cycle, these processes are still poorly constrained. This is largely caused by a lack of observations of key biogeochemical properties and associated processes on relevant space and time scales.

Copyright 2015 by the American Geophysical Union.
0148-0227/15/$9.00

Traditional ship-based sampling and measurement methods provide direct, detailed information on biogeochemical properties of the water column, but with insufficient space-time coverage.

Recent advances in optical sensors implemented on in situ and remote-sensing platforms allow the study of biogeochemical variables and processes in the open ocean over a broad range of temporal and spatial scales. The increasing use of such optical tools has led the scientific community to develop optical proxies for estimating key biogeochemical parameters. Specifically the particulate backscattering coefficient and the particulate beam attenuation coefficient are widely used as proxies of POC [Bishop and Wood, 2009 ; Bishop, 2009 ; Gardner et al., 2006]. The particulate backscattering coefficient (b_{bp}) has received much attention in the recent years because it can be continuously measured in situ from autonomous platforms [Behrenfeld and Boss, 2003, 2006] or retrieved from satellite remote-sensing of ocean color [Behrenfeld et al., 2005 ; Siegel et al., 2005 ; Westberry et al., 2008]. Aside from being a relevant proxy of POC [Balch et al., 2001 ; Cetinić et al., 2012 ; Loisel et al., 2001, 2002 ; Stramski et al., 1999, 2008], this bio-optical property can be used as an index of the particulate load, and its spectral dependence as an index of particle size [e.g. DallOlmo and Mork, 2014 ; Loisel et al., 2006] and phytoplankton size structure [Kostadinov et al., 2010]. Although still debated, several recent studies have shown that b_{bp} could also be considered as an indicator of phytoplankton carbon [Behrenfeld et al., 2005 ; Graff et al., 2015 ; Martinez-Vicente et al., 2013]. This would make b_{bp} an interesting alternative to chlorophyll a concentration for monitoring phytoplankton biomass in situ or from space. Therefore b_{bp} appears as a key bio-optical property to study the space-time dynamics of POC and possibly of phytoplankton biomass, a prerequisite for ultimately improving the characterization and quantitative assessment of biologically mediated carbon fluxes in the global open ocean.

Satellite remote sensing of ocean color, coupled to relevant algorithms, has the potential to provide a quasi-synoptic view of b_{bp} which, in turn, can be interpreted in terms of POC [Loisel et al., 2001, 2002 ; Stramski et al., 1999, 2008]. We note that satellite-derived products of POC may also be obtained from reflectance- or beam attenuation-based algorithms [e.g. Gardner et al., 2006 ; Stramski et al., 2008]. Several studies have used this potential to examine the spatial and temporal distribution of POC in the open ocean [Gardner et al., 2006 ; Loisel et al., 2002 ; Stramska, 2009]. However, such satellite-based estimates are restricted to the ocean surface layer and, in the context of global carbon cycle studies, need to be extended to depth. To our knowledge, this has been attempted only by Dufort-Gaurier et al. [2010] who based their study on a relatively small database of POC vertical profiles.

Because b_{bp} is tightly linked to the stock of biologically derived carbon (POC), its vertical distribution must be in some way driven by nutrient availability and light regime, which are in turn influenced by the physical forcing of the water column. Hence, one may expect that combining the satellite-derived surface data of b_{bp} with available information on the physical state of the water column will help extending surface b_{bp} to depth and constraining its vertical distribution. Since the launch of the Argo program, temperature and salinity profiles are measured continuously with high spatio-temporal resolution throughout the worlds oceans [Roemmich et al., 2009]. Now mature, with more than 3800 active floats, the Argo array provides a unique high-resolution view of hydrological properties in the upper 2000 m of the ocean. These data represent an ideal candidate for merging with satellite ocean color products. Therefore, in this study, we propose to develop and examine the potential of a new global method for merging satellite ocean color and physical Argo data to infer the vertical distribution of b_{bp} with a relatively high spatio-temporal resolution, i.e. the resolution of Argo-to-satellite matchup data.

In the past few years, the number of in situ observations of the vertical distribution of b_{bp} has dramatically increased. This results from the integration of optical sensors on autonomous platforms, especially Bio-Argo profiling floats which almost all measure b_{bp} [Boss 2008 ; Claustre et al., 2010a, 2010b ; Mignot et al., 2014 ; Xing et al., 2014]. Concurrently with b_{bp} profiles, Bio-Argo floats collect vertical profiles of temperature and salinity from which the physical status of the water column can be determined (e.g. density profiles, mixed layer depth). Hence the numerous vertical profiles collected by Bio-Argo floats offer a new path for developing a global parameterization of the vertical distribution of this key bio-optical property. Our study aims to use the large database of physical and bio-optical vertical profiles collected by the Bio-Argo fleet within the global open ocean to establish the proposed method.

Artificial neural networks (ANNs) are very powerful methods for approximating any differentiable and continuous functions [Hornik et al., 1989] and have been widely used for biogeochemical, geophysical and remote sensing applications [e.g. Bricaud et al., 2007 ; Friedrich and Oschlies, 2009 ; Gross et al., 2000 ; Jamet et al., 2012 ; Krasnopolsky, 2009 ; Niang et al., 2006 ; Palacz et al., 2013 ; Raitsos et al., 2008 ; Sauzède et al., 2015 ; Telszewski et al., 2009]. These methods have a large potential to model complex and nonlinear relationships that are characteristic of ecological datasets [Lek and Guégan, 1999]. Furthermore, ANNs are relatively insensitive to reasonable noise levels of input data, because noise is accounted for during the training process of the ANN. Therefore, we selected this method as the most appropriate for reaching our goal.

In summary, this study presents a new ANN-based method that uses merged satellite ocean color-based products and physical Argo data to retrieve the vertical distribution of b_{bp} at the global scale. Hereafter the method is referred to as SOCA-BBP for Satellite Ocean Color merged with Argo data to infer the vertical distribution of the Particulate Backscattering coefficient. SOCA-BBP uses three main input components : (1) a surface component composed of satellite-based estimates of b_{bp} and chlorophyll a concentration, (2) vertically-resolved physical quantities derived from Argo temperature and salinity profiles, and (3) the day of the year of the considered satellite-to-Argo matchup. Our analysis utilizes a large database of concurrent in situ vertical profiles of temperature, salinity and b_{bp} collected by 78 Bio-Argo profiling floats (or 4725 profiles), matched up with satellite ocean color observations. The resulting database is representative of various trophic conditions, making the method largely applicable to the global open ocean.

2. Data presentation and processing

Below we present the Bio-Argo database used in this study, which is composed of concurrent vertical profiles of temperature, salinity and particulate backscattering coefficient. Then we present the procedure for matching up the Bio-Argo vertical profiles with satellite-based bio-optical products. We finally describe the resulting database used to develop and validate the SOCA-BBP algorithm.

2.1. Bio-Argo database of concurrent vertical profiles of temperature, salinity and particulate backscattering coefficient

In addition to the standard conductivity-temperature-depth (CTD) sensors mounted on physical Argo profiling floats, Bio-Argo floats are equipped with sensors dedicated

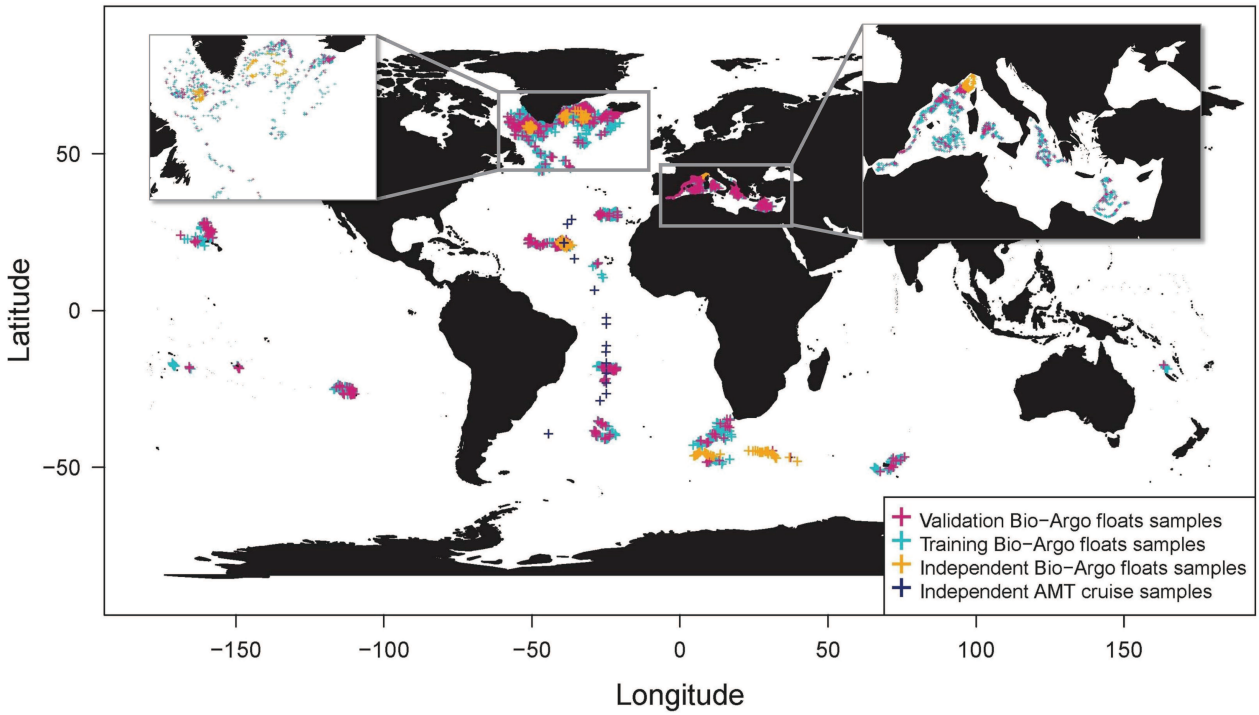


Figure 1. Geographic distribution of the 4725 stations used in the present study. For each station, concurrent profiles of temperature, salinity and b_{bp} collected by Bio-Argo floats were matched up with concomitant MODIS-Aqua-derived products. Turquoise and purple crosses indicate the location of the profiles of the so-called training and validation datasets, respectively (see text). The vertical profiles collected by the four independent Bio-Argo floats are shown as orange crosses. The vertical profiles collected during the AMT oceanographic cruise (also used for an independent validation of the method) are shown as dark blue crosses.

to the measurement of bio-optical properties that can be used as proxies of major biogeochemical variables. Specifically, the Bio-Argo floats are fitted with a CTD (Seabird) ; a sensor package (Satlantic OCR) that measures downwelling irradiance at three wavelengths and PAR (Photosynthetically Available Radiation) ; and a sensor package (WET Labs ECO Puck Triplet) composed of a chlorophyll *a* fluorometer, a CDOM (Colored Dissolved Organic Matter) fluorometer and a sensor measuring the particulate backscattering coefficient at a wavelength of 700 nm (79 floats) or 532 nm (4 floats). In the present study, we use exclusively measurements of temperature, salinity and particulate backscattering coefficient to train and validate the method.

The Bio-Argo floats used in this study typically collect measurements from 1000 m to the surface with a ∼1 m resolution every 10 days, 3 days, or even three times per day depending on the float mission configuration. When the float surfaces, data are transmitted in real time using Iridium communication. Thanks to this communication system the float mission parameters can also be modified in real time (e.g. time interval between two profiling cycles).

The volume scattering function (VSF), $\beta(\theta,\lambda)$ $(m^{-1}sr^{-1})$, is defined as the angular distribution of scattering relative to the direction of light propagation θ at the optical wavelength λ. The backscattering sensor of Bio-Argo floats measures $\beta(124^\circ,\lambda)$ with λ = 700 nm or 532 nm. The contribution of particles to the VSF, β_p, is calculated by subtracting the contribution of pure seawater, β_{sw}, from $\beta(124^\circ,\lambda)$:

$$\beta_p(124^\circ,\lambda) = \beta(124^\circ,\lambda) - \beta_{sw}(124^\circ,\lambda) \qquad (1)$$

with β_{sw} depending on temperature and salinity and computed using a depolarization ratio of 0.039 [Zhang et al., 2009].

Then, the particulate backscattering coefficient at 700 and 532nm, $b_{bp}(\lambda)$, is determined from $\beta_p(124^\circ,\lambda)$ and a conversion factor, χ, [Boss and Pegau, 2001 ; Kokhanovsky, 2012 ; Sullivan and Twardowski, 2009] as follows :

$$b_{bp}(\lambda) = 2\pi\chi(\beta(124^\circ,\lambda) - \beta_{sw}(124^\circ,\lambda)) \qquad (2)$$

The value of χ for an angle of 124 ° is 1.076 [Sullivan and Twardowski, 2009]. In order to harmonize the b_{bp} profiles of the Bio-Argo database, the profiles of $b_{bp}(532)$ were converted to $b_{bp}(700)$. The uncertainty associated to the conversion procedure is limited because the database includes ten times more $b_{bp}(700)$ profiles than $b_{bp}(532)$ profiles. The conversion was performed using a power law model of the particulate backscattering coefficient spectral dependency :

$$b_{bp}(\lambda) = b_{bp}(\lambda_0)\cdot(\frac{\lambda}{\lambda_0})^{-\gamma} \qquad (3)$$

We use a value of γ =2 for the Bio-Argo profiles collected in the North Pacific subtropical gyre and a value of γ =3 for those collected in the South Pacific Subtropical Gyre. These values are based on Loisel et al., [2006] who showed that the low chlorophyll waters of the subtropical gyres are typically associated with high γ values (between 2 and 3) whereas low or even negative γ values are found in the more productive areas of the ocean (between -1.5 and 1). The CTD data were quality controlled following the standard Argo protocol [Wong et al., 2014]. A quality control procedure was applied to each profile of $b_{bp}(700)$ (hereafter b_{bp} ; see Table 1 for a list of symbols) : (1) the manufacturer-supplied offsets and scaling factors were applied to each raw profile ; (2) high-frequency spikes were removed using a median filter ; (3) b_{bp} values above 0.03 m^{-1} were discarded as considered outside

Table 1. Abbreviations used in the present study and their significance.

Abbreviations	Significance
b_{bp}	Optical particulate backscattering coefficient at 700 nm (m^{-1})
Chl	Chlorophyll a concentration ($mg\ m^{-3}$)
R_{rs}	Remote-sensing reflectance (sr^{-1})
b_{bp_MODIS}	MODIS-Aqua-derived b_{bp} using QAA algorithm [Lee et al., 2002, 2009] (m^{-1})
b_{bp_VIIRS}	VIIRS-derived b_{bp} using QAA algorithm [Lee et al., 2002, 2009] (m^{-1})
Chl_{MODIS}	MODIS-Aqua-derived Chl ($mg\ m^{-3}$)
Chl_{VIIRS}	VIIRS-derived Chl ($mg\ m^{-3}$)
b_{bp_SOCA}	Vertically resolved values of b_{bp} retrieved by SOCA-BBP (m^{-1})
b_{bp_Floats}	Vertically resolved values of b_{bp} collected by Bio-Argo profiling floats (m^{-1})
b_{bp_AMT}	Vertically resolved values of b_{bp} collected during the AMT cruise (m^{-1})
z	Geometrical depth (m)
Z_{norm}	Depth at which the Chl profile returns to a constant background value at depth (m)
ζ	Depth normalized with respect to Z_{norm}, $\zeta = z/Z_{norm}$ (dimensionless)
Z_m	Mixed layer depth (m)
Z_e	Euphotic layer depth (m)
K_{d490}	Diffuse attenuation coefficient at 490 nm (m^{-1})
K_{PAR}	Diffuse attenuation coefficient for photosynthetically available radiation (m^{-1})
Day	Day of the year
Day_{rad}	Day transformed into radians
Z_{pd}	Penetration depth defined as $Z_{pd} = Z_e/4.6$ [Morel and Berthon, 1989] (m)
MAPD	Median Absolute Percent Difference (%)

of the sensor range of operation. The final Bio-Argo database of concurrent b_{bp} and temperature-salinity measurements is composed of 8330 vertical profiles collected by 83 Bio-Argo profiling floats.

2.2. Bio-Argo and satellite ocean color matchup database

Each profile of the Bio-Argo database described above was matched up with satellite data of surface b_{bp} and chlorophyll *a* concentration (Chl) using the closest pixel from 8-day MODIS-Aqua images with a 9-km resolution. For consistency with Bio-Argo b_{bp} data measured at (or converted to) 700 nm, the satellite-derived b_{bp} data were estimated for a wavelength of 700 nm. This transformation was computed by linear extrapolation using b_{bp} values obtained from the Quasi-Analytical Algorithm [QAA, Lee et al., 2002, 2009] applied to MODIS-Aqua remote-sensing reflectance, R_{rs}, at 7 wavelengths (i.e. 412, 443, 469, 488, 531, 555 and 667 nm ; see Figure S1).

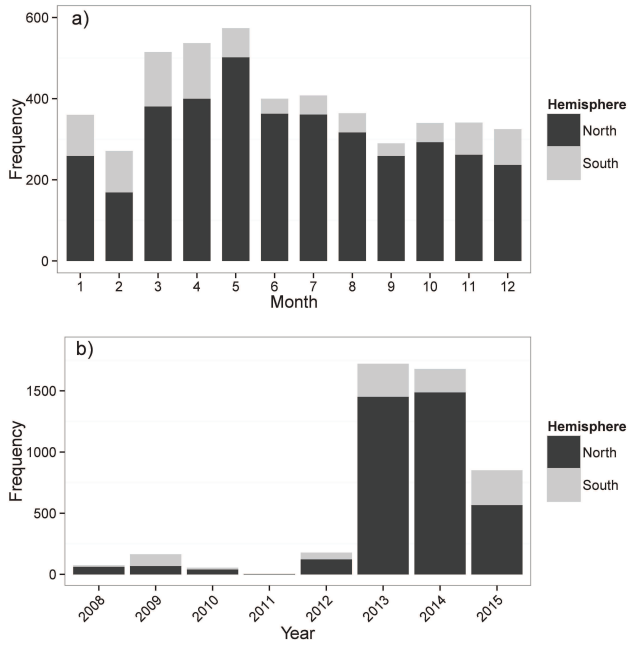


Figure 2. Temporal distribution of the 4725 stations for which both Bio-Argo and satellite data were simultaneously available as a function of (a) months and (b) years with black and grey colors indicating the hemisphere of data acquisition.

The matchup procedure led to discarding 43% of the profiles from the initial Bio-Argo database (see discarding rate for the major oceanic basins in Table 2). The geographic distribution of the 4725 remaining Bio-Argo profiles with concomitant MODIS-Aqua-derived products is presented in Figure 1. The database used in this study covers most of the major ocean basins (i.e. Southern Ocean, Indian Ocean, Mediterranean Sea, North Pacific, South Pacific, North Atlantic and South Atlantic ; see Figure S2 for details of basin boundaries). The under-representation of the southern hemisphere is apparent in Figure 2. However, we note that a substantial sampling effort has started in 2015 with multiple floats deployed in the Southern Ocean. On the annual scale, more profiles are available for spring and summer than for autumn and winter for the northern hemisphere (Figure 2a). This temporal bias of data acquisition is mostly due to a lack of satellite images at high latitudes during winter and autumn. For the southern hemisphere, the acquired data appear evenly distributed (Figure 2a). The annual distribution of the vertical profiles in the database covers 8 years from 2008 to 2015 (Figure 2b) ; most of the observations were nevertheless collected since 2013.

The resulting Bio-Argo and satellite matchup database appears to be representative of a broad variety of hydrological and biogeochemical conditions prevailing in the global

Table 2. Summary of the number of profiles rejected after the satellite versus Bio-Argo matchup procedure for the different Bio-Argo float sampling regions. The seven major oceanic basins boundaries used to compute this table are presented in Figure S2.

Area	Number of profiles	Number of profiles after matchup	% of profiles removed
North Atlantic	3985	1791	56%
South Atlantic	1073	620	43%
North Pacific	364	173	53%
South Pacific	275	188	32%
Southern Ocean	203	47	77%
Mediterranean Sea	2235	1839	18%
Indian Ocean	195	67	66%
Global ocean (total)	8330	4725	56%

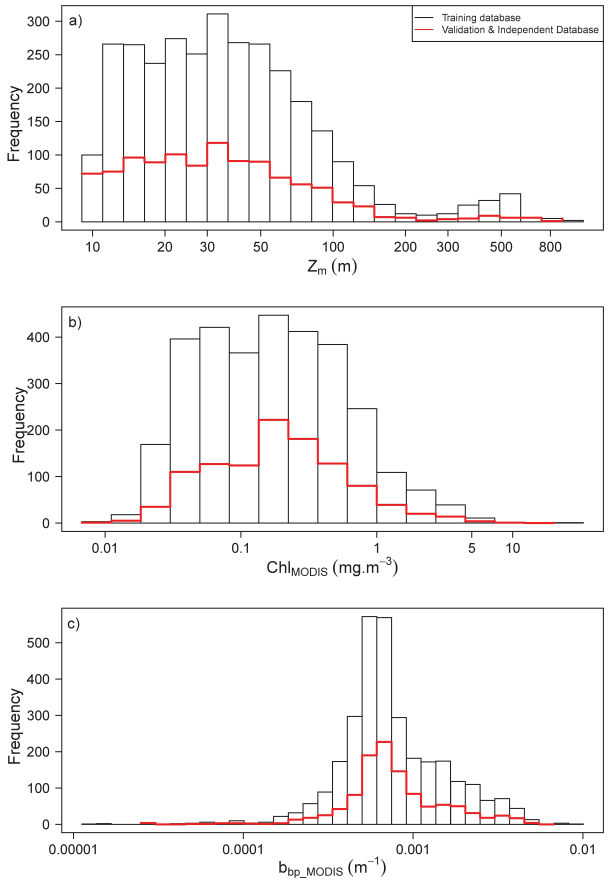


Figure 3. Histogram of frequency of (a) the mixed layer depth, Z_m (m), (b) the satellite-derived chlorophyll *a* concentration, Chl_{MODIS} ($mg\ m^{-3}$), and (c) the satellite-derived particulate backscattering coefficient, b_{bp_MODIS} (m^{-1}). The black histogram represents the distribution of the data used to train the method and the red histogram is for the independent dataset (20% of the initial database) used to validate the method.

open ocean (Figure 3). For instance, the values of mixed layer depth, Z_m, acquired by the Bio-Argo floats vary between 15 and 900 m (measurements from the North Atlantic Subtropical Gyre in spring and the North Atlantic in winter, respectively). Z_m is calculated from the density profiles using a density criterion of 0.03 $kg\ m^{-3}$ as in de Boyer Montégut et al. [2004]. The database is also representative of most trophic conditions observed in the open ocean (i.e. from oligotrophic to eutrophic waters, see Figure 3b). The MODIS-Aqua-estimated Chl, Chl_{MODIS}, covers 4 orders of magnitude (i.e. from 0.01 to 10 $mg\ m^{-3}$). The most oligotrophic conditions were found in the South Atlantic Subtropical Gyre in autumn and the most eutrophic in the North Atlantic, especially in the Labrador Sea during the spring bloom. Similarly, the MODIS-Aqua-derived b_{bp}, b_{bp_MODIS}, covers 4 orders of magnitude (from 0.00001 to 0.01 m^{-1}; Figure 3c).

Before splitting the Bio-Argo and satellite matchup database into two subsets for developing the neural network (i.e. the training and validation datasets), 314 profiles collected by four Bio-Argo floats were removed from the database to create an independent dataset used for an additional validation of the method. These four floats were chosen in four major oceanic basins : the Subpolar North Atlantic, the North Atlantic Subtropical Gyre, the Southern Ocean and the Mediterranean Sea.

Finally, the resulting matchup database was randomly split into two independent subsets, including 80% (3525 profiles used for training the MLP) and 20% (886 profiles used for validating the MLP) of the data. Similar to the training dataset, the validation dataset is representative of the hydrological and biogeochemical conditions prevailing in the global open ocean (see histograms in Figure 3). The geographical distribution of the sampling stations included in the training, validation and independent 4-float subsets is shown in Figure 1.

2.3. Normalization of the vertical profiles of the particulate backscattering coefficient

SOCA-BBP is designed to predict the vertical distribution of b_{bp} within the so-called productive layer. This corresponds essentially to the layer where most particle and phytoplankton stocks are confined. In mixed conditions, the thickness of the productive layer roughly coincides with that of the mixed layer. In stratified conditions (typically associated with the presence of a deep chlorophyll maximum) the productive layer is more linked to the euphotic layer. Here the productive layer is described through the introduction of a dimensionless depth, ζ [Sauzède et al., 2015], with ζ defined as the geometrical depth, z, divided by a normalization depth, Z_{norm} :

$$\zeta = \frac{z}{Z_{norm}} \tag{4}$$

with Z_{norm} defined as the depth at which the Chl vertical profile returns to a constant background value (depth of the bottom of the productive layer). As fluorescence profiles are always collected simultaneously with temperature, salinity and b_{bp} profiles by the Bio-Argo floats, Z_{norm} can be computed with precision using the fluorescence profiles [see Sauzède et al., 2015 for details].

The application of SOCA-BBP to merged satellite-Argo data does not require the availability of fluorescence profile anymore. For application purposes, we developed a statistical relationship between Z_{norm} and two parameters accessible or derivable from our input dataset : (1) the euphotic layer depth, Z_e, the depth at which irradiance is reduced to 1% of its surface value ; and (2) the mixed layer depth, Z_m. Z_e is computed with the following procedure : (1) the attenuation coefficient at 490 nm, K_{d490}, is determined using the satellite-derived chlorophyll *a* concentration [Morel and Maritorena, 2001] ; (2) the total attenuation coefficient, K_{PAR}, is retrieved from K_{d490} [Rochford et al., 2001] ; (3) finally,

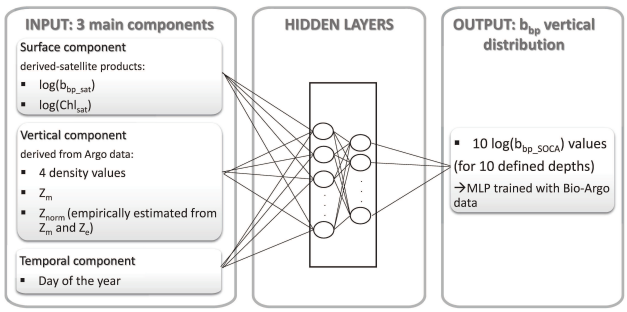


Figure 4. Schematic overview of the SOCA-BBP MLP-based algorithm that retrieves the vertical distribution of b_{bp} from merged ocean color satellite and Argo data associated with the day of the year of the considered satellite-to-Argo matchup.

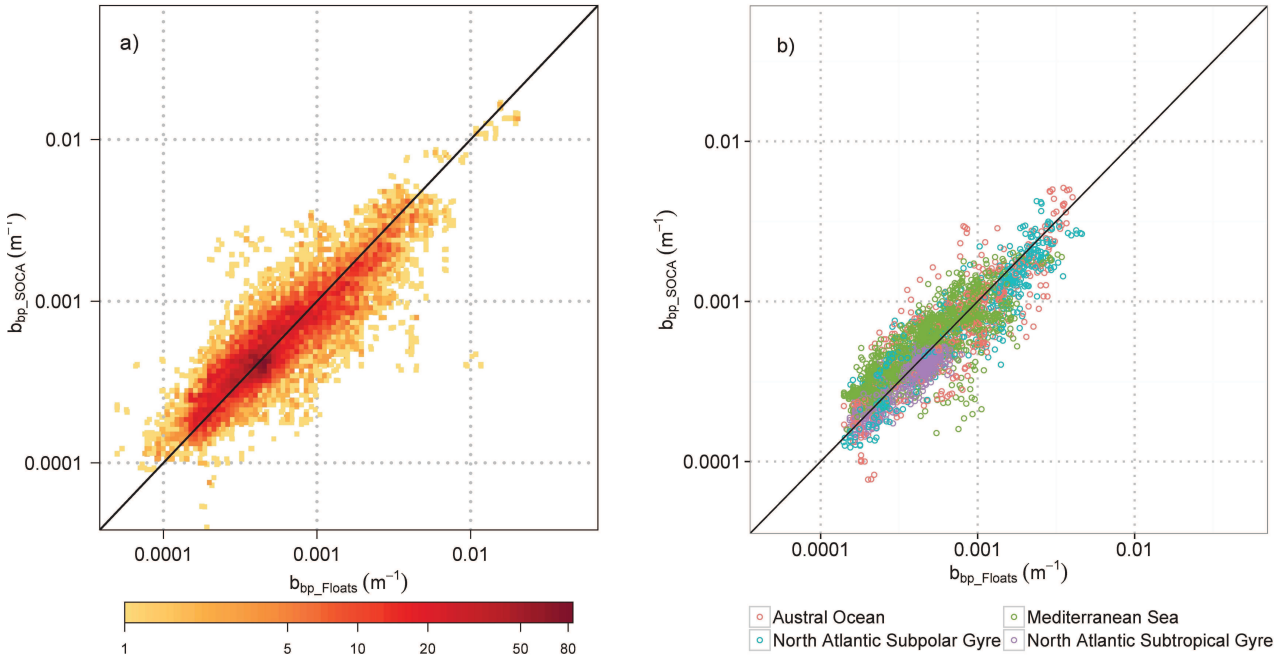


Figure 5. Comparison of the bbp values retrieved by SOCA-BBP (b_{bp_SOCA}) to the reference b_{bp} measurements acquired by the Bio-Argo floats (b_{bp_Floats}) using two different datasets : (a) the validation database (i.e. 20% of the entire database chosen randomly) with the color code indicating the number of points in each pixel ; (b) the independent data acquired by four Bio-Argo floats not integrated in the training and validation databases with the color code indicating the oceanic basins in which the Bio-Argo floats were deployed. The 1 :1 line is shown in each plot.

Z_e is retrieved from K_{PAR} using the exponential decrease of light over depth. The most statistically significant relationship between Z_{norm} and both the Z_e and Z_m parameters was found when stratified conditions are discriminated from mixed conditions based on the ratio of Z_e to Z_m [i.e. $Z_e > Z_m$: stratified ; $Z_e < Z_m$: mixed ; Morel and Berthon, 1989 ; Uitz et al., 2006]. We obtain the following optimal statistical relationships for a stratified water column :

$$log(Z_{norm}) = 0.12 \cdot log(Z_m) + 1.04 log(Z_e) \quad (5)$$

and for a mixed water column :

$$log(Z_{norm}) = 0.64 \cdot log(Z_m) + 0.51 log(Z_e) \quad (6)$$

The relationship between Z_{norm} computed from the fluorescence in situ profiles measured by the Bio-Argo floats and modeled with the statistical relationships presented above for the two hydrological regime of the water column (stratified or mixed) is satisfactory with a median absolute percent difference, MAPD, of 14% (for more details see Figure S3). Finally, Z_{norm} used to scale b_{bp} profiles ranges from 20 m to 805 m in the Bio-Argo database (see Figure S4). Scaling the b_{bp} profiles with respect to ζ enables the merging of all the profiles regardless of their vertical shape and range of variation while simultaneously accounting for their variability.

3. SOCA-BBP algorithm development

3.1. General principles of Multi-Layer Perceptron (MLP)

The type of artificial neural network chosen in this study is a Multi-Layer Perceptron [MLP ; Bishop, 1995 ; Rumelhart et al., 1988]. A MLP is composed of several layers : one input layer, one output layer and one or more intermediate levels (i.e. the so-called hidden layers). Each layer is composed of neurons, which are elementary transfer functions that provide outputs when inputs are applied. Each neuron is interconnected with the others by weights (Figure 4). The matrix of these weight values is iteratively adjusted during the training phase of the MLP and is computed by minimizing a cost function defined as the quadratic difference between the desired and computed outputs. The technique used for this minimization is the back-propagation conjugate-gradient, which is an iterative optimization method adapted to MLP development [Bishop, 1995 ; Hornik et al., 1989].

To determine the weights of the MLP, the training dataset is randomly split into two subsets (50% of the data each), the so-called learning and test datasets. These two subsets are used during the training process of the MLP to prevent overlearning [Bishop, 1995]. The validation dataset used to evaluate independently the final performance of the MLP is composed of 20% of the entire initial database.

3.2. Developing a MLP to retrieve the vertical distribution of b_{bp}

After multiple tests, the following set of three input components was selected as optimal (see Figure 4) : (1) a temporal component, i.e. the day of year ; (2) a surface component, defined by the satellite-derived log-transformed particulate backscattering coefficient (b_{bp_MODIS}) and chlorophyll *a* concentration (Chl_{MODIS}) (see Sect. 2.2) ; and (3) a vertical component, i.e. the normalization depth Z_{norm}, the mixed layer depth Z_m and four potential density values along the vertical profile, of which three taken at shallow depths and one at depth. The dimensionless depths according to density inputs were chosen using a principal component analysis to minimize redundancy in the selected input data (not shown). The MLP returns simultaneously 10 normalized values of $\log(b_{bp})$ as output according to ten depths taken at regular intervals within the $0 - 1.3\zeta$ layer. The elementary transfer function (sigmoid nonlinear function) that provides outputs when inputs are applied to the

MLP varies within the range [-1 ;1]. Therefore, to take advantage of the nonlinearity of this function, the inputs and outputs of the MLP ($x_{i,j}$) are centered and reduced to match the [-1 ;1] domain as follows :

$$x_{i,j} = \frac{2}{3} \cdot \frac{x_{i,j} - mean(x_{i,j})}{\sigma(x_{i,j})} \quad (7)$$

with σ the standard deviation of the considered input or output variable x. Obviously, the outputs need to be denormalized using the above equation with appropriate mean and standard deviation for each dimensionless depth of restitution. With respect to the temporal component, we applied a specific normalization procedure that accounts for the periodicity of the day of the year (i.e. day one of the year is very similar from a seasonal perspective to day 365). Thus, similar to the method developed by Sauzède et al. [2015], the temporal input is transformed in radians using the following equation :

$$Day_{rad} = \frac{Day \cdot \pi}{182.625} \quad (8)$$

where Day_{rad} is the day of the year in radian units and the coefficient 182.625 accounts for half the number of days per year (365.25). Once the optimal input and output variables were determined, additional tests were performed to establish the best architecture of the MLP : one or two hidden layers with a number of neurons in each layer varying between 1 and 50 and 1 and 20, respectively. The architecture with minimum error of validation and minimum number of neurons was selected as optimal. The resulting optimal MLP is here composed of two hidden layers with 8 neurons in the first hidden layer and 6 neurons in the second one. In order to evaluate the MLP robustness, different subsets of the training dataset have been tested and no significant difference in the predictive skills of the MLP was observed.

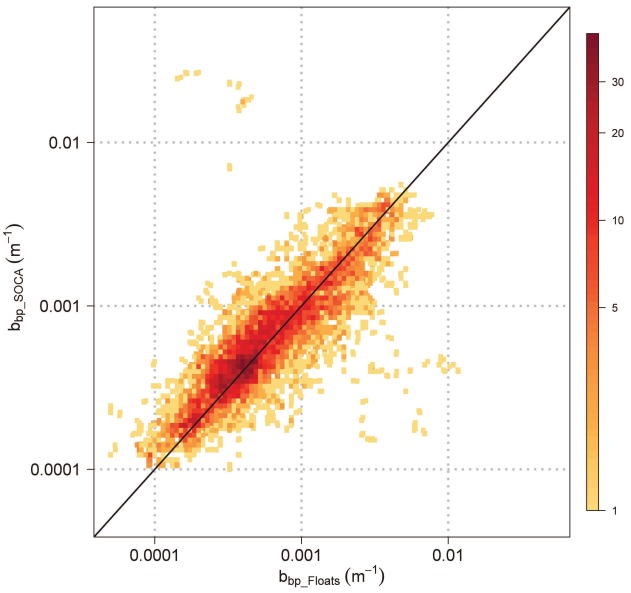


Figure 6. Predictive skills of SOCA-BBP when using as satellite input VIIRS-derived products instead of MODIS-Aqua products. Comparison of b_{bp} retrieved by SOCA-BBP (b_{bp_SOCA}) to reference b_{bp} measurements acquired by Bio-Argo floats (b_{bp_Floats}) with the color code indicating the number of data points in each pixel. The 1 :1 line is represented in black.

3.3. Evaluation of method performance

The SOCA-BBP method is validated using independent datasets of b_{bp} profiles acquired by Bio-Argo floats or as part of an AMT field cruise (see later). For each profile, the 10 b_{bp} values simultaneously retrieved by SOCA-BBP (b_{bp_SOCA}), associated with the 10 dimensionless depths taken at regular intervals within the $0 - 1.3\zeta$ layer, are compared to b_{bp} values measured by the Bio-Argo floats (b_{bp_Floats}) or measured during the AMT cruise (b_{bp_AMT}) at each corresponding depth. To evaluate the performance of SOCA-BBP in inferring the vertical distribution of b_{bp}, several statistical parameters are considered. First, the determination coefficient (R^2) and the slope of the linear regression between the log-transformed values of b_{bp_SOCA} and b_{bp_Floats} (or b_{bp_AMT}) are computed. Second, we estimate the model error using the Median Absolute Percent Difference (MAPD, %) calculated as follows :

$$MAPD = median\left(\frac{|b_{bp_SOCA} - b_{bp_Floats}|}{b_{bp_Floats}}\right) \quad (9)$$

Note that b_{bp_Floats} is replaced by b_{bp_AMT} for the validation against b_{bp} profiles from AMT cruise. We also evaluate the sensitivity of SOCA-BBP to uncertainties in the origin of the satellite data used as input to the MLP. For this purpose, a test is performed which consists in replacing the MODIS-Aqua-derived bio-optical products by VIIRS-derived products.

4. Results and Discussion

4.1. Retrieval of the vertical distribution of the particulate backscattering coefficient

Using the validation database (i.e. 20% of our initial database), the ability of the method is evaluated through a comparison of the 10 values retrieved from SOCA-BBP

Table 3. Statistics of the comparison of the b_{bp} values predicted by SOCA-BBP to Bio-Argo reference measurements. Determination coefficient (R^2) and slope of the linear regression between the retrieved and reference values. The MAPD (Median Absolute Percent Difference) between the retrieved and reference values is also indicated.

	R^2	Slope	MAPD (%)
Validation dataset	0.78	0.8	21
Independent dataset : total	0.78	0.8	22
Independent dataset : Mediterranean Sea	0.71	0.7	30
Independent dataset : North Atlantic Subtropical Gyre	0.81	0.81	12
Independent dataset : Austral Ocean	0.72	0.84	23
Independent dataset : North Atlantic Subpolar Gyre	0.85	0.85	21

Table 4. Statistics of the comparison of the b_{bp} values predicted by SOCA-BBP using MODIS-Aqua or VIIRS-derived derived products as input. Determination coefficient (R^2) and slope of the linear regression between the retrieved and reference values. The MAPD (Median Absolute Percent Difference) between the retrieved and reference values is also indicated.

Type of satellite products	R^2	Slope	MAPD (%)
MODIS-Aqua	0.79	0.79	21
VIIRS	0.72	0.78	21

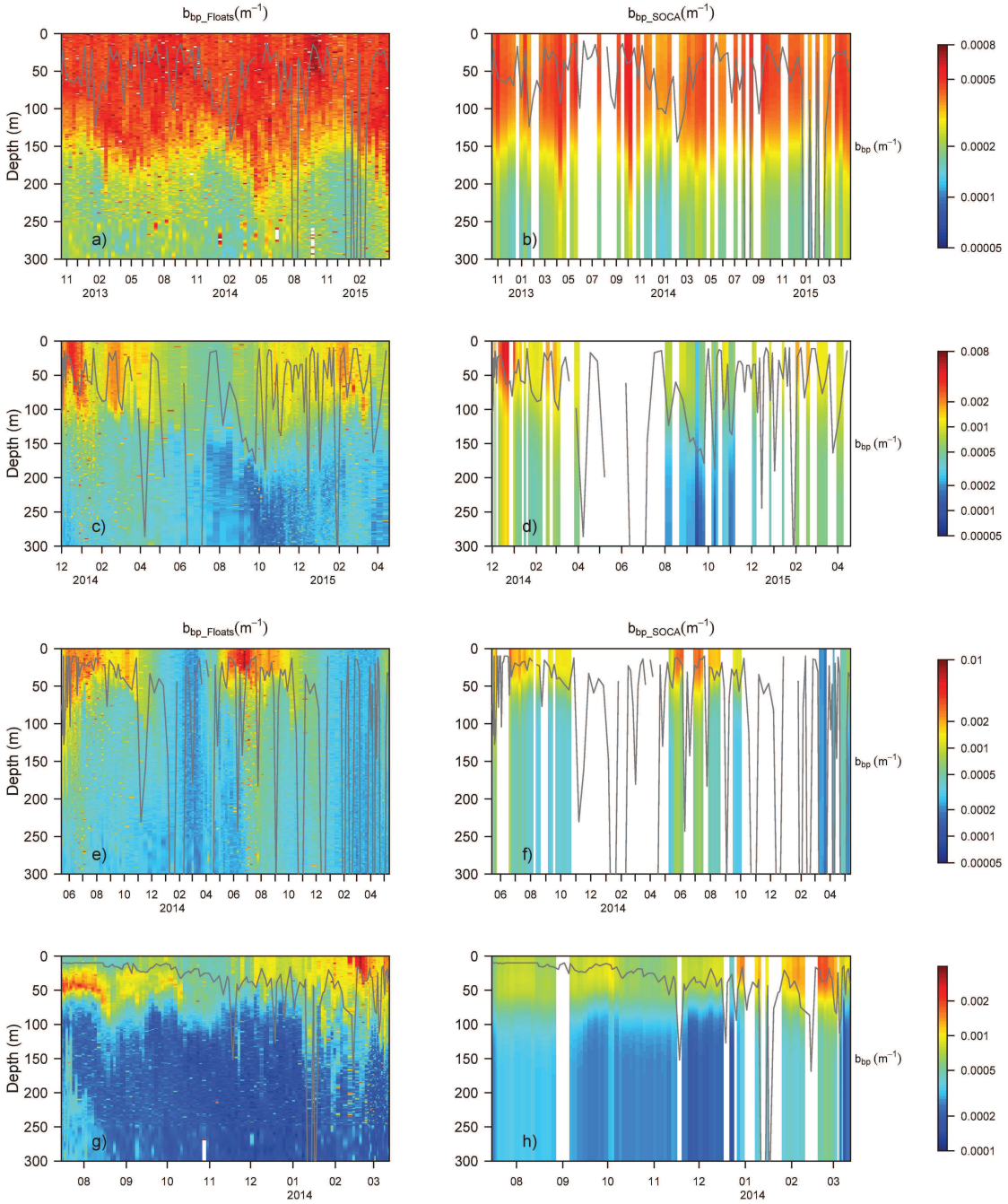


Figure 7. Comparison of the reference b_{bp} measurements acquired by Bio-Argo floats, b_{bp_Floats} (a, c, e and g) with the values predicted by SOCA-BBP, b_{bp_SOCA} (b, d, f and h). Time series for the Bio-Argo floats deployed in the North Atlantic Subtropical Gyre (a-b, WMO=6901472), in the Southern Ocean (c-d, WMO=6901493), in the North Atlantic Subpolar Gyre (e-f, WMO=6901523) and in the Mediterranean Sea (g-h, WMO=6901496). The WMO numbers are official numbers of the World Meteorological Organization. The location of time series for each float is represented in orange in Figure 1. The grey line in each panel indicates the depth of the mixed layer.

(b_{bp_SOCA}) with corresponding values measured by the Bio-Argo floats (b_{bp_Floats}). The scatterplot of b_{bp_SOCA} versus b_{bp_Floats} reveals that SOCA-BBP predicts b_{bp} without systematic bias (i.e. global error of retrieval of 21% ; see Figure 5a and Table 3). The density of the data points shows that most of the b_{bp_SOCA} values are retrieved with substantial accuracy and that only a limited number of data points diverge significantly from the 1 :1 line (Figure 5a). We also tested the performance of the method with respect to the vertical dimension (Figure S5). The median of APD is ~20% for each of the ten dimensionless output depths. The APD appears somewhat lower for the $0-0.7\zeta$ layer suggesting that the method performs slightly better for the upper layers. For the deep layers, the b_{bp} values are very low, which may lead to large errors even when the difference between the predicted and reference values is minor. Nevertheless, the APD remains still low for these deep layers (~22% for the median).

4.2. Sensitivity of SOCA-BBP to satellite input data

We evaluate the sensitivity of the method to satellite input data by replacing in the validation dataset the MODIS-Aqua products by VIIRS-derived products. The VIIRS-to-Argo matchups are computed using VIIRS images with a 4-km resolution and 8-day binning period (9-km resolution as for MODIS-Aqua not available). Among the 886 Bio-Argo profiles of the validation dataset, 649 profiles had concomitant VIIRS and MODIS-Aqua products available. The sensitivity of the method to both types of satellite input data is therefore evaluated using these 649 profiles (see statistics in Table 4). Although the use of VIIRS data seems to slightly reduce the skills of the method (i.e. decrease in the determination coefficient by 0.07), the scatterplot of b_{bp_SOCA} (using VIIRS data as input) versus b_{bp_Floats} shows that the data points are still fairly well scattered around the line 1 :1 (Figure 6). The accuracy of the method remains satisfactory when VIIRS data are used as input (Table 4). Finally, SOCA-BBP appears robust to reasonable noise in the input satellite data. Use of VIIRS- instead of MODIS-Aqua-derived products yield accurate results (global retrieval error of 21% ; Table 4) despite a MAPD in the VIIRS products compared to the MODIS-Aqua products of 44% and 15% for b_{bp} and Chl, respectively.

4.3. Additional validation of SOCA-BBP using independent datasets

Time series of the vertical profile of the particulate backscattering coefficient collected from four Bio-Argo profiling floats deployed in several oceanic basins (Southern Ocean, North Atlantic Subtropical Gyre, North Atlantic Subpolar Gyre, North Western Mediterranean Sea) were removed from the initial database to obtain an independent dataset for further validation of SOCA-BBP (see Sect. 2.2). A comparison of b_{bp_SOCA} with the corresponding b_{bp_Floats} in each basin is presented in Figure 5b. This comparison suggests that the method has similar accuracy when tested with this independent dataset (Table 3) as with the validation dataset comprising 20% of the initial database. The skill of the method is slightly reduced in the Mediterranean Sea compared to other areas (increase of MAPD by $\sim$ 8%).

To obtain fully depth-resolved vertical profiles of b_{bp}, we applied a linear interpolation between each of the 10 b_{bp_SOCA} values provided by the MLP. The resulting predicted time series are compared to their float counterparts for the four basins (Figure 7). We note that the absence of b_{bp_SOCA} data (white bands in Figures 7b, d, f and h) reflect missing satellite-to-Argo matchups caused by a lack of satellite image in cloudy areas/seasons (i.e. usually high latitude environments in winter). The vertical patterns of b_{bp} predicted by SOCA-BBP are very consistent with those observed by the profiling floats in the 0-300 m layer. For the North Atlantic Subtropical Gyre, an area with extremely low b_{bp} values, SOCA-BBP reproduces the seasonal deepening of the b_{bp} maximum in spring and early summer and the shoaling in June-July (Figures 7a-b). For the Southern Ocean, the b_{bp_SOCA} values are consistent with the Bio-Argo float measurements with respect to an increase of b_{bp} from December to April in the 0-100 m layer (Figures 7c and d). In the North Atlantic Subpolar Gyre, the retrieved b_{bp} values agree with float measurements for both years of the time series (i.e. 2013 and 2014, Figures 7e and f). SOCA-BBP seems to underestimate b_{bp} in the Mediterranean Sea, especially from July to October when a deep b_{bp} maximum develops at ~50 m (Figures 7g and h). This underestimation is possibly caused by the relatively coarse vertical resolution of the SOCA-BBP outputs. The b_{bp} maximum may be missed by the 10 output depths from which the entire vertical profile is derived. Finally, a comparison of bbp values integrated within the $0 - Z_m$ layer estimated from SOCA-BBP and measured by the Bio-Argo floats show good agreement for the four examined areas (R^2 of 0.92 and MAPD of 20% ; Figure 8a). Yet, this result is not surprising because the four floats used for this additional validation exercise, although not used for the training process, were deployed in the same areas as the floats represented in the training and validation datasets (see Figure 1).

The performance of the method was also evaluated using a totally independent set of data from a Atlantic Meridional Transect (AMT) cruise conducted in 2009. The geolocation of the 16 profiles of temperature, salinity and b_{bp} with concomitant MODIS-Aqua-derived products is shown in Figure 1 (dark blue crosses). The SOCA-BBP retrieved b_{bp} values, b_{bp_SOCA}, were compared to the reference b_{bp} ship-based measurements integrated within the $0 - Z_m$ layer, b_{bp_AMT} (Figure 8b). Similar to previous validation tests, the results appear highly satisfactory (i.e. median absolute percent difference of 18%). In addition, this validation exercise demonstrates that SOCA-BBP may be applicable to conditions/regions not included in the training database (e.g. Atlantic Equatorial Zone).

4.4. Potential applications of the SOCA-BBP method

4.4.1. Developing climatologies of b_{bp}

Before the emergence and use of profiling floats equipped with backscattering sensors, vertical profiles of b_{bp} in the ocean were quite scarce and highly heterogeneous. SOCA-BBP provides a way to estimate vertical profiles of b_{bp} using basic ocean color products merged with Argo data. A natural application of this method is the development of depth-resolved b_{bp} climatologies. As an example, we develop a climatology of b_{bp} for the global ocean for the months of June and December. We use as input satellite-based monthly composites of b_{bp} and Chl. These data are merged with monthly temperature and salinity data from the Argo global climatologies [Roemmich and Gilson, 2009]. As the Mediterranean sea is not represented in the Argo climatology, we use climatological data from the World Ocean Database [WOD, Levitus et al., 2013] for this basin. We compare the b_{bp_SOCA} surface values with the corresponding satellite estimates, with the b_{bp_SOCA} surface values defined as b_{bp} averaged within the layer comprised between the surface and the penetration depth, $Z_{pd} = Z_e/4.6$ [Morel and Berthon, 1989]. This comparison is conducted at a 1°-resolution (i.e. Argo climatology resolution). Overall, the geographical patterns of the SOCA-BBP retrieved surface b_{bp} values for the months of June and December (Figures 9a and b) are consistent with those observed by the satellite (Figures 9c and d). In June high values of b_{bp} are consistently recorded in the high latitude regions of the northern hemisphere. Reciprocally, high values of b_{bp} are recorded in the high latitudes of the southern hemisphere in December. Unsurprisingly, the equatorial band, the upwelling systems associated with Eastern Boundary Currents and other near-coastal areas (Figure 9a) show high b_{bp} values with weak seasonal variability. SOCA-BBP yields low b_{bp} values compared to satellite-based estimates in the subtropical gyres (Figure 9e). Figure 10 shows a comparison of the surface b_{bp} values retrieved by SOCA-BBP and derived from MODIS-Aqua climatological products for the month of June. We observe a general bias between the SOCA-retrieved surface b_{bp} values and the corresponding satellite values. This observation might result either from a global overestimation of the predicted b_{bp}, or from a global underestimation of b_{bp_MODIS} derived from the QAA model. The SOCA-retrieved b_{bp} versus VIIRS-estimated climatological b_{bp} for the month of June is displayed in Figure S6 for a comparison with Figure 10. There is no general bias between the SOCA-BBP retrieved and satellite-derived values using the VIIRS matchup data.

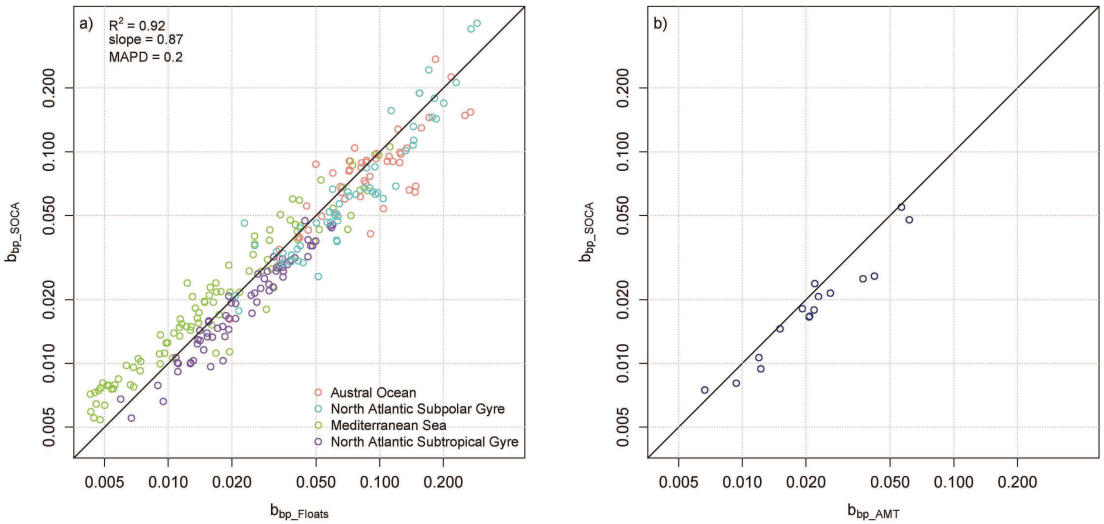


Figure 8. Comparison of b_{bp} integrated within the $0-Z_m$ layer (scale given in optical thickness) predicted by the SOCA-BBP method (b_{bp_SOCA} dimensionless) and calculated from the reference measurements collected by the Bio-Argo floats (b_{bp_Floats} dimensionless) or during an AMT cruise (b_{bp_AMT} dimesnionless). This comparison makes use of reference measurements from two independent datasets : (a) collected by the four Bio-Argo floats not represented in the training and validation databases of the MLP with the color code indicating the float deployment basins ; and (b) acquired during an AMT oceanographic cruise in 2009. The identity line is shown in black in each panel.

This suggests that the bias observed in Figure 10 is probably caused by an underestimation of b_{bp} using the QAA algorithm with satellite MODIS-Aqua products.

Besides the bias between the two estimates of b_{bp}, Figure 10 also indicates that the range of the b_{bp} values retrieved from SOCA-BBP is broader than the range of the satellite-based estimates (when taking into account pixels with > 80 points). In particular, b_{bp} levels yield by the model are lower than the satellite-derived values in the subtropical gyres. Using in situ data from the BIOSOPE cruise [Biogeochemistry and Optics South Pacific Experiment ; Claustre et al., 2008], several studies have shown an overestimation of backscattering satellite estimates in the most oligotrophic conditions of the South Pacific Subtropical Gyre [Brown et al., 2008 ; Huot et al., 2008]. In fact, inherent optical properties are very difficult to estimate from satellite-based measurements in these extremely clear waters and it is now acknowledged that semi-analytical algorithms of b_{bp} retrieval from satellite data lead to a systematic overestimation [e.g. Brown et al., 2008 ; Lee and Huot, 2014]. Interestingly, the model improves the retrieval of the surface b_{bp} in the subtropical gyres (Figures 9 and 10), even though satellite-based surface b_{bp} estimates used as input show overestimation. This is principally because the learning of the neural network is based on accurate Bio-Argo in situ b_{bp} profiles hence constraining the retrieved b_{bp} surface values. In addition, as the subtropical gyres are characterized by low Chl, the results shown in this study are consistent with the formulations of Morel and Maritorena [2001] and Huot et al. [2008] that account for a continuous decrease in b_{bp} with decreasing chlorophyll a concentrations for low chlorophyll a concentrations ($< 0.1\ mg\ m^{-3}$) instead of the constant b_{bp} with decreasing chlorophyll a concentrations as reported by Behrenfeld et al. [2005].

4.4.2. Estimating POC and phytoplankton carbon from b_{bp}

In addition to the production of climatologies of the vertical distribution of b_{bp}, the subsequent conversion of b_{bp} into POC or phytoplankton carbon has the potential to support investigations dedicated to carbon cycle, including carbon production and export. Several published relationships link POC to b_{bp} either regionally or at the global scale [Balch et al., 2001 ; Cetinić et al., 2012 ; Loisel et al., 2001, 2002 ; Stramski et al., 1999, 2008]. A systematic and routine acquisition of b_{bp} vertical profile has started only recently (a decade ago) so that the number of concurrent b_{bp} and POC measurements for establishing robust regional or global b_{bp}-to-POC relationships is still limited. Collecting systematic measurements of POC and b_{bp} is of critical need for refining the previously published relationships and, ultimately, exploit in an optimal manner the growing b_{bp} dataset acquired by Bio-Argo floats and the SOCA-BBP climatological products.

Apart from deriving POC, recent studies have highlighted the potential of b_{bp} as a phytoplankton carbon proxy [Graff et al., 2015]. Actually b_{bp} might be a more reliable proxy of phytoplankton carbon than Chl or POC. Hence, using the b_{bp}-to-phytoplankton carbon relationships presently available in the literature in combination with SOCA-BBP, it appears possible to propose global estimates of the vertical distribution of phytoplankton carbon with high space-time resolution. A potential consequence of obtaining improved estimates of the phytoplankton biomass is a possible reassessment of the sources of variability in the Chl. Using phytoplankton carbon estimates derived from satellite-based data of b_{bp}, some studies have indeed shown that temporal changes in Chl over large oceanic regions may be predominantly caused by physiologically-driven modifications in the cellular Chl-to-carbon ratio rather than by actual changes in phytoplankton biomass [Behrenfeld et al., 2005, 2009 ; Siegel et al., 2013]. The combination of SOCA-BBP with other methods, which infer the vertical distribution of Chl from space [Sauzède et al., In prep. ; Uitz et al., 2006], could permit the variability in the phytoplankton carbon-to-Chl relationship to be examined over the vertical dimension. This would represent a significant step towards a better understanding of light and nutrient control of phytoplankton biomass and physiological status, a perequiste for improving the characterization of the distribution and variability in primary production and carbon export.

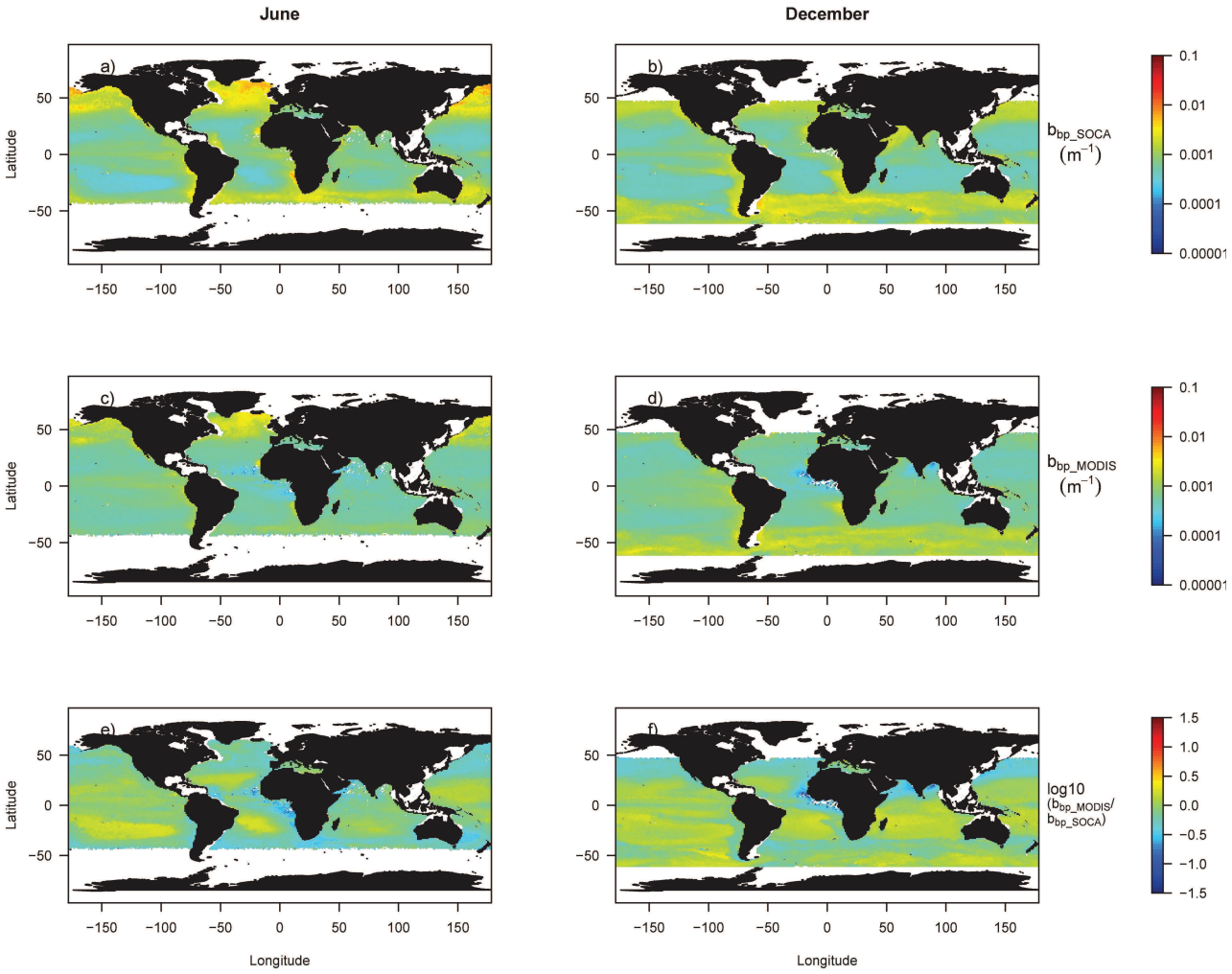


Figure 9. Surface climatology of the particulate backscattering coefficient with a 1 ° resolution for the month of June and December (left and right panels, respectively). (a-b) Surface b_{bp} (i.e. averaged over the $0 - Z_{pd}$ layer) obtained from the SOCA-BBP algorithm ; (c-d) MODIS-Aqua-derived estimates of b_{bp} ; (e-f) log_{10} ratio of the satellite-based b_{bp} to the SOCA-BBP-retrieved b_{bp}.

5. Conclusion and perspectives

We have demonstrated that, using a Multi-Layered Perceptron method, we can merge ocean color-based products with temperature-salinity Argo data to infer the vertical distribution of a bio-optical property estimated from both satellite and robotic platform measurements. The proposed method, SOCA-BBP, infers the vertical distribution of the particulate backscattering coefficient using three main input components : (1) a surface component, i.e., satellite-derived products ; (2) a vertical component derived from temperature and salinity profiles measured by Argo floats ; and (3) a temporal component, i.e., the day of the year of the considered satellite-to-Argo matchup. Because the training of the MLP-based method was conducted using a dataset representative of the hydrologic and biogeochemical conditions prevailing in the global open ocean, the method is expected to be applicable to most open-ocean environments. Nevertheless, we note that SOCA-BBP has not been developed for applications on a profile-per-profile basis, where a single satellite-to-Argo matchup associated with a specific day would be used to retrieve an accurate vertical profile of b_{bp}. Instead SOCC-BBP should be considered as a method dedicated to relatively large-scale applications, such as the development of climatological products (see, e.g., Sect. 4.4).

The natural variability of the vertical distribution of b_{bp} makes the prediction of this bio-optical parameter challenging. Compared to the reference measurements acquired by the Bio-Argo floats from the training dataset (i.e. used to establish the SOCA-BBP underlying relationship), the retrieved b_{bp_SOCA} values from the training dataset (i.e. used to establish the underlying relationship of the MLP) are retrieved with a median absolute percent difference of 19% (i.e. intrinsic error of the model). Therefore, the error of SOCA-BBP in retrieving the vertical distribution of b_{bp} (i.e. 21%) seems to be essentially induced by the natural variability of b_{bp}. The uncertainties associated with ocean color-based bio-optical products may also generate additional uncertainties in the retrieval of b_{bp}. Our analysis of the sensitivity of SOCA-BBP to the satellite data used as input indicates limited changes in the prediction of b_{bp} as one uses VIIRS products instead of MODIS-Aqua products (Figure 6 and Table 4). This suggests that the MLP is relatively insensitive to reasonable noise levels in the input satellite data because noise is accounted for in the training of the MLP. Based on this sensitivity analysis, we expect that the method can be safely used with satellite products other than those derived from MODIS-Aqua (e.g. SeaWiFS, MERIS, VIIRS, OLCI) or with merged products (e.g. GlobColour, CCI-OC).

The present study provides an invaluable source of information on the vertical distribution of the b_{bp} allowing

this key bio-optical property to be comprehensively described at a global scale. A major application of the method is obviously linked to the creation of a depth- resolved global proxy of POC and, possibly, phytoplankton carbon with high space-time resolution. This is a prerequisite for improving the characterization and quantification of key carbon fluxes such as net primary production or export fluxes. In particular, the data resulting from SOCA-BBP are valuable for the initialization or validation of biogeochemical models. The climatological data retrieved from SOCA-BBP also have the potential to serve as benchmarks against which temporal or regional trends could be evidenced.

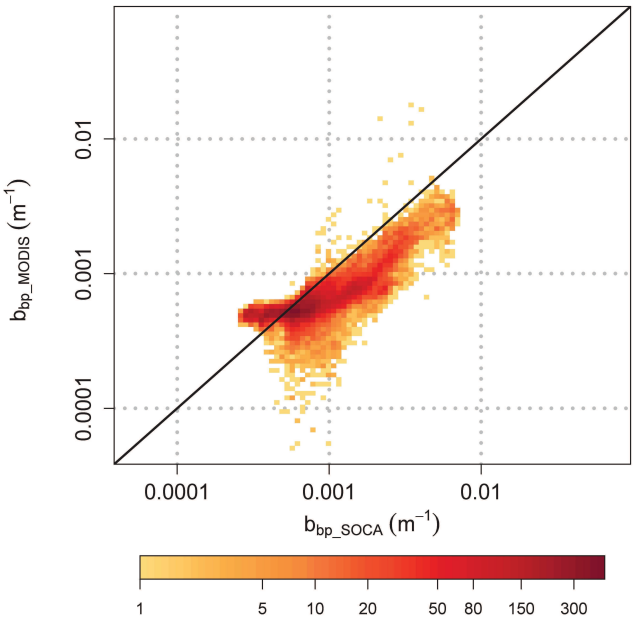


Figure 10. Comparison of the global climatological values of surface b_{bp} retrieved by SOCA-BBP (b_{bp_SOCA}) to those derived from MODIS-Aqua data (b_{bp_MODIS}) for the month of June with the color code indicating the number of data points in each pixel. Both b_{bp_SOCA} and b_{bp_MODIS} are scaled to a resolution of 1°.

Along with the progressive development of a global Bio-Argo program and associated float deployments, additional measurements of concurrent density and b_{bp} profiles will help to improve the relationship established in the MLP. This is especially expected for the regions currently under-sampled in the Bio-Argo database used in this study (e.g. Indian Ocean Gyre, Arctic Ocean). It is indeed important to stress out the evolving aspect of this database and of the quality of the products that can be retrieved from it. This study has shown that neural network-based methods can link the vertical distribution of a given bio-optical property (i.e. particulate backscattering coefficient) to the corresponding near-surface value merged with vertically resolved physical properties. The development of analogous methods for other bio-optical properties, measured from both Bio-Argo floats and ocean color satellites (e.g. chlorophyll *a* concentration, CDOM), appears as a natural extension of the present study [Sauzède et al., In prep.].

Acknowledgments. This paper is a contribution to the Remotely Sensed Biogeochemical Cycles in the Ocean (remOcean) project, funded by the European Research Council (grant agreement 246777), to the ATLANTOS EU project (grant agreement 2014-633211) funded by H2020 program, to the French Bio-Argo project funded by CNES-TOSCA and to the French Equipement davenir NAOS project (ANR J11R107-F). This study is a contribution to the international IMBER project and was supported by the UK Natural Environment Research Council National Capability funding to Plymouth Marine Laboratory and the National Oceanography Centre, Southampton. This is contribution number 276 of the AMT program. GDO acknowledges funding from the UK National Centre for Earth Observation. Density and particulate backscattering coefficient profiles acquired by Bio-Argo floats used in this study are freely available (http ://www.coriolis.eu.org). Temperature and salinity data were collected and made freely available by the International Argo Program and the national programs that contribute to it (http ://www.argo.ucsd.edu, http ://argo.jcommops.org). The Argo Program is part of the Global Ocean Observing System. The authors deeply acknowledge the NASA for MODIS-Aqua and VIIRS imagery. The International Argo Program and the World Ocean Atlas are acknowledged for making freely available data that have allowed computation of the temperature and salinity climatologies used for our application.

References

Balch, W. M., D. T. Drapeau, J. J. Fritz, B. C. Bowler, and J. Nolan (2001), Optical backscattering in the Arabian Seacontinuous underway measurements of particulate inorganic and organic carbon, Deep Sea Res. Part I Oceanogr. Res. Pap., 48(11), 24232452, doi :10.1016/S0967-0637(01)00025-5.

Behrenfeld, M. J., and E. Boss (2003), The beam attenuation to chlorophyll ratio : An optical index of phytoplankton physiology in the surface ocean ?, Deep. Res. Part I Oceanogr. Res. Pap., 50(12), 15371549, doi :10.1016/j.dsr.2003.09.002.

Behrenfeld, M. J., and E. Boss (2006), Beam attenuation and chlorophyll concentration as alternative optical indices of phytoplankton biomass, J. Mar. Res., 64(3), 431451.

Behrenfeld, M. J., E. Boss, D. A. Siegel, and D. M. Shea (2005), Carbon-based ocean productivity and phytoplankton physiology from space, Global Biogeochem. Cycles, 19(1), n/an/a, doi :10.1029/2004GB002299. [online] Available from : http ://doi.wiley.com/10.1029/2004GB002299 (Accessed 21 January 2014)

Behrenfeld, M. J. et al. (2009), Satellite-detected fluorescence reveals global physiology of ocean phytoplankton, Biogeosciences, 6(5). [online] Available from : http ://escholarship.org/uc/item/9fh4s430 (Accessed 28 July 2015)

Bishop, C. M. (1995), Neural Networks for Pattern Recognition, Oxford University Press, Inc. [online] Available from : http ://dl.acm.org/citation.cfm ?id=525960 (Accessed 20 March 2014)

Bishop, J. K. B. (2009), Autonomous observations of the ocean biological carbon pump, Lawrence Berkeley Natl. Lab.

Bishop, J. K. B., and T. J. Wood (2009), Year-round observations of carbon biomass and flux variability in the Southern Ocean, Global Biogeochem. Cycles, 23(2), GB2019, doi :10.1029/2008GB003206.

Boss, E., and W. S. Pegau (2001), Relationship of Light Scattering at an Angle in the Backward Direction to the Backscattering Coefficient, Appl. Opt., 40(30), 5503, doi :10.1364/AO.40.005503.

De Boyer Montgut, C., G. Madec, A. S. Fischer, A. Lazar, and D. Iudicone (2004), Mixed layer depth over the global ocean : An examination of profile data and a profile-based climatology, J. Geophys. Res., 109(C12), C12003, doi :10.1029/2004JC002378.

Bricaud, A., C. Mejia, D. Blondeau-Patissier, H. Claustre, M. Crepon, and S. Thiria (2007), Retrieval of pigment concentrations and size structure of algal populations from their absorption spectra using multilayered perceptrons., Appl. Opt., 46(8), 125160.

Brown, C., Y. Huot, P. WerdellL, B. Gentili, and H. Claustre (2008), The origin and global distribution of second order variability in satellite ocean color and its potential applications to algorithm development, Remote Sens. Environ., 112(12), 41864203, doi :10.1016/j.rse.2008.06.008.

Cetinić, I., M. J. Perry, N. T. Briggs, E. Kallin, E. A. DAsaro, and C. M. Lee (2012), Particulate organic carbon and inherent optical properties during 2008 North

Atlantic Bloom Experiment, J. Geophys. Res., 117(C6), C06028, doi :10.1029/2011JC007771. [online] Available from : http ://adsabs.harvard.edu/abs/2012JGRC..117.6028C (Accessed 9 February 2015) Claustre, H., A. Sciandra, and D. Vaulot (2008), Introduction to the special section Bio-optical and biogeochemical conditions in the South East Pacific in late 2004 : the BIOSOPE program, Biogeosciences Discuss., 5(1), 605640.

Claustre, H. et al. (2010a), Bio-optical profiling floats as new observational tools for biogeochemical and ecosystem studies : potential synergies with ocean color remote sensing., in Proceedings of the OceanObs 09 : Sustained Ocean Observations and Information for Society Conference (Vol.2), edited by J. Hall, D. E. Harrison, and D. Stammer, ESA Publ., Venice, Italy.

Claustre, H. et al. (2010b), Guidelines towards an integrated ocean observation system for ecosystems and biogeochemical cycles, in Proceedings of the OceanObs 09 : Sustained Ocean Observations and Information for Society Conference (Vol.1), edited by J. Hall, D. E. Harrison, and D. Stammer, ESA Publ., Venice, Italy.

DallOlmo, G., and K. A. Mork (2014), Carbon export by small particles in the Norwegian Sea, Geophys. Res. Lett., 41(8), 29212927, doi :10.1002/2014GL059244.

Dufort-Gaurier, L., H. Loisel, D. Dessailly, K. Nordkvist, and S. Alvain (2010), Estimates of particulate organic carbon over the euphotic depth from in situ measurements. Application to satellite data over the global ocean, Deep Sea Res. Part I Oceanogr. Res. Pap., 57(3), 351367, doi :10.1016/j.dsr.2009.12.007.

Falkowski, P. G., T. R. Barber, and V. Smetacek (1998), Biogeochemical Controls and Feedbacks on Ocean Primary Production, Science (80-.)., 281(5374), 200206, doi :10.1126/science.281.5374.200.

Friedrich, T., and a. Oschlies (2009), Neural network-based estimates of North Atlantic surface pCO 2 from satellite data : A methodological study, J. Geophys. Res., 114(C3), C03020, doi :10.1029/2007JC004646.

Gardner, W. D., A. V. Mishonov, and M. J. Richardson (2006), Global POC concentrations from in-situ and satellite data, Deep Sea Res. Part II Top. Stud. Oceanogr., 53(5-7), 718740, doi :10.1016/j.dsr2.2006.01.029.

Graff, J. R., T. K. Westberry, A. J. Milligan, M. B. Brown, G. DallOlmo, V. van Dongen-Vogels, K. M. Reifel, and M. J. Behrenfeld (2015), Analytical phytoplankton carbon measurements spanning diverse ecosystems, Deep Sea Res. Part I Oceanogr. Res. Pap., doi :10.1016/j.dsr.2015.04.006.

Gross, L., S. Thiria, R. Frouin, and B. G. Mitchell (2000), Artificial neural networks for modeling the transfer function between marine reflectance and phytoplankton pigment concentration, J. Geophys. Res., 105(C2), 3483, doi :10.1029/1999JC900278.

Hornik, K., M. Stinchcombe, and H. White (1989), Multilayer Feedforward Networks are Universal Approximators, Neural Networks, 2, 359366.

Huot, Y., A. Morel, M. S. Twardowski, D. Stramski, and R. A. Reynolds (2008), Particle optical backscattering along a chlorophyll gradient in the upper layer of the eastern South Pacific Ocean, Biogeosciences, 5(2), 495507.

Jamet, C., H. Loisel, and D. Dessailly (2012), Retrieval of the spectral diffuse attenuation coefficient K d ( ) in open and coastal ocean waters using a neural network inversion, J. Geophys. Res., 117(C10), C10023, doi :10.1029/2012JC008076.

Kokhanovsky, A. A. (2012), Light Scattering Reviews 7 : Radiative Transfer and Optical Properties of Atmosphere and Underlying Surface.

Kostadinov, T. S., D. A. Siegel, and S. Maritorena (2010), Global variability of phytoplankton functional types from space : assessment via the particle size distribution, Biogeosciences, 7(10), 32393257, doi :10.5194/bg-7-3239-2010.

Krasnopolsky, V. M. (2009), Neural Network Applications to Solve Forward and Inverse Problems in Atmospheric and Oceanic Satellite Remote Sensing, , 191205.

Lee, Z., and Y. Huot (2014), On the non-closure of particle backscattering coefficient in oligotrophic oceans., Opt. Express, 22(23), 2922333, doi :10.1364/OE.22.029223.

Lee, Z., K. L. Carder, and R. A. Arnone (2002), Deriving Inherent Optical Properties from Water Color : a Multiband Quasi-Analytical Algorithm for Optically Deep Waters, Appl. Opt., 41(27), 5755, doi :10.1364/AO.41.005755.

Lee, Z., B. Lubac, J. Werdell, and R. Arnone (2009), An update of the quasi-analytical algorithm (QAA_v5), Int. Ocean Color Gr. Softw. Rep.

Lek, S., and J. F. Gugan (1999), Artificial neural networks as a tool in ecological modelling, an introduction, Ecol. Modell., 120(2-3), 6573, doi :10.1016/S0304-3800(99)00092-7.

Levitus, S. et al. (2013), The World Ocean Database, Data Sci. J., 12(0), WDS229WDS234. Loisel, H., E. Bosc, D. Stramski, K. Oubelkheir, and P.-Y. Deschamps (2001), Seasonal variability of the backscattering coefficient in the Mediterranean Sea based on satellite SeaWiFS imagery, Geophys. Res. Lett., 28(22), 42034206, doi :10.1029/2001GL013863.

Loisel, H., J.-M. Nicolas, P.-Y. Deschamps, and R. Frouin (2002), Seasonal and inter-annual variability of particulate organic matter in the global ocean, Geophys. Res. Lett., 29(24), 2196, doi :10.1029/2002GL015948.

Loisel, H., J.-M. Nicolas, A. Sciandra, D. Stramski, and A. Poteau (2006), Spectral dependency of optical backscattering by marine particles from satellite remote sensing of the global ocean, J. Geophys. Res., 111(C9), C09024, doi :10.1029/2005JC003367.

Martinez-Vicente, V., G. DallOlmo, G. Tarran, E. Boss, and S. Sathyendranath (2013), Optical backscattering is correlated with phytoplankton carbon across the Atlantic Ocean, Geophys. Res. Lett., 40(6), 11541158, doi :10.1002/grl.50252.

Morel, A., and J.-F. Berthon (1989), Surface Pigments, Algal Biomass Profiles, and Potential Production of the Euphotic Layer : Relationships Reinvestigated in View of Remote-Sensing Applications, Limnol. Oceanogr., 34(8), 15451562.

Morel, A., and S. Maritorena (2001), Bio-optical properties of oceanic waters : A reappraisal, J. Geophys. Res., 106(C4), 7163, doi :10.1029/2000JC000319. [online] Available from : http ://doi.wiley.com/10.1029/2000JC000319 (Accessed 20 March 2014)

Niang, a., F. Badran, C. Moulin, M. Crpon, and S. Thiria (2006), Retrieval of aerosol type and optical thickness over the Mediterranean from SeaWiFS images using an automatic neural classification method, Remote Sens. Environ., 100(1), 8294, doi :10.1016/j.rse.2005.10.005.

Palacz, A. P., M. A. St. John, R. J. W. Brewin, T. Hirata, and W. W. Gregg (2013), Distribution of phytoplankton functional types in high-nitrate, low-chlorophyll waters in a new diagnostic ecological indicator model, Biogeosciences, 10(11), 75537574, doi :10.5194/bg-10-7553-2013.

Raitsos, D. E., S. J. Lavender, C. D. Maravelias, J. Haralabous, A. J. Richardson, and P. C. Reid (2008), Identifying four phytoplankton functional types from space : An ecological approach, Limnol. Oceanogr. Ocean., 53(2), 605613.

Rochford, P. A., A. B. Kara, A. J. Wallcraft, and R. A. Arnone (2001), Importance of solar subsurface heating in ocean general circulation models, J. Geophys. Res., 106(C12), 30923, doi :10.1029/2000JC000355.

Roemmich, D., and J. Gilson (2009), The 20042008 mean and annual cycle of temperature, salinity, and steric height in the global ocean from the Argo Program, Prog. Oceanogr., 82(2), 81100, doi :10.1016/j.pocean.2009.03.004.

Roemmich, D., G. Johnson, S. Riser, R. Davis, J. Gilson, W. B. Owens, S. Garzoli, C. Schmid, and M. Ignaszewski (2009), The Argo Program : Observing the Global Oceans with Profiling Floats, Oceanography, 22(2), 3443, doi :10.5670/oceanog.2009.36.

Rumelhart, D. E., G. E. Hinton, and R. J. Williams (1988), Learning representations by back-propagating errors, Cogn. Model., 5.

Sauzède, R., H. Claustre, C. Jamet, J. Uitz, J. Ras, A. Mignot, and F. DOrtenzio (2015), Retrieving the vertical distribution of chlorophyll a concentration and phytoplankton community composition from in situ fluorescence profiles : A method based on a neural network with potential for global-scale applications, J. Geophys. Res. Ocean., 120(1), 451470, doi :10.1002/2014JC010355.

Sauzède, R., H. Claustre, J. Uitz, C. Jamet, F. DOrtenzio, B. Gentili, C. Schmechtig, and A. Poteau (In prep.), A neural network-based method for merging ocean color and Argo data to extend surface bio-optical properties to depth : chlorophyll a concetration retrieval

Siegel, D. A., S. Maritorena, N. B. Nelson, and M. J. Behrenfeld (2005), Independence and interdependencies among global ocean color properties : Reassessing the bio-optical assumption, J. Geophys. Res., 110(C7), C07011, doi :10.1029/2004JC002527.

Siegel, D. A. et al. (2013), Regional to global assessments of phytoplankton dynamics from the SeaWiFS mission, Remote Sens. Environ., 135, 7791, doi :10.1016/j.rse.2013.03.025.

Stramska, M. (2009), Particulate organic carbon in the global ocean derived from SeaWiFS ocean color, Deep Sea Res. Part I Oceanogr. Res. Pap., 56(9), 14591470, doi :10.1016/j.dsr.2009.04.009.

Stramski, D., R. A. Reynolds, M. Kahru, and B. G. Mitchell (1999), Estimation of Particulate Organic Carbon in the Ocean from Satellite Remote Sensing, Science (80-.)., 285(5425), 239242, doi :10.1126/science.285.5425.239.

Stramski, D. et al. (2008), Relationships between the surface concentration of particulate organic carbon and optical properties in the eastern South Pacific and eastern Atlantic Oceans, Biogeosciences, 5(1), 171201.

Sullivan, J. M., and M. S. Twardowski (2009), Angular shape of the oceanic particulate volume scattering function in the backward direction., Appl. Opt., 48(35), 68119, doi :10.1364/AO.48.006811.

Telszewski, M. et al. (2009), Estimating the monthly pCO2 distribution in the north Atlantic using a self-organizing neural network, Biogeosciences.

Uitz, J., H. Claustre, A. Morel, and S. B. Hooker (2006), Vertical distribution of phytoplankton communities in open ocean : An assessment based on surface chlorophyll, J. Geophys. Res., 111(C8), C08005, doi :10.1029/2005JC003207.

Volk, T., and M. I. Hoffert (1985), The Carbon Cycle and Atmospheric CO2 : Natural Variations Archean to Present, Geophysical Monograph Series, edited by E. T. Sundquist and W. S. Broecker, American Geophysical Union, Washington, D. C.

Westberry, T., M. J. Behrenfeld, D. A. Siegel, and E. Boss (2008), Carbon-based primary productivity modeling with vertically resolved photoacclimation, Global Biogeochem. Cycles, 22(2), n/an/a, doi :10.1029/2007GB003078.

Wong, A., R. Keeley, T. Carval, and the Argo Data Management Team (2014), Argo quality control manual, Tech. Rep. Zhang, X., L. Hu, and M.-X. He (2009), Scattering by pure seawater : Effect of salinity, Opt. Express, 17(7), 5698, doi :10.1364/OE.17.005698.

Corresponding author: R. Sauzède, Laboratoire d'Ocanographie de Villefranche (LOV), FRANCE. (sauzede@obs-vlfr.fr)

Supplementary Material

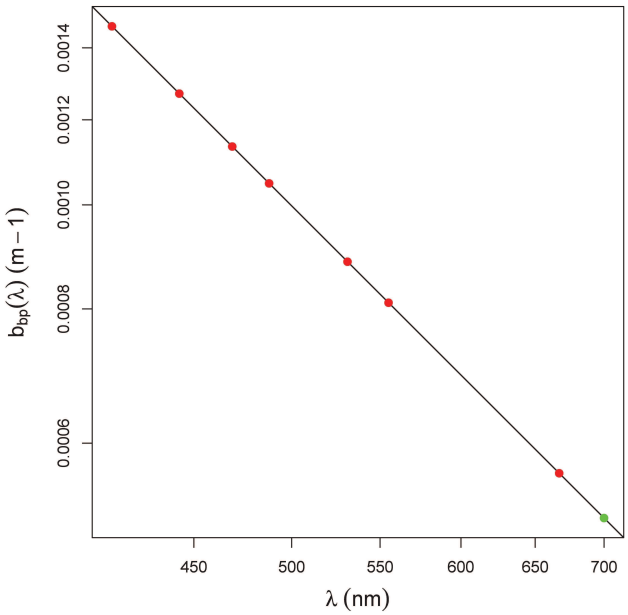


Figure 1. Example of the spectral backscattering coefficient $b_{bp}(\lambda)$ retrieved at 7 wavelengths (red points) from the Quasi-Analytical Algorithm [QAA, see Lee et al., 2002, 2009] applied to MODIS-Aqua R_{rs} data. The green point represent the bbp retrieved at 700 nm using the power law relationship of the 7 QAA-derived bbp (black line).

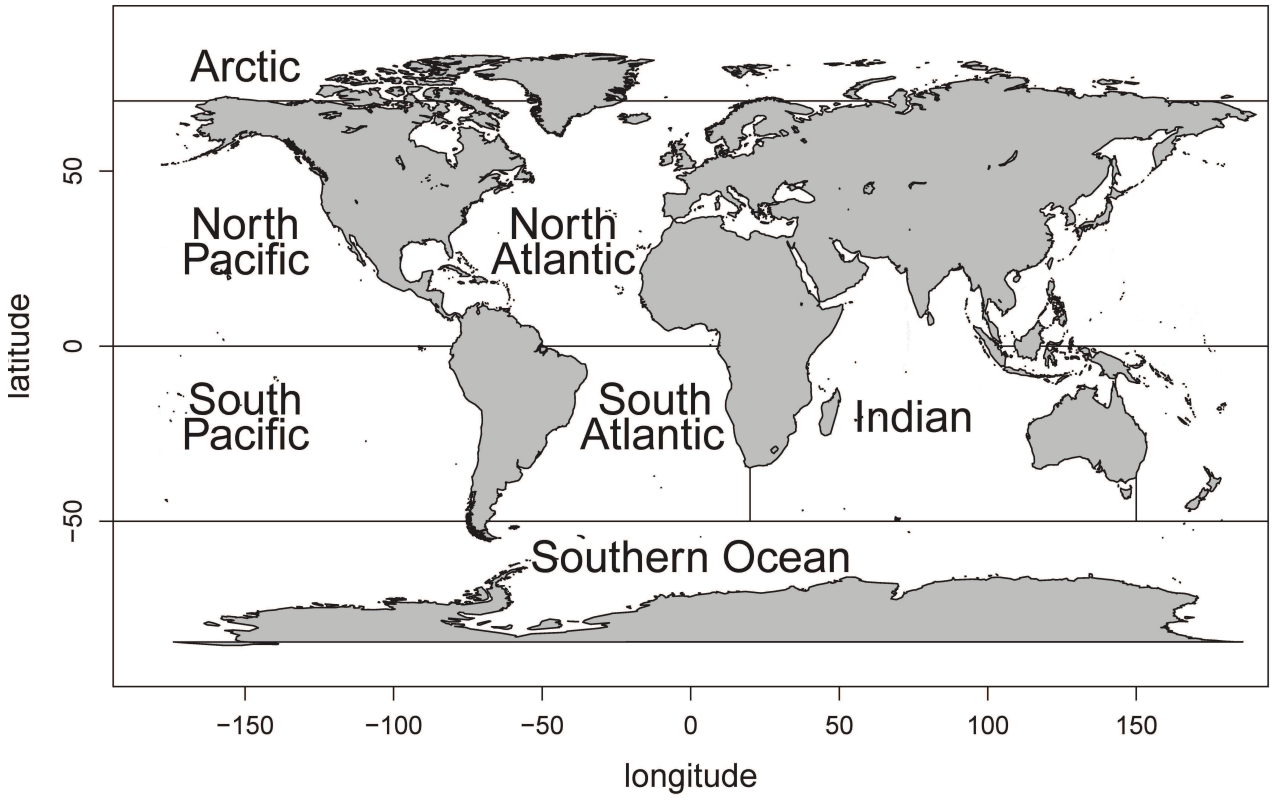


Figure 2. Geographic boundaries of the oceanic basins referred to in Table 2.

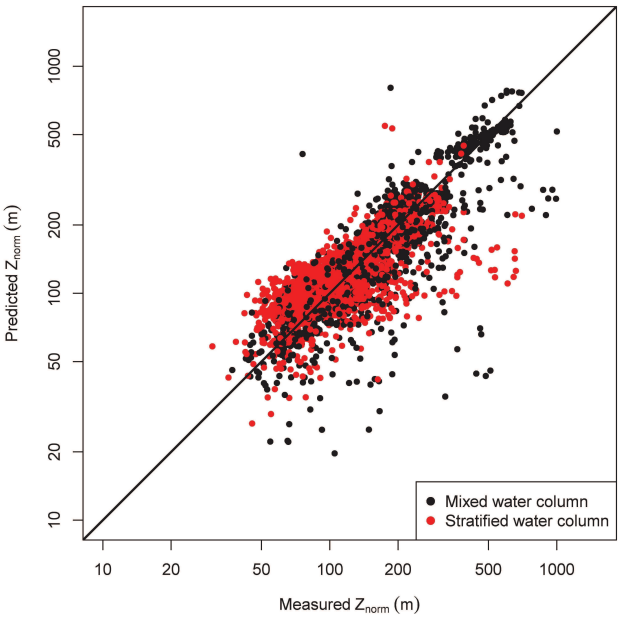


Figure 3. Comparison of the normalization depth, Z_{norm}, predicted with Equations 5 and 6 with the reference value computed from the Bio-Argo fluorescence profiles [see text and Sauzède et al., 2015]. The red dots indicate stratified water conditions (computed with Equation 5) and the black dots indicate mixed water conditions (computed with Equation 6). The 1 :1 line is shown in black.

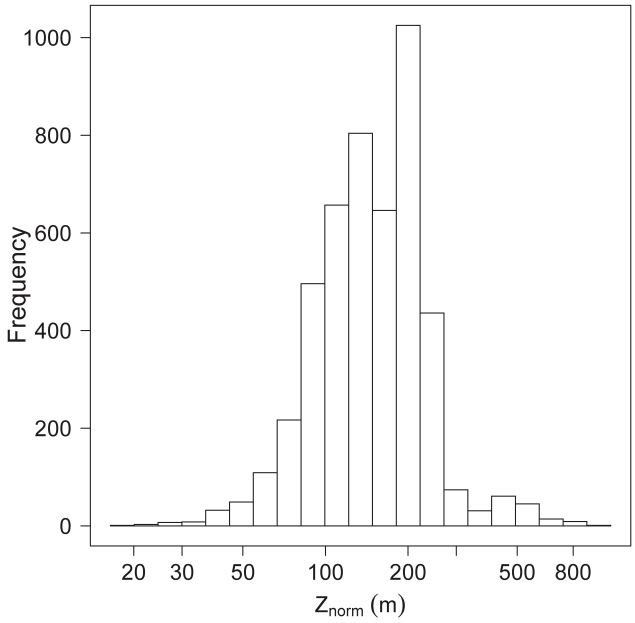


Figure 4. Frequency distribution of the computed dimensionless depth, Z_{norm} (m), for the 4725 stations used in the present study.

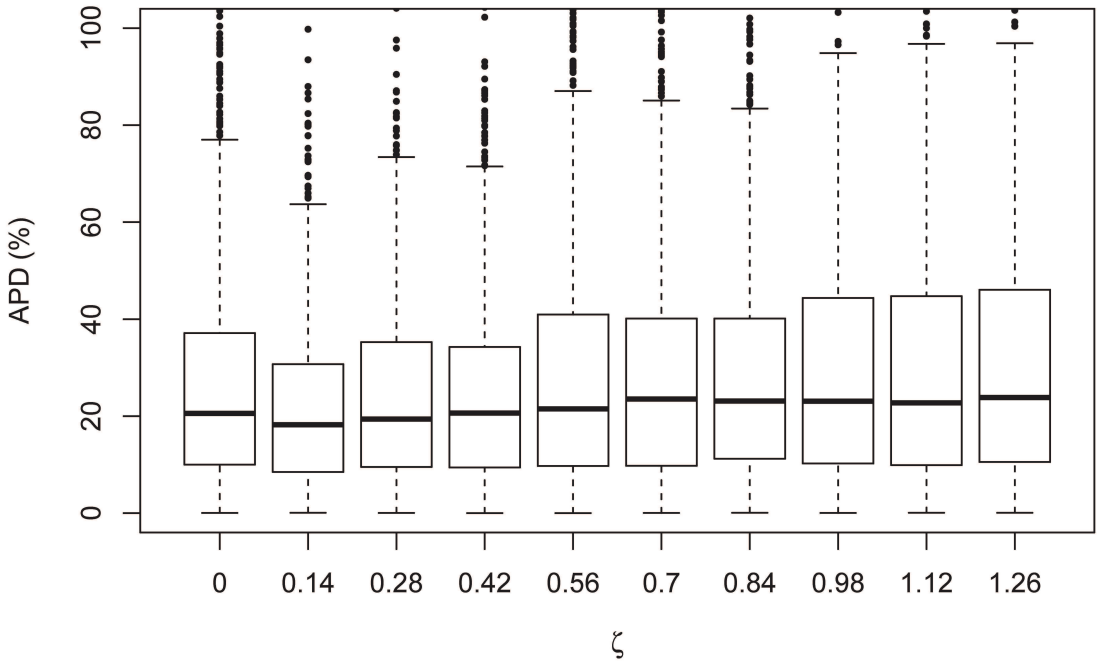


Figure 5. Boxplots illustrating the distribution of the Absolute Percent Difference, APD (%), between the b_{bp} values predicted by SOCA-BBP and the reference Bio-Argo b_{bp} measurements at the ten dimensionless depths, ζ, at which b_{bp} is predicted.

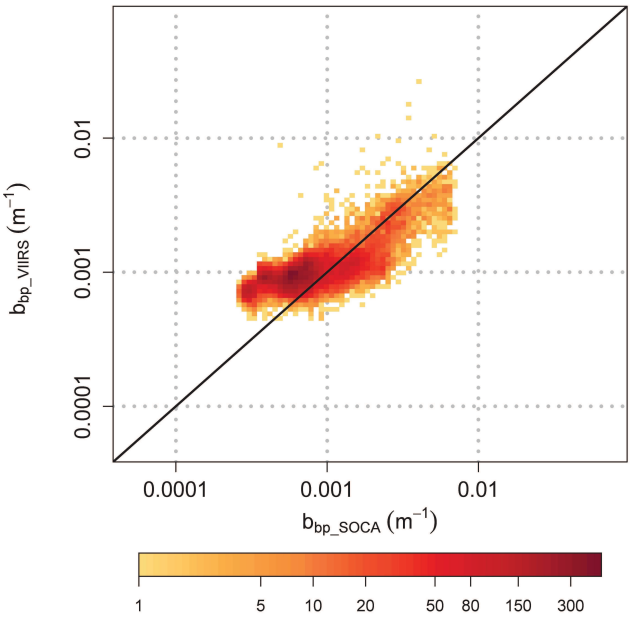


Figure 6. Comparison of the global climatological values of surface b_{bp} retrieved by SOCA-BBP (b_{bp_SOCA}) to those derived from VIIRS data b_{bp_VIIRS}) for the month of June with the color code indicating the number of data points in each pixel. Both b_{bp_SOCA} and b_{bp_VIIRS} are scaled to a resolution of 1 $^\circ$.

3.2.3 Conclusions et perspectives de l'étude

De manière générale, cette étude a permis de montrer qu'en utilisant un réseau de neurones il était possible de dériver la distribution verticale du b_{bp} à partir de produits satellitaires co-localisés à des profils verticaux hydrologiques collectés par flotteurs Argo avec une erreur globale d'environ 20%.

Même si les données utilisées pour entrainer le réseau de neurones de SOCA-BBP sont représentatives des conditions hydrologiques et biogéochimiques pouvant être rencontrées dans la plupart des eaux océaniques, les nouvelles mesures acquises dans le futur par les flotteurs Bio-Argo permettront surement **d'améliorer** les relations sous-jacentes à notre modèle. Notre méthode a donc un caractère **évolutif** qu'il faut mettre en avant puisque de plus en plus de profils bio-optiques sont attendus grâce aux plateformes autonomes.

Nous avons évalué la **sensibilité** de SOCA-BBP aux données satellites d'entrée. En effet, nous avons pu voir qu'en utilisant les produits issus de VIIRS au lieu de MODIS Aqua, la méthode était relativement insensible à des niveaux de bruits raisonnables des données satellites. Le bruit des données d'entrées est en effet surement pris en compte pendant l'entrainement puisque les réseaux de neurones sont connus pour réussir à modéliser des relations très bruitées. Cette analyse suggère qu'il est possible d'utiliser SOCA-BBP avec des produits satellitaires issus **d'autres satellites** (e.g. SeaWiFS, VIIRS, MERIS, OLCI) ou bien également de produits satellitaires **fusionnés** (e.g. GlobColour, CCI-OC) sans trop détériorer la performance de la méthode.

Nous avons pu montrer que cette méthode semblait **améliorer les produits satellites** de b_{bp} dans certaines régions océaniques oligotrophiques comme le gyre subtropical du Pacifique Sud (voir Figure 9 de l'article soumis ci-dessus). En effet, les IOPs (propriétés optiques inhérentes) sont très difficiles à estimer dans les **eaux claires et oligotrophiques**. Plusieurs études ont montré que le b_{bp} était largement surestimé par les mesures satellitales dans le gyre du Pacifique Sud (Brown *et al.*, 2008, Huot *et al.*, 2008). Notre méthode permettrait donc, grâce à son entrainement basé sur des profils *in situ* de b_{bp} collectés par les flotteurs Bio-Argo, de corriger le biais des estimations satellitales de b_{bp} en surface.

Il a été montré que le b_{bp} était un bon estimateur du POC (Balch *et al.*, 2001, Cetinić *et al.*, 2012b, Loisel *et al.*, 2001, 2002, Stramski *et al.*, 1999, 2008) ainsi que du carbone phytoplanctonique (Graff *et al.*, 2015, Behrenfeld *et al.*, 2005, Martinez-Vicente *et al.*, 2013). En utilisant SOCA-BBP, il est

donc possible d'obtenir les distributions **verticales** du **POC** et du **carbone phytoplanctonique** à l'échelle globale avec une haute résolution spatio-temporelle. Cette information est essentielle pour la **quantification et la caractérisation des flux de carbone** à l'échelle globale (production primaire nette et flux d'export). De plus, le potentiel de la méthode à l'échelle climatologique a été mis en évidence dans l'article ci-dessus. En particulier, ces climatologies pourraient être très utiles pour l'initialisation et la validation des modèles biogéochimiques mais également pour servir de référence afin de déceler d'éventuelles anomalies temporelles ou spatiales.

Grâce à SOCA-BBP utilisé conjointement avec des méthodes qui estiment la distribution verticale de la **concentration en chlorophylle *a*** à partir des données couleur de l'eau (e.g. Uitz *et al.*, 2006), la variabilité du rapport phytoplanctonique carbone/chlorophylle pourrait être étudié sur la dimension **verticale**, ce qui représenterait un premier pas vers une meilleure compréhension du contrôle des conditions environnementales sur la biomasse phytoplanctonique.

Finalement, cette méthode a montré qu'il était possible de relier la distribution verticale d'une propriété bio-optique donnée (i.e. b_{bp}) à sa valeur de sub-surface estimée par satellite fusionnée avec les propriétés physiques de la colonne d'eau. Le développement de méthodes **analogues** pour estimer la distribution verticale d'autres propriétés bio-optiques à la fois mesurées par **satellite** et par **flotteurs Bio-Argo** (e.g. concentration en chlorophylle *a*) est une extension naturelle de l'étude présentée ici. Ceci fait l'objet de la Section suivante.

3.3 Paramétrisation de la distribution verticale de la concentration en chlorophylle *a*

Cette section présente un article en préparation pour Geophysical Research Letters :
A neural network-based method for merging ocean color and Argo data to extend surface bio-optical properties to depth : chlorophyll a concentration retrieval.

3.3.1 Introduction

The assessment of the distribution and space-time variability of phytoplankton across the global ocean is essential for biogeochemical studies. The chlorophyll-*a* concentration (Chl) is a universal

proxy of phytoplankton biomass. Remote sensing of Ocean Color Radiometry provides estimates of Chl for the surface layer of the global ocean quasi-synoptically (Gordon et McCluney, 1975). This technique led to a better understanding of phytoplankton dynamics in the upper ocean at various space and time scales (e.g. McClain, 2009, Siegel *et al.*, 2013). However, remote sensing of ocean color measures only one fifth of the so-called euphotic layer where most of the photosynthetic activity takes place (e.g. Morel et Berthon, 1989). Thus several studies have examined the potential of a global relationship linking the vertical distribution of phytoplankton biomass to ocean color-based estimates of surface Chl (Morel et Berthon, 1989, Uitz *et al.*, 2006). Nevertheless, these methods typically rely on the analysis of databases including a limited number of discrete measurements of Chl determined by fluorometry (Morel et Berthon, 1989) or High Performance Liquid Chromatography (HPLC, Uitz *et al.*, 2006).

The fluorescence of chlorophyll-*a* (hereafter fluorescence) is a widely used bio-optical proxy of Chl (Lorenzen, 1966). Since the launch of the Bio-Argo program, bio-optical profiling floats measure vertical profiles of fluorescence with high spatio-temporal resolution. Extensive databases of fluorescence profiles representative of the global ocean are now available. Although several studies have exploited Bio-Argo measurements to examine specific processes associated to limited space and time scales (e.g. Boss *et al.*, 2008, Boss et Behrenfeld, 2010, Mignot *et al.*, 2014, Xing *et al.*, 2014), to our knowledge, such information has not been used in global-scale studies.

The main limitation of fluorescence measurements is the variability of the fluorescence-to-Chl ratio, which depends on the physiological properties of phytoplankton as well as on community composition (Kiefer, 1973, Falkowski *et al.*, 1985, Cunningham, 1996). Thus fluorescence measurements need to be calibrated in Chl equivalent. A new fluorescence calibration method has been developed to transform into Chl equivalent and combine a large number of fluorescence profiles collected in the global open ocean (Sauzède *et al.*, 2015a). Using this method, all the available Bio-Argo fluorescence profiles could be converted into Chl, which would yield an extensive Chl-calibrated Bio-Argo database pertinent for global scale studies.

Using this global Bio-Argo database, and with a similar objective in mind as Morel et Berthon (1989) and Uitz *et al.* (2006), the present study examines the potential of extend bio-optical properties estimated from satellite to depth. The vertical distribution of phytoplankton biomass is driven by the

nutrient and light regime, itself strongly influenced by the physical state of the water column. Thus, ocean color surface data products merged with Argo hydrological profiles appear as good candidates to improve previous parameterizations of the vertical distribution of the phytoplankton biomass.

In this context artificial neural networks were chosen as the most appropriate technique because they are capable of modeling complex and nonlinear relationships (Hornik *et al.*, 1989). In addition a neural network-based method has proven to be useful to parameterize the vertical distribution of the particulate backscattering coefficient from merged ocean color and Argo data (Sauzède *et al.*, 2015b). Therefore, we expect that an analogous method can be developed to estimate the vertical distribution of Chl from satellite-to-Argo matchups.

In brief this study examines the potential of improving the previously published parameterizations of the vertical distribution of Chl (1) using a neural network-based method to link the Chl vertical distribution to the corresponding near-surface satellite-based value merged with vertically-resolved physical properties ; (2) using Argo-derived physical properties to better constrain the vertical distribution of Chl ; and (3) using the extensive global database of fluorescence measurements collected by Bio-Argo floats to train and validate the neural network-based method. The proposed method is referred as SOCA-CHL (for Satellite Ocean Color merged with Argo data to infer the vertical distribution of Chlorophyll-*a* concentration).

3.3.2 Databases and data processing

We assembled a database comprising 8330 concurrent vertical profiles of temperature, salinity and fluorescence collected by 83 Bio-Argo floats in the world's open oceans. Each profile of the Bio-Argo database was matched up with MODIS-Aqua derived data of Chl and PAR using the closest pixel from 8-day images with a 9-km resolution. The matchup procedure led to discarding 43% of the profiles from the initial Bio-Argo database. The geographic distribution of the 4725 remaining Bio-Argo profiles with concomitant MODIS-Aqua-derived products is shown in Figure 3.1. The Bio-Argo database used in this study covers most of the major oceanic basins.

We used the FLAVOR (Fluorescence to Algal communities Vertical distribution in the Oceanic Realm) method for calibrating the Bio-Argo fluorescence database. The FLAVOR method was trained and validated using a large database including 896 concomitant *in situ* vertical profiles of Chl determi-

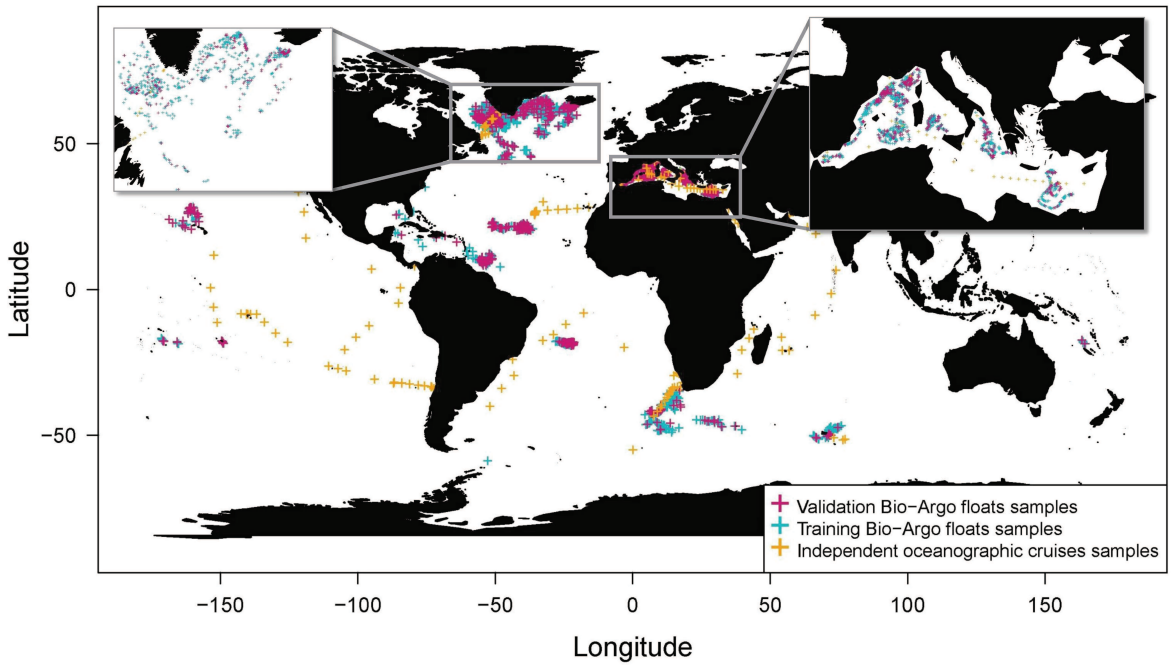


FIGURE 3.1 – *Geographic distribution of the 4725 stations sampled by the Bio-Argo floats and 262 stations sampled during 10 oceanographic cruises used in the present study. For each station, concurrent profiles of temperature, salinity and fluorescence were matched up with concomitant MODIS-Aqua-derived products. For each station of the oceanographic cruises, HPLC-determined pigment profiles are also available. Turquoise and purple crosses indicate the location of the profiles of the so-called training and validation datasets, respectively. The vertical profiles from oceanographic cruises used for an independent validation of the method are in orange.*

ned by HPLC and chlorophyll fluorescence. These profiles were collected as part of 22 oceanographic cruises, representative of the global ocean in terms of trophic and oceanographic conditions, making the method applicable to most oceanic waters. In summary, FLAVOR is a neural network-based method which uses as inputs (1) the shape of the *in situ* measured fluorescence profile (ten values from the normalized profile with values ranging between 0 and 1) ; (2) a normalization depth Z_{norm} which is the depth at which the fluorescence profile returns to a constant background value ; (3) the geolocation (latitude and longitude) and the day of acquisition of the fluorescence profile. The method yields as output the vertical distribution of the Chl. More specifically, FLAVOR is a 2-step method (see flowchart in Figure 4 of Sauzède *et al.* (2015a)). First, ten points of the Chl profile distributed within the water column are obtained from a neural network. Second the shape of the fluorescence profile is used to retrieve a quasi-continuous vertical profile of Chl using a scaling factor α computed from the ten points obtained in the first step. The global absolute error of FLAVOR Chl retrieval is 40% (Sauzède *et al.*, 2015a).

A limitation of FLAVOR is that it does not account for the process of Non-Photochemical Quenching (NPQ, e.g. Cullen et Lewis, 1995), responsible for a decrease in chlorophyll fluorescence values at high irradiance independently of the Chl. In the Bio-Argo database used in the present study, density profiles are available together with fluorescence profiles so the NPQ correction method of Xing *et al.* (2012a) can be applied. This correction involves substituting the fluorescence values acquired within the mixed layer by the maximum value within this layer. Hence we used a slightly modified version of the FLAVOR method : (1) the network was applied to each (uncorrected) Bio-Argo fluorescence profile to retrieve an α scaling factor ; (2) α was applied to the corresponding fluorescence profile corrected for NPQ to obtain a Chl-calibrated quasi-continuous vertical profile.

Each vertical Chl profile was scaled to a normalization depth, Z_{norm} ; i.e. the geometric depths of each profile was divided by Z_{norm}. This procedure enables to merge all the Chl profiles of the database regardless of their vertical shape and range of magnitude while simultaneously accounting for their variability (see above and Sauzède *et al.*, 2015a).

The characteristics of the resulting Bio-Argo database of concurrent profiles of temperature, salinity and Chl, matched up in space and time with satellite-derived products are presented in Figure 3.2. The Bio-Argo sampling appears evenly distributed among the twelve months of the year (Figure 3.2-a). Yet a lack of satellite matchups at high latitudes during winter and autumn causes a small bias between winter and the spring/summer season. Our database covers 8 years from 2008 to 2015 (Figure 3.2-b) with most of observations collected since 2013 (sharp increase in the frequency of Bio-Argo sampling).

The Bio-Argo database appears to be representative of most hydrological and biogeochemical conditions prevailing in the global open ocean (Figures 3.2-c and 3.2-d). The mixed layer depth, Z_m, computed from Bio-Argo density profiles using the criterion of $0.03\ kg.m^{-3}$ (de Boyer Montégut *et al.*, 2004), ranges between 15 and 900 m. In terms of trophic regimes, the surface Chl derived from MODIS-Aqua data, Chl_{MODIS}, covers 4 order of magnitude from the most oligotrophic to the most eutrophic waters (i.e. from 0.01 to 10 $mg.m^{-3}$). These characteristics make the method developed here applicable to most conditions encountered in open-ocean waters.

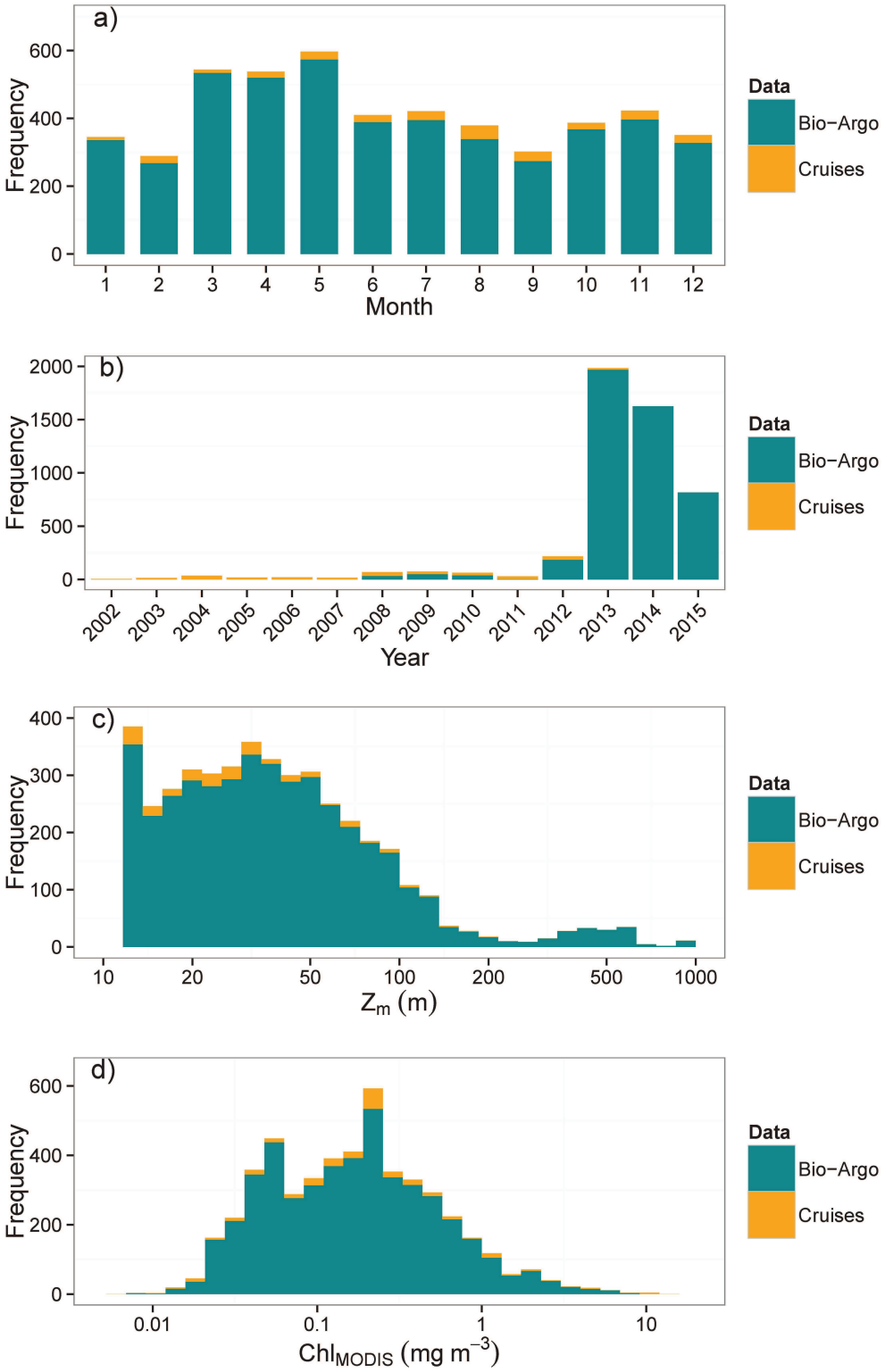


FIGURE 3.2 – *Characteristics of the Bio-Argo database (turquoise) and of the cruise database (orange). Distribution of the 4725 sampling stations on a monthly (a) and annual basis (b). Histogram of frequency of the mixed layer depth, Z_m (c) and of the satellite-derived surface chlorophyll-a concentration, Chl_{MODIS} (d).*

3.3.3 SOCA-CHL algorithm development

General principles of Multi-Layered Perceptron (MLP)

The neural network chosen in this study is a Multi-Layered Perceptron (MLP, Rumelhart *et al.*, 1988, Bishop, 1995). Several layers compose a MLP : the input layer, one or two hidden layers and the output layer. Each layer is composed of neurons that are elementary transfer functions which provide outputs when inputs are applied. The weights that interconnect all the neurons in the MLP are iteratively adjusted during the training phase of the MLP using the minimization of a cost function. To do that, a common backpropagation algorithm was used (Hornik *et al.*, 1989, Bishop, 1995). To prevent from overlearning, the training database (composed of 80% of the initial database) is randomly split in two subsets : the so-called learning and test datasets (Bishop, 1995). Finally, the validation dataset, composed of 20% of the initial database chosen randomly is used to evaluate independently the performance of the MLP.

Development of the MLP using the Bio-Argo and satellite matchup database

The following set of input components were selected as optimal for the MLP developed in this study : (1) a surface component defined as MODIS-Aqua-derived log-transformed data of Chl and PAR ; (2) a vertical component comprising the normalization depth Z_{norm}, the depth of the mixed layer depth Z_m, and four potential density values taken along the vertical Bio-Argo profile (three at shallow depths and one at depth ; note that these depths were chosen from a principal component analysis) ; and (3) a temporal component, i.e. the day of the year of the considered matchup. Z_{norm} is estimated from Z_m and Z_e, the depth of the euphotic zone (see Sauzede *et al.*, Submitted, for details of calculation). The MLP returns simultaneously as output ten $log(Chl)$ values corresponding to ten dimensionless depths taken at regular interval within the water column.

3.3.4 Results and Discussion

Retrieval of the vertical distribution of the chlorophyll *a* concentration

The optimal MLP (with minimum error of validation and minimum number of neurons) is composed of two hidden layers with 3 neurons on the first hidden layer and 2 neurons on the second one.

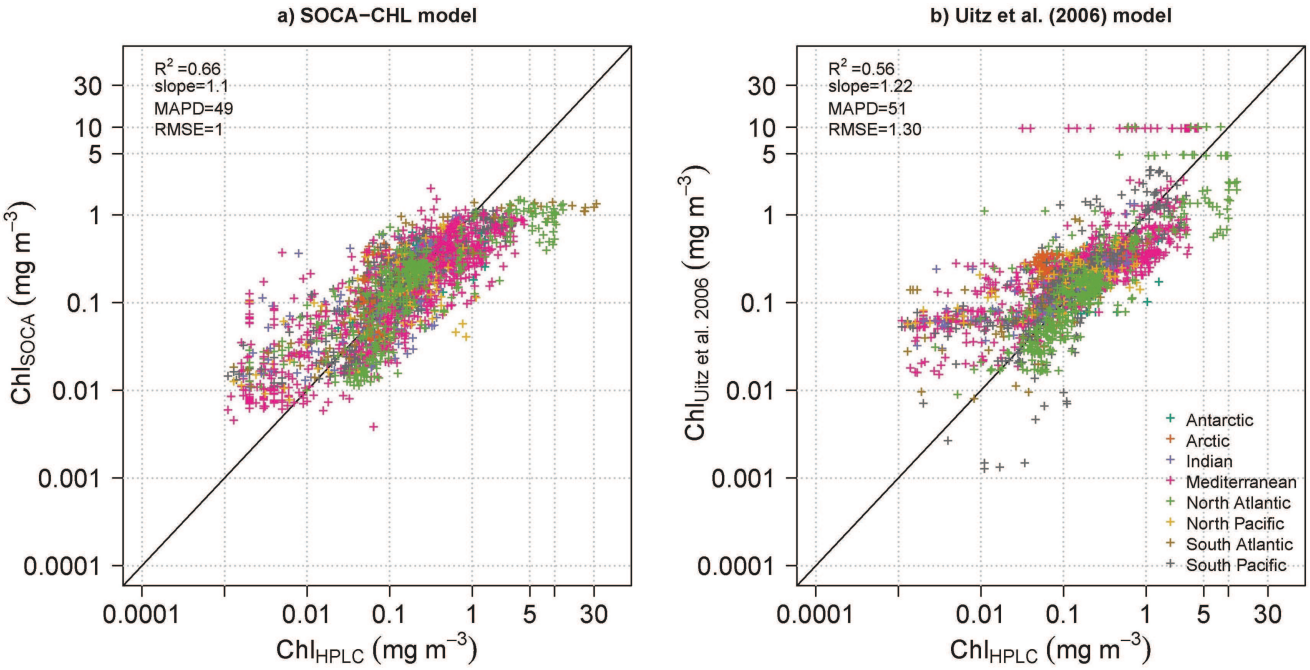


FIGURE 3.3 – *Comparison of the retrieval of the Chl vertical distribution using SOCA-CHL or Uitz* et al. *(2006) using as reference HPLC-determined Chl profiles collected during 262 oceanographic field cruises : (a) Chl values retrieved from SOCA-CHL (Chl_{SOCA}) compared to reference HPLC measurements (Chl_{HPLC}) and (b) Chl values retrieved from Uitz* et al. *(2006) ($Chl_{Uitzetal.2006}$) compared to the reference HPLC measurements (Chl_{HPLC}). Statistics are shown in each panel : R^2 and slope of the linear model between predicted and reference values, MAPD (Median Absolute Percent Difference) and RMSE (Root Mean Squared Error). The 1 :1 line is shown in black in each panel.*

First the skills of the SOCA-CHL method in predicting the Chl vertical profile were evaluated using the validation database (i.e. 20% of the initial Bio-Argo and satellite matchup database chosen randomly). For this purpose, we compared the 10 Chl values retrieved from SOCA-CHL with the corresponding Chl values from the (FLAVOR chlorophyll-calibrated) Bio-Argo database. The Chl retrieval is rather accurate with a determination coefficient (R^2) of the linear model between predicted and measured values of 0.67. The global Median Absolute Percent Difference (MAPD) is 44%. Nevertheless, we note that the Chl data used as a reference in the present validation exercise were derived from the FLAVOR calibration method, which introduces some level of uncertainties.

Thus, a second independent validation was performed using a database of concurrent *in situ* profiles of temperature, salinity and Chl determined by HPLC, matched up with satellite-derived products. These measurements were collected during 10 oceanographic cruises from 2002 to 2013 (see Figure 3.2-b) in various regions of the world's ocean (Figure 3.1). Similar to the Bio-Argo and satellite matchup database, the database of field measurements and satellite products can be considered

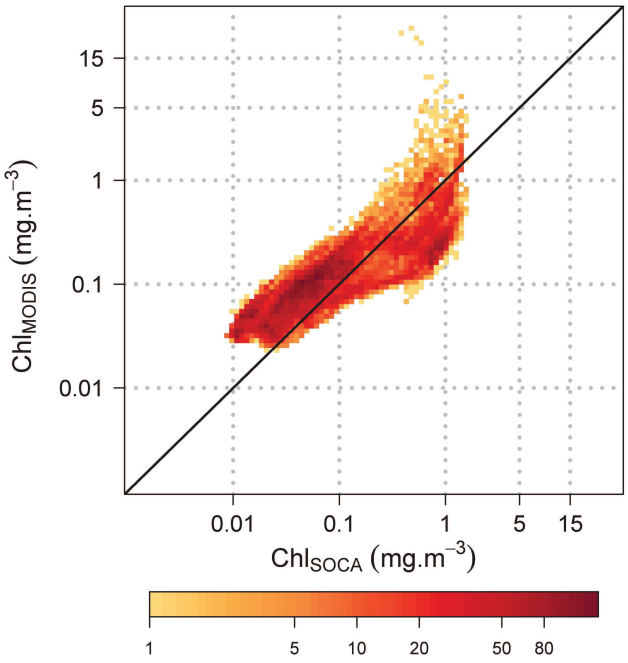


FIGURE 3.4 – *Comparison of the global climatological values of surface Chl retrieved by SOCA-CHL (Chl_{SOCA}) to those derived from MODIS-Aqua data (Chl_{MODIS}) for the month of June. The color code indicates the number of data points in each pixel. Both Chl_{SOCA} and Chl_{MODIS} are scaled to a resolution of 1 °.*

representative of the global ocean (Figure 3.2). The retrieval of the vertical distribution of Chl using SOCA-CHL is compared to HPLC-determined Chl profiles (Figure 3.3-a). The data appear fairly well distributed along the 1 :1 line. Yet SOCA-CHL seems to overestimate Chl at low values (< 0.01 $mg.m^{-3}$) and underestimate Chl at high values(> 5 $mg.m^{-3}$). The SOCA-CHL-retrieved Chl values are compared to those obtained from the parameterization of the vertical distribution of Chl of Uitz *et al.* (2006) (Figure 3.3-b). SOCA-CHL appears to improve the estimation of the Chl vertical distribution from ocean color data compared to the model of Uitz *et al.* (2006) (see statistics in the Figure 3.3-a and 3.3-b). We also note that our model has been developed using a database of fluorescence profiles calibrated in Chl with the FLAVOR method, which has a global average 40% error. In contrast, the model of Uitz *et al.* (2006) was based on accurate HPLC-determined Chl measurements. This suggests that with improved fluorescence calibration methods, the predictive skills of SOCA-CHL would improve as well and might be substantially better than those of the Uitz *et al.* (2006) method.

Example of SOCA-CHL application : production of a global Chl climatology

A natural application of the SOCA-CHL method is the development of Chl climatologies. As an

example, we developed a global climatology of the Chl for the month of June. We used as input data satellite-based climatological data of Chl and PAR for the month of June merged with temperature and salinity June climatological Argo data (Roemmich et Gilson, 2009). As the Mediterranean Sea is not represented in Argo climatologies, we used the WOD (World Ocean Database, Levitus *et al.*, 2013) climatological data for this basin. We show a direct comparison between the SOCA-CHL-retrieved Chl values averaged within the surface layer (defined as the layer comprised between the surface and the penetration depth $Z_{pd} = Z_e/4.6$, Morel et Berthon, 1989) and the satellite-based estimates (Figure 3.4). Overall, SOCA-CHL estimates of Chl for the surface layer are consistent with satellite-based Chl data. However, the results show that low Chl values are slightly underestimated by SOCA-CHL compared to MODIS-Aqua estimates.

Figure 3.5 presents the distribution of the Chl estimated from SOCA-CHL at three different depths of the global ocean. For the surface layer (Figure 3.5-a), high latitude environments of the northern hemisphere, the equatorial band, upwelling zones associated with Eastern Boundary Currents and other near-coastal areas are unsurprisingly characterized by high Chl values, whereas subtropical gyres are characterized by low values. SOCA-CHL seems to overestimate surface Chl values for high latitudes in the Southern hemisphere. Figures 3.5-b and 3.5-c suggest that the Chl decreases with depth in all areas of the world's oceans, except in the South Pacific and Atlantic Subtropical Gyres where Chl increases with depth. This increase is associated with the so-called Deep Chlorophyll Maximum (DCM), which is a typical feature of these oligotrophic regions (e.g. Cullen, 1982, Mignot *et al.*, 2014).

3.3.5 Conclusions and perspectives

The present study proposes a novel neural network-based method that predicts the vertical distribution of the Chl in the global open ocean using as input satellite-derived surface data of Chl and PAR and Argo density profiles. Our results show that previous parameterizations of the Chl vertical distribution based on satellite surface Chl could be improved using the large databases of bio-optical proxies of phytoplankton biomass (e.g. fluorescence) collected by Bio-Argo profiling floats. In the future, advances in the global Bio-Argo program and associated float deployments will bring additional measurements of concurrent density and fluorescence profiles in a broad range of oceanic conditions,

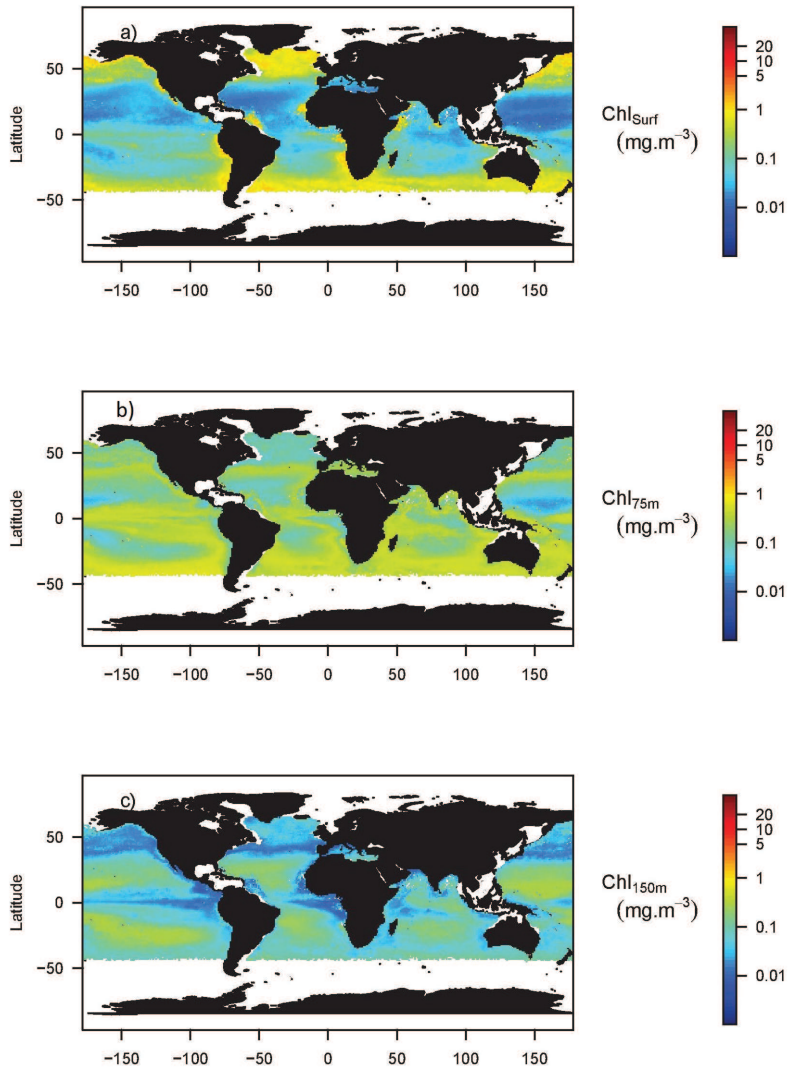


FIGURE 3.5 – *Depth-resolved climatology of chlorophyll-a concentration obtained from the SOCA-CHL method for the month of June with a 1° resolution. a) Surface Chl (i.e. averaged over the $0 - Z_{pd}$ layer), b) Chl at 75 m depth (i.e. averaged +/- 10 m) and c) Chl at 150 m (i.e. averaged +/- 10 m).*

which will lead to improve the relationship established by our MLP. The climatological application presented here indicates that in some areas the predictive skills of SOCA-CHL are less accurate (e.g. high latitudes in the Southern hemisphere for the month of June). This points to regions that are certainly under-sampled in the Bio-Argo database and thus deserve more attention. Hence our method could help to define new areas of the world's oceans where future Bio-Argo deployments would be critical. More generally, the present study highlights the need to develop an accurate calibration method specific to fluorescence profiles acquired from Bio-Argo profiling floats in order to improve their use in global-scale applications. Finally, the FLAVOR calibration method has the advantage that it retrieves from the shape of a fluorescence profile not only the Chl but also the composition of phy-

toplankton communities in terms of three size classes (i.e. micro-, nano- and pico-phytoplankton). One of the perspectives of this study is to use the Bio-Argo fluorescence database combined with the FLAVOR method to develop a new parameterization of the vertical distribution of the composition of phytoplankton communities. Such parameterization could represent a significant improvement compared to that of Uitz *et al.* (2006), based on a relatively reduced HPLC database.

3.4 Conclusion générale

Dans ce chapitre, nous avons développé des **méthodes neuronales** permettant d'estimer la **distribution verticale** de deux indicateurs bio-optiques de la biomasse phytoplanctonique (i.e. coefficient de rétrodiffusion particulaire, b_{bp}, et concentration en chlorophylle *a*, $[Chl]$) à partir de leurs mesures télédétectées par **couleur de l'eau** couplées à des profils verticaux **hydrologiques** acquis par flotteurs Argo. Les deux méthodes présentées dans ce chapitre, SOCA-CHL et SOCA-BBP respectivement pour la paramétrisation de la distribution verticale du b_{bp} et de la $[Chl]$, ont un grand potentiel. En effet, nous avons vu plusieurs fois au cours de cette thèse que la concentration en chlorophylle *a* était un estimateur largement reconnu de la **biomasse phytoplanctonique**. Il a été montré que le coefficient de rétrodiffusion était un proxy du **POC** et du **carbone phytoplanctonique** (e.g. Cetinić *et al.*, 2012a, Graff *et al.*, 2015). Ces deux méthodes SOCA permettent donc de dériver à partir de données couleur de l'eau les distributions verticales de propriétés biogéochimiques clés pour l'étude de la **biomasse phytoplanctonique** et de **l'export de carbone** à **l'échelle globale**.

La nécessité de calibrer les profils de fluorescence collectés par les flotteurs Bio-Argo en $[Chl]$ a rendu plus compliqué le développement de SOCA-CHL que celui de SOCA-BBP. En effet, l'avantage de l'utilisation du b_{bp} pour cette étude est qu'il n'a pas besoin d'être calibré a posteriori puisque les calibrations du constructeur sont considérées pour l'instant, comme valables a priori pour cette propriété bio-optique. Cela n'empêche que ces bases de données Bio-Argo de profils de b_{bp} puissent être un jour modifiées pour mieux prendre en compte des calibrations plus élaborées (e.g. meilleure prise en compte du « dark » instrumental). La base de données globale des profils de b_{bp} semble donc être a priori soumise à moins **d'incertitudes** que celle de la $[Chl]$ (que nous avons estimée à partir de FLAVOR, méthode ayant ses propres incertitudes). Le b_{bp} apparaît donc comme une mesure alternative intéressante à la fluorescence pour ce genre d'études.

Il serait intéressant de pouvoir estimer la distribution verticale de la **composition de l'assemblage phytoplanctonique** en plus du b_{bp} et de la $[Chl]$. En effet, cette information est essentielle dans les études liées aux flux de carbone. En utilisant l'estimation des distributions verticales de la composition des communautés phytoplanctoniques à partir de la forme d'un profil de fluorescence (méthode FLAVOR, voir Sect. 2.2.2 du chapitre 2), il serait envisageable de développer le même type d'approche que SOCA-CHL utilisant en entrées un matchup de données satellite/Argo pour estimer la distribution verticale de la composition des communautés phytoplanctoniques.

Enfin, l'utilisation **conjointe** des deux méthodes présentées ici offre la possibilité d'étudier la variabilité du rapport carbone/$[Chl]$ sur la dimension verticale pour l'océan global. Cette étude représenterait une étape importante vers une meilleure évaluation du contrôle de la lumière et des nutriments sur la biomasse phytoplanctonique et son état physiologique, ce qui permettrait d'améliorer la caractérisation de la distribution et de la variabilité de la production primaire et de l'export de carbone.

Dans ce contexte, le chapitre 4 présente une application **régionale** de ces méthodes utilisées de manière **conjointe** afin de tester leur cohérence et d'illustrer leur potentiel.

CHAPITRE 4

Étude régionale en Atlantique Nord : variabilité saisonnière de la biomasse phytoplanctonique et de la production primaire

Sommaire

4.1 Introduction 123

4.2 Données et méthodes 126

4.2.1 Données utilisées 126

4.2.2 Bio-régionalisation de l'Atlantique Nord 126

4.2.3 Distribution verticale de plusieurs paramètres biogéochimiques à partir des méthodes SOCA 127

4.2.4 Modèle de production primaire 128

4.2.5 Réalisation de séries temporelles climatologiques pour plusieurs paramètres biogéochimiques 130

4.3 Résultats et discussion 132

4.3.1 Bio-régionalisation de l'Atlantique Nord en 5 zones écologiques 132

4.3.2 Validation des modèles utilisés 133

4.3.3 Variabilité saisonnière du coefficient de rétrodiffusion 136

4.3.4 Variabilité saisonnière de la concentration en chlorophylle *a* 138

4.3.5 Comparaison des cycles saisonniers du coefficient de rétrodiffusion et de la concentration en chlorophylle *a* 141

4.3.6 Variabilité saisonnière de la composition des communautés phytoplanctoniques 142

4.3.7 Variabilité saisonnière de la production primaire jusqu'au taux de croissance phytoplanctonique 143

4.4 Conclusion générale et perspectives 147

4.1 Introduction

Ce chapitre propose une **application régionale** des méthodes SOCA [1] développées au cours de cette thèse. Nous avons choisi d'appliquer ces méthodes à **l'Atlantique Nord** qui est défini ici comme la zone allant de 40° à 70° Nord et de 0° à 80° Est. Malgré sa relative faible surface (~ 2% de l'océan global), l'Atlantique Nord est une zone **clé** dans la **séquestration du carbone** d'origine anthropique dans l'océan (~20% de la séquestration globale, Takahashi *et al.*, 2009, Sanders *et al.*, 2014). En effet, au Nord de l'Atlantique, les eaux chaudes associées au Gulf Stream sont progressivement refroidies et entrainent avec elles, lors de leurs plongées, une quantité significative de carbone (i.e. **pompe physique de carbone**, voir la circulation thermo-haline sur la Figure 4.1, e.g. Schmitz et McCartney, 1993). Quantifier et déterminer les incertitudes de ces flux de carbone dans cette zone et étudier leurs variabilités interannuelles et saisonnières est donc indispensable (e.g. Gruber, 2002).

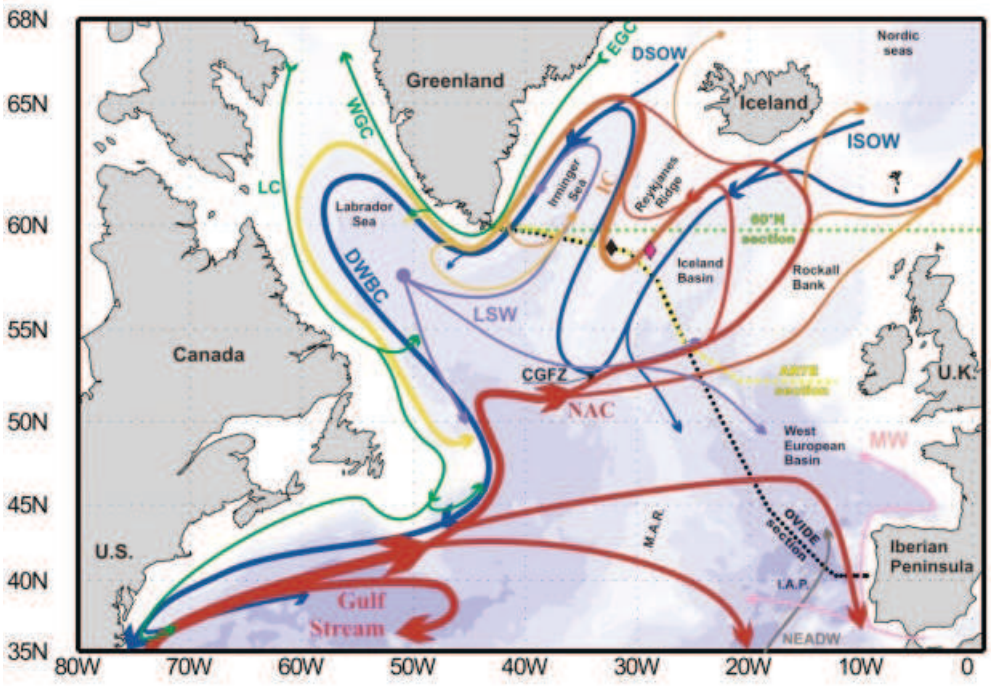


FIGURE 4.1 – *Schéma de circulation des courants en Atlantique Nord. Source : http ://www.insu.cnrs.fr/, Ifremer, Pascale Lherminier*

En parallèle, l'Atlantique Nord est une région très productive. Le cycle saisonnier de la biomasse

1. SOCA-BBP présentée dans la Sect. 3.2 et SOCA-CHL présentée dans la Sect. 3.3 permettent d'estimer la distribution verticale 1) du coefficient de rétrodiffusion et 2) de la concentration en chlorophylle *a* en utilisant comme seules données d'entrée des données biogéochimiques de surface estimées par satellite couplées à un profil vertical hydrologique acquis par un flotteur profileur Argo.

phytoplanctonique y est marqué et caractérisé par une efflorescence printanière (**bloom**, i.e. Longhurst, 2010). En plus de la pompe physique de carbone due à la circulation thermo-haline, le bloom favorise également l'exportation du carbone vers les couches profondes de l'océan (**pompe biologique de carbone**). Dans le futur, le changement climatique pourraient affecter les conditions environnementales qui contrôlent la **dynamique du bloom** printanier (Edwards et Richardson, 2004). C'est pour ces raisons que beaucoup d'études récentes se sont axées sur l'étude et la compréhension de la dynamique du bloom, et plus particulièrement sur son initiation. En effet **l'initiation du bloom** (e.g. date et intensité) a un impact sur toute la dynamique saisonnière de la biomasse phytoplanctonique « d'après bloom » (Henson *et al.*, 2006). Malgré l'importance de comprendre et décrire les processus associés à l'initiation du bloom, les théories restent très **controversées** pour le moment.

La première **théorie** expliquant le démarrage du bloom printanier a été proposée par **Sverdrup (1953)**. Ce modèle décrit l'initiation du bloom (pendant le **printemps**) par la **remontée de la couche de mélange** jusqu'à une **profondeur critique**, définie comme étant la profondeur à partir de laquelle la lumière est suffisante à la croissance nette de la biomasse phytoplanctonique. Depuis la présentation de ce modèle, de nombreuses études se sont focalisées sur les théories d'initiation du bloom. Certaines études ont nuancée ou redéfinie la théorie de Sverdrup (1953) (e.g. Taylor et Ferrari, 2011), d'autres l'ont confirmée (e.g. Henson *et al.*, 2006, Chiswell, 2011) et enfin certaines l'ont remise en cause (e.g. Behrenfeld, 2010).

La majorité de ces études sont basées sur des observations **satellitaires**. Afin de mieux comprendre et évaluer la **dynamique saisonnière de la biomasse phytoplanctonique** dans la zone de l'Atlantique Nord, il est important de prendre en compte la dimension **verticale** des observations (e.g. Boss et Behrenfeld, 2010) en complément de celles issues des satellites.

Nous proposons ici une application **climatologique** des modèles SOCA présentés dans le chapitre 3 afin d'étudier les variations saisonnières de la **distribution verticale** de la biomasse phytoplanctonique et de la production primaire en Atlantique Nord. Nos méthodes permettent en effet d'aborder cette approche climatologique à partir de laquelle les mécanismes **de premier ordre** peuvent être mis en évidence de manière relativement **robuste**. Par ailleurs et sur la base de ces études climatologiques, la variabilité interannuelle de ces mécanismes pourrait être étudiée en analysant les déviations par rapport aux tendances de premier ordre.

Afin de mener cette étude climatologique, comme l'Atlantique Nord est une zone très complexe et hétérogène spatialement (voir Figure 4.1), il est important de définir des **zones adaptées** afin d'y **aggréger** les données et d'y étudier leurs cycles saisonniers. Ces zones doivent être **écologiquement comparables** en termes de variations saisonnières phytoplanctoniques (i.e. bio-régions). Longhurst (2010) a proposé une classification de l'océan global en regroupant les régions présentant des similitudes écologiques. Cette **bio-régionalisation** est notamment basée sur la caractérisation des **cycles biogéochimiques de phytoplancton** dérivés d'observations satellitaires couplées à des données *in situ* (e.g. évolution de la couche de mélange, grazing). Ainsi, d'après les provinces de Longhurst (2010), L'Atlantique Nord peut être divisé en 7 grandes provinces écologiques : 1) BPLR (Boreal Polar Province), 2) ARCT (Atlantic Arctic Province), 3) SARC (Atlantic Subarctic Province), 4) NWCS (Northwest Atlantic Shelves Province), 5) NADR (North Atlantic Drift Province), 6) GEST (Gulf Stream Province) et 7) NAST-E (North Atlantic Subtropical Gyre Provine Est). Ces zones sont représentées sur la Figure 4.2.

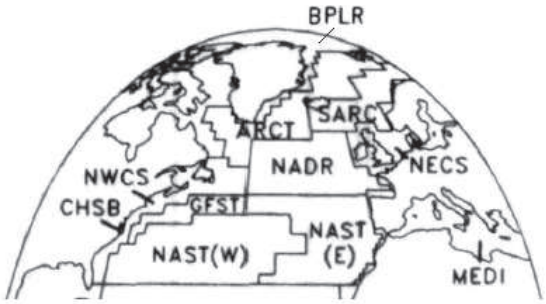


FIGURE 4.2 – *Distribution géographique des 7 provinces de l' Atlantique Nord définies par Longhurst (2010) : BPLR, ARCT, SARC, NWCS, NADR, GEST, NAST-E. Source : Longhurst (2010)*

Même si ces provinces peuvent constituer une forme de **référence** pour un certain nombre d'études, elles n'en ont pas moins été définies que de manière relativement arbitraire. Leurs contours rectilignes ne sont probablement pas indicatifs de frontières « réelles ». De **nouvelles approches** méthodologiques permettent d'envisager des bio-régionalisations avec des frontières plus fines et mieux définies dans l'espace (e.g. D'Ortenzio et Ribera d'Alcalà, 2008). Similairement à la bio-régionalisation de D'Ortenzio et Ribera d'Alcalà (2008), nous définirons dans cette étude des **régions écologiques** comparables afin d'y étudier les **cycles saisonniers climatologiques de la distribution verticale** de deux indicateurs bio-optiques de la biomasse phytoplanctonique (i.e. la concentration en chlorophylle *a* et le coefficient de rétrodiffusion particulaire).

Notons ici que, même si l'application présentée ici n'est qu'une étude préliminaire, notre approche vise à souligner le **potentiel** à l'échelle régionale de l'utilisation conjointe des deux méthodes SOCA développées pendant cette thèse. De plus, l'Atlantique Nord est une zone particulièrement adaptée à **l'application** des méthodes de ce type car un nombre très significatif de flotteurs Bio-Argo y ont été déployés (e.g. Figure 1 de l'article présenté dans la Sect. 3.2.2 du chapitre 3). Grâce à la **densité** des bases de données utilisées pour développer nos méthodes, nous émettons l'hypothèse que les réseaux de neurones sont bien entrainés pour cette zone et donc que les données modélisées qui en résultent sont réalistes.

4.2 Données et méthodes

4.2.1 Données utilisées

Nous rappelons que les méthodes **SOCA** (présentées en détails dans le chapitre 3) utilisent pour estimer la **distribution verticale des propriétés bio-optiques** (i.e. coefficient de rétrodiffusion particulaire ou concentration en chlorophylle *a*) comme principales entrées : 1) des **produits satellitaires**, 2) la distribution verticale des **propriétés hydrologiques** de la colonne d'eau mesurées par un flotteur profileur Argo et 3) le jour de l'année des données concomitantes et géo-localisées (notées **matchup** par la suite) considérées.

Dans cette application, afin d'étudier la variabilité saisonnière de la biomasse phytoplanctonique à **l'échelle régionale** (Atlantique Nord), nous avons utilisé des produits **climatologiques mensuelles**. Les entrées des méthodes SOCA sont donc : 1) les **climatologies satellitaires « couleur de l'eau »** mensuelles provenant de MODIS Aqua moyennée sur 9km (i.e. concentration en chlorophylle *a*, coefficient de rétrodiffusion particulaire mesuré à 700 nm et PAR, voir détails dans le chapitre 2) et 2) les **propriétés hydrologiques** de la colonne d'eau climatologiques mensuelles moyennées tous les 1° et dérivées de données collectées par **flotteurs Argo** (Roemmich et Gilson, 2009). La climatologie T/S (température et salinité) de Roemmich et Gilson (2009) a été établie en utilisant des données acquises de 2004 à 2008. Les différentes données satellitaires climatologiques ont été obtenues à partir de tout le composite MODIS Aqua (2002-2015). Le coefficient de rétrodiffusion particulaire mesuré à 700nm ($b_{bp}(700nm)$ noté par la suite b_{bp}), climatologique a été calculé de la même manière que décrit

précédemment, c'est-à-dire en utilisant les réflectances climatologiques à plusieurs longueurs d'ondes avec le modèle « Quasi-Analytical Algorithm » (QAA Lee *et al.*, 2002, 2009). Les détails de calculs du b_{bp} satellitaire sont présentés dans l'article présenté en Sect.3.2.2 du chapitre 3.

4.2.2 Bio-régionalisation de l'Atlantique Nord

Similairement à la bio-régionalisation de D'Ortenzio et Ribera d'Alcalà (2008), nous avons tenté de définir des **régions écologiques** définies par le même cycle saisonnier de concentration en chlorophylle *a* ($[Chl]$) de surface estimée par satellite pour l'Atlantique Nord. Pour chaque matchup satellite/Argo disponible, nous avons dérivé un cycle saisonnier climatologique de biomasse phytoplanctonique à partir de données de $[Chl]$ estimées par MODIS Aqua. En utilisant une méthode de **clusterisation** (i.e. k-means), nous avons ensuite réalisé une bio-régionalisation de l'Atlantique Nord. Cette clusterisation a été appliquée aux cycles saisonniers normalisés de $[Chl]$, c'est-à-dire que les valeurs de $[Chl]$ de chaque série temporelle ont été divisées par la valeur maximum au sein de cette série. Ceci permet de contraindre la clusterisation au **cycle saisonnier de la biomasse phytoplanctonique** et non à ses valeurs quantitatives.

4.2.3 Distribution verticale de plusieurs paramètres biogéochimiques à partir des méthodes SOCA

Grâce aux climatologies satellitaires couleur de l'eau et Argo (voir Sect. 4.2.1), en utilisant SOCA, il est possible de déterminer pour chaque matchup (satellite/Argo) la distribution verticale du **coefficient de rétrodiffusion** (b_{bp}) et de la **concentration en chlorophylle *a*** ($[Chl]$). Les matchups correspondent à des pixels de 1˚x 1˚.

Il serait d'une grande utilité de pouvoir estimer, en plus de la distribution verticale de la $[Chl]$ associée à la biomasse phytoplanctonique totale, la fraction de $[Chl]$ associée à chaque **classe de taille** phytoplanctonique (i.e. micro-, nano-, picophytoplancton). En effet, comme nous l'avons plusieurs fois rappelé dans cette thèse, la nature des communautés phytoplanctoniques est une variable indispensable à prendre en compte dans les études concernant les **flux de carbone** (e.g. Le Quere *et al.*, 2005).

A ce jour, nous n'avons pas eu le temps de développer une méthode neuronale qui permet de mo-

déliser la distribution verticale de la composition des communautés phytoplanctoniques à l'échelle globale (voir Sect. 3.4 du chapitre 2). Toutefois et dans le cadre de la présente étude spécifique à l'Atlantique Nord, nous avons décidé de développer un **réseau de neurones régional** pour l'Atlantique Nord qui permet de dériver les distributions verticales des fractions de $[Chl]$ associées aux trois classes de taille phytoplanctoniques en plus de la $[Chl]$ associée à la biomasse phytoplanctonique totale. Ce modèle régional sera nommé par la suite **SOCA-CHL-NA**.

Développement d'un réseau de neurones régional : SOCA-CHL-NA

Le réseau de neurones (MLP) régional, SOCA-CHL-NA, développé dans le but de modéliser la distribution verticale des communautés phytoplanctoniques est basé sur les mêmes entrées que celles définies pour la méthode SOCA-CHL présentée dans le chapitre 3 (i.e. les produits satellitaires de $[Chl]$ et de PAR, le profil T/S Argo et le jour de l'année du matchup considéré). SOCA-CHL-NA a pour spécificité ses sorties qui correspondent à la **forme du profil de** $[Chl]$ et non plus à des valeurs quantitatives de $[Chl]$. Ainsi, SOCA-CHL-NA restitue 10 valeurs de $[Chl]$ normalisée entre 0 et 1. Ensuite, en appliquant **FLAVOR** (voir détails Sect. 2.2.2 du chapitre 2) à chacune des formes de profil restituées par le réseau de neurones, nous pouvons estimer pour chaque matchup les distributions verticales de : 1) la $[Chl]$ associée à la **biomasse totale**, $[TChl]$, et 2) la $[Chl]$ associée à chaque **classe de taille** phytoplanctonique (i.e. $[microChl]$, $[nanoChl]$ et $[picoChl]$).

4.2.4 Modèle de production primaire

Nous utiliserons dans cette étude le **modèle de production primaire** développé par (Morel, 1991). Ce modèle estime la quantité de carbone organique fixé par photosynthèse, par unités de volume d'eau et de temps, en terme de taux de production primaire ($g.C.m^{-3}.s^{-1}$) ayant lieu localement à une profondeur z et un instant t pour une longueur d'onde λ :

$$P(z,t,\lambda) = 12 \cdot [Chl](z,t) \cdot a^*(z,t,\lambda) \cdot PAR(z,t,\lambda) \cdot \phi_c(z,t,\lambda) \tag{4.1}$$

où $[Chl]$ est la concentration en chlorophylle a en $mg.m^{-3}$, a^* est le coefficient d'absorption spécifique en $m^2(mg.[Chl])^{-1}$, PAR est la lumière pouvant être utilisée pour la photosynthèse en mol

quanta $m^{-2}.s^{-1}$ et ϕ_c est le rendement quantique de fixation de carbone dépendant de l'irradiance en mol$C$(mol quanta)$^{-1}$. Ce modèle prend en compte deux processus principaux de la **photosynthèse** : l'absorption de l'énergie lumineuse par les pigments photosynthétiques ainsi que sa transformation en produits synthétiques (Antoine *et al.*, 1996). Le facteur 12 permet la conversion des moles en grammes de carbone. L'intégration de cette équation 4.1 en fonction de λ, t et z nous renseigne sur la **production primaire quotidienne intégrée verticalement sur la couche productive** ($P.g.C.m^{-2}.d^{-1}$). La profondeur de cette couche productive est définie ici comme étant 1.5 fois la profondeur de la couche euphotique (Z_e). Rappelons que la couche euphotique correspond à la couche de l'océan où la photosynthèse peut être réalisée. Z_e est définie comme étant la profondeur à laquelle le PAR vaut 1% de sa valeur de surface. Cependant, comme le profil vertical de PAR mesuré *in situ* n'est pas disponible pour notre étude, Z_e a été calculée d'après la même procédure que celle détaillée dans la Sect. 3.2.2 du chapitre 3.

Le modèle de production primaire décrit précédemment a été utilisé au niveau des **groupes phytoplanctoniques** similairement à Uitz *et al.* (2010) afin de calculer la production primaire associée à chaque **classe de taille** phytoplanctonique (i.e. PP_{micro}, PP_{nano} et PP_{pico} respectivement pour le micro-, nano- et pico-phytoplancton). La production primaire total (PP_{tot}), attribuée à la totalité de la biomasse phytoplanctonique, est définie comme étant la **somme** des contributions de chaque classe. Finalement, le modèle nécessite les entrées suivantes : 1) la concentration en chlorophylle *a* de surface dérivée des produits satellitaires, 2) le PAR de surface estimé par satellite, 3) les profils verticaux de $[Chl]$ associée à chaque classe de taille phytoplanctonique ainsi que 4) leurs propriétés physiologiques. Les trois premières entrées sont accessibles à partir des données d'entrée et de sortie du modèle **SOCA-CHL-NA**. Les **propriétés physiologiques** nécessaires en entrée du modèle de production primaire sont estimées en utilisant les profils verticaux des propriétés physiologiques de chaque classe de taille phytoplanctonique proposés par Uitz *et al.* (2008) pour l'océan global. Ainsi, pour chaque profil de $[microChl]$, $[nanoChl]$ et $[picoChl]$ estimé par le modèle SOCA-CHL-NA, nous avons estimé la **production primaire intégrée dans la couche productive** associée à la biomasse phytoplanctonique totale (PP_{tot}) et associée aux trois classes de taille phytoplanctoniques (PP_{micro}, PP_{nano} et PP_{pico}).

4.2.5 Réalisation de séries temporelles climatologiques pour plusieurs paramètres biogéochimiques

En utilisant les climatologies mensuelles satellites et Argo, grâce à l'application des méthodes SOCA-BBP et SOCA-CHL-NA à la zone de l'Atlantique Nord, pour chaque matchup satellite/Argo (pixels de 1°x 1°), nous avons estimé les profils verticaux de b_{bp}, $[Chl]$, $[microChl]$, $[nanoChl]$ et $[picoChl]$. A partir de ces trois derniers profils, nous avons pu estimer pour chaque pixel la **production primaire** intégrée dans la couche $0 - 1.5.Z_e$ associée à la biomasse phytoplanctonique totale et aux communautés phytoplanctoniques : PP_{tot}, PP_{micro}, PP_{nano} et PP_{pico} (voir Sect. 4.2.4).

De plus, en utilisant la relation entre le carbone phytoplanctonique et le $b_{bp}(470nm)$ récemment publiée pour l'océan global (Graff *et al.*, 2015), un profil de **carbone phytoplanctonique** (C_{phyto} en $mg.C.m^{-3}$) a été dérivé du profil modélisé de b_{bp} pour chaque pixel de 1°x 1°. Les profils moyens climatologiques mensuels de $b_{bp}(700nm)$ ont d'abord été transformés en $b_{bp}(470nm)$ en utilisant la dépendance spectrale du b_{bp} (Loisel *et al.*, 2006, Kostadinov *et al.*, 2009) (voir Equation 1.11 dans la Sect. 1.2.3 du chapitre 1). Nous avons utilisé pour l'Atlantique Nord $\gamma = 1$ d'après l'étude de Loisel *et al.* (2006), γ étant la pente de la dépendance spectrale de b_{bp}. Enfin, la relation de Graff *et al.* (2015) a été appliqué aux profils de $b_{bp}(470nm)$ pour estimer les profils de C_{phyto}.

Pour chaque matchup, le **taux de croissance phytoplanctonique** (μ en d^{-1}) est estimé en divisant PP_{tot} ($mg.C.m^{-2}.d^{-1}$) par le carbone phytoplanctonique intégré dans la couche productive, ($C_{phyto1.5.Z_e}$ en $mg.C.m^{-2}$).

Le PAR (Photosynthetically Available Radiation) moyen dans la couche de mélange est estimé à partir du coefficient d'atténuation totale, K_{PAR}, (voir détails des calculs dans la Sect 4.2.4), ainsi que du PAR de sub-surface, $PAR(0-)$, calculé en utilisant une diminution de 7.6% du PAR estimé par satellite au-dessus de la surface (Morel, 1991). Le PAR moyenné dans la couche de mélange, PAR_{ML}, est calculé selon :

$$PAR_{ML} = \frac{1}{Z_m} \cdot \int_0^{Z_m} PAR(0-) \cdot exp(-K_{PAR} \cdot z) dz \tag{4.2}$$

avec Z_m la profondeur de la couche de mélange estimée à partir d'un profil de densité climatologique en utilisant le critère de densité $0.03kg.m^{-3}$ défini par de Boyer Montégut *et al.* (2004).

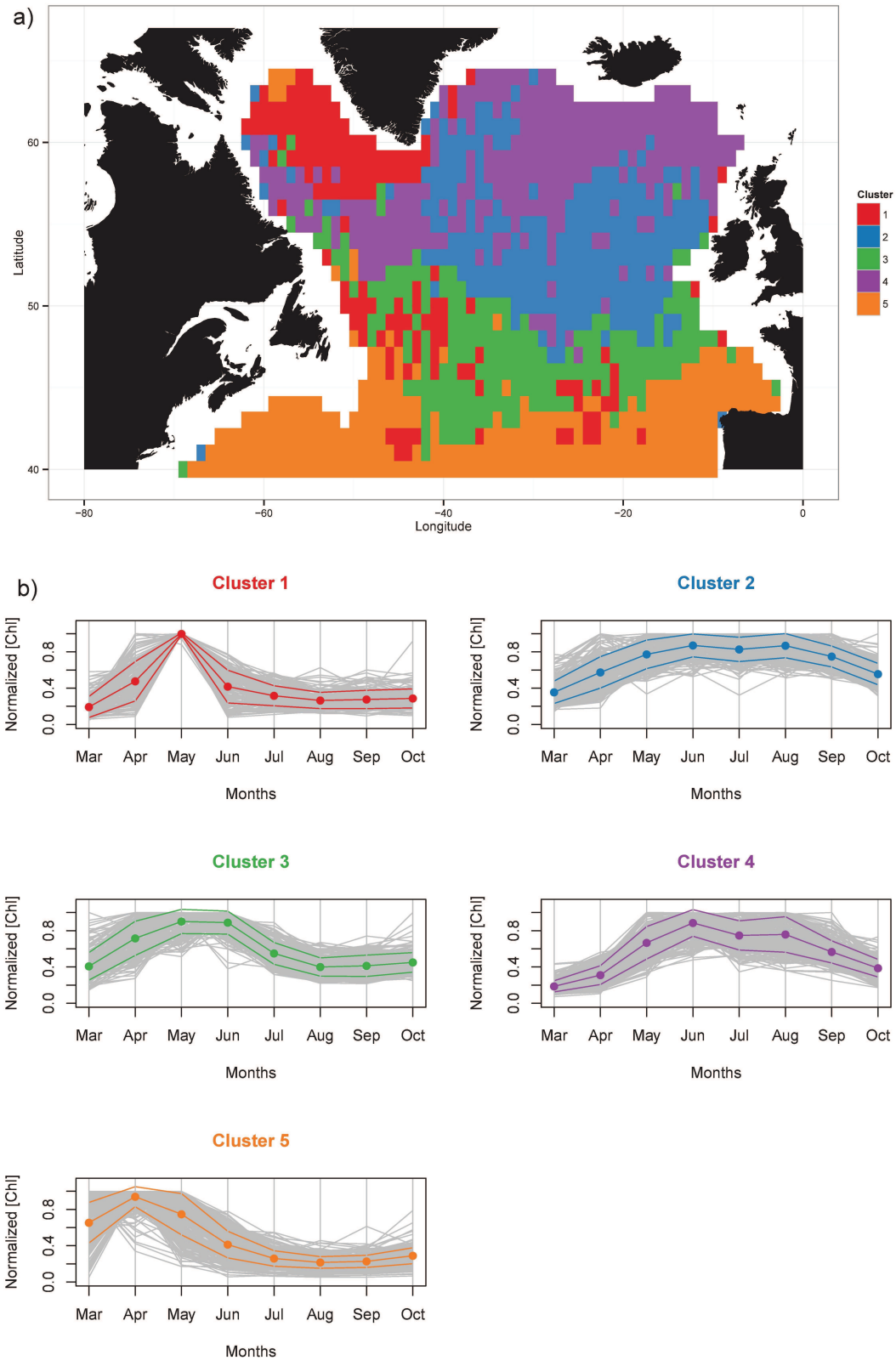


FIGURE 4.3 – *Biorégionalisation de l'Atlantique Nord : a) distribution géographique des 5 clusters obtenus à travers l'analyse des k-means et b) pour chaque cluster les points correspondent au cycle annuel de la* $[Chl]$ *normalisée en moyenne, les courbes de couleurs correspondent à* $\pm$ *l'écart-type de cette moyenne et enfin les courbes grises correspondent à tous les cycles saisonniers de* $[Chl]$ *normalisée utilisés pour calculer la moyenne et l'écart type.*

Finalement, ces différents indicateurs biogéochimiques ont été moyennés pour chaque **bio-région** déterminée par la méthode de clusterisation présentée dans la Sect. 4.2.2. Enfin, des **séries mensuelles climatologiques moyennes** permettant d'étudier la **dynamique saisonnière** des différents paramètres biogéochimiques modélisés ont été générées pour chaque bio-région de l'Atlantique Nord.

4.3 Résultats et discussion

4.3.1 Bio-régionalisation de l'Atlantique Nord en 5 zones écologiques

La bio-régionalisation de l'Atlantique Nord présentée sur la Figure 4.3 révèle la **variabilité saisonnière** du **cycle saisonnier** de la biomasse phytoplanctonique à la surface de l'océan en utilisant une **approche climatologique**. Chaque bio-région (ou cluster) est **spécifique** quant à la date, l'intensité et la durée du bloom observées à partir des données satellites de $[Chl]$.

Afin de mieux comprendre la dynamique temporelle de la $[Chl]$ pour chaque cluster, il est important de la mettre en relation avec la variabilité temporelle des propriétés **hydrologiques** de la colonne d'eau. Les Figures 4.4-a et 4.4-b présentent les cycles saisonniers moyens de Z_m et de Z_e.

Les clusters 5 et 3 sont typiques des **régions tempérées** et sont caractérisés par une augmentation de $[Chl]$ à partir respectivement d'avril et mai. La différence d'un mois observée dans le maximum de $[Chl]$ entre les deux clusters est due à la propagation vers le Nord du bloom printanier (Siegel, 2002b). Ces deux clusters sont caractérisés par de faibles couches de mélange ainsi que des couches euphotiques relativement profondes pendant l'été (entre 50 et 65 m, voir Figure 4.4). Cette dynamique correspond bien à la connaissance que l'on a des zones tempérées (voir détails dans la Sect. 1.1.1 et la Figure 1.2-b du chapitre 1). Les clusters 2 et 4 sont représentatifs des **régions subpolaires** (voir Figure 1.2-c du chapitre 1) avec une augmentation de biomasse phytoplanctonique visible pendant environ 3 mois (de juin à août). Ces bio-régions sont caractérisées par de forts mélanges hivernaux (~ 300 m en moyenne) et des couches euphotiques relativement peu profondes pendant l'été (~30 m, voir Figure 4.4). Finalement, le cluster 1, principalement représentatif du **Nord de la mer du Labrador**, est toujours caractérisé par une augmentation abrupte de $[Chl]$ en mai, suivie d'une rapide diminution. Bien que que ce cluster se trouve à la même latitude que les clusters des régions subpolaires, il est caractérisé par un cycle saisonnier de Z_m et Z_e plus proche de celui des régions tempérées (voir Figure

4.4). Ceci est dû au fait que cette zone est influencée par l'apport d'eau douce provenant de la fonte des glaces Arctiques (Lacour *et al.*, 2015, Devred *et al.*, 2007).

Ces 5 clusters sont très cohérents avec la bio-régionalisation établie récemment par (Lacour *et al.*, 2015) pour l'Atlantique Nord. Leur bio-régionalisation est basée sur des données satellitaires climatologiques de $[Chl]$ (provenant de GlobColour) ayant une résolution **plus fine spatialement et temporellement** que les données utilisées pour l'étude présentée ici. En effet, nous avons utilisé des données avec une résolution spatiale et temporelle moins fine afin d'obtenir une bio-régionalisation cohérente spatialement et temporellement avec les données provenant de nos modèles (résolution spatiale de 1° et résolution temporelle mensuelle). Le fait que la bio-régionalisation présentée ici soit **cohérente** avec celle de Lacour *et al.* (2015) suggère que la résolution spatio-temporelle utilisée pour notre analyse est suffisante.

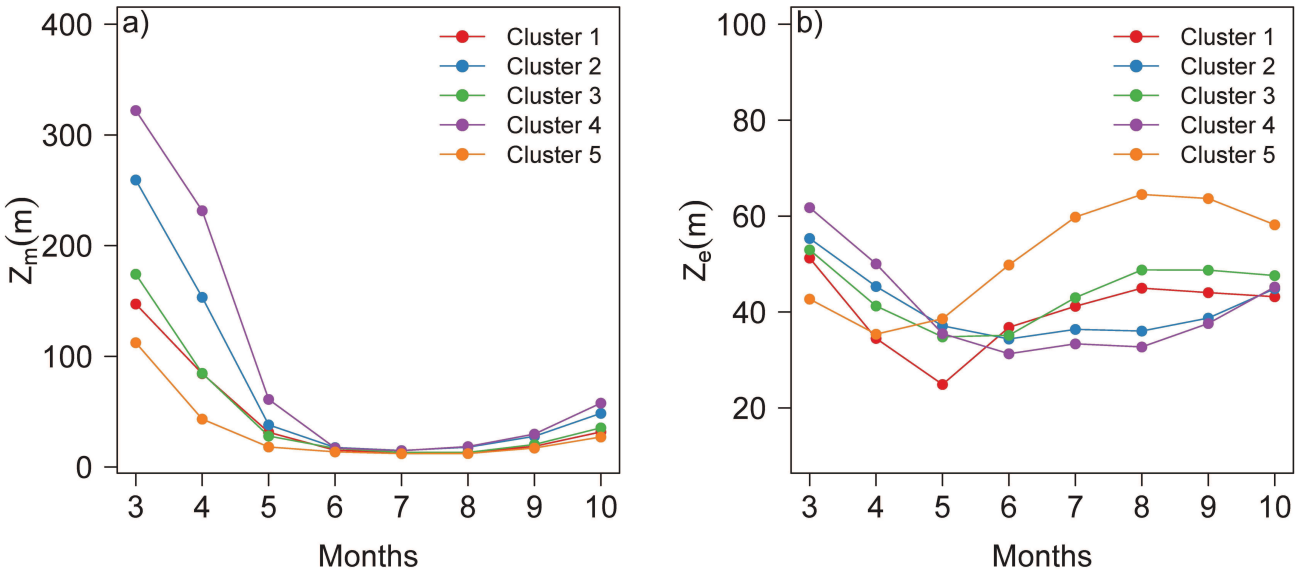


FIGURE 4.4 – *Cycles saisonniers moyen a) de la profondeur de la couche de mélange (Z_m en m) et b) de la profondeur de la couche euphotique (Z_e en m) pour 5 clusters de l'Atlantique Nord.*

4.3.2 Validation des modèles utilisés

Distribution verticale du coefficient de rétrodiffusion : évaluation du modèle SOCA-BBP

La validation du modèle SOCA-BBP à l'échelle globale a été présentée dans l'article de la Sect. 3.2.2 du chapitre 3. Une validation indépendante a également été proposée pour des flotteurs Bio-Argo déployés dans quatre zones océaniques majeures, dont le gyre subpolaire de **l'Atlantique Nord**

(voir les détails de cette validation sur les Figures 5-b, 7-e, 7-f, 8-a ainsi que la Table 3 dans l'article présenté en Sect. 3.2.2).

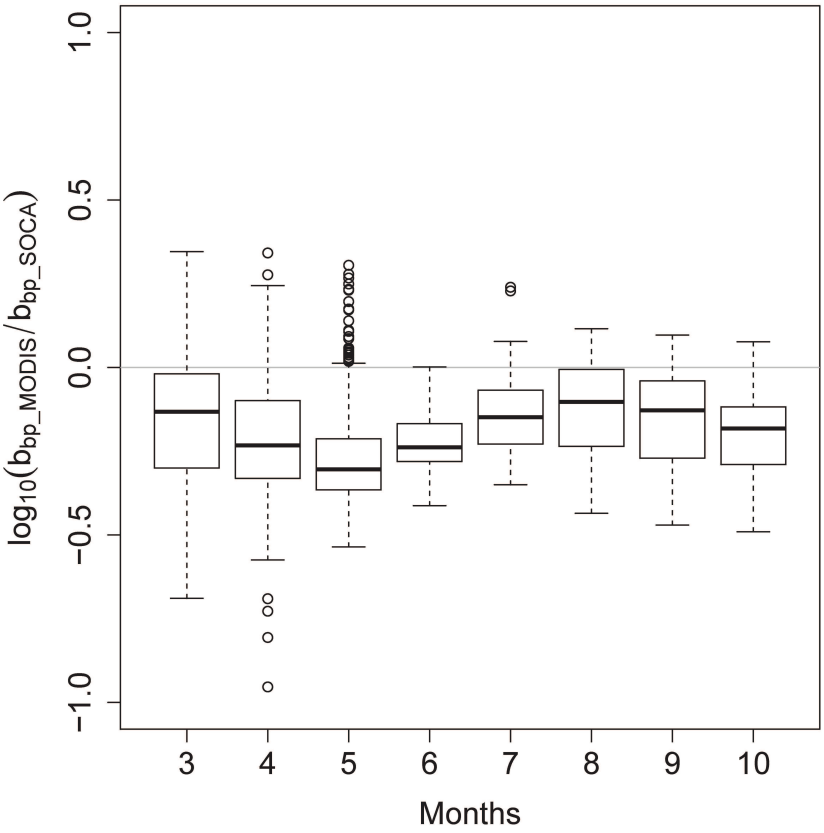


FIGURE 4.5 – *Boîte à moustaches (« boxplots ») mensuelles des rapports transformés en log_{10} entre le coefficient de rétrodiffusion estimé par satellite (b_{bp_MODIS}) et celui estimé par SOCA-BBP à la surface (b_{bp_SOCA}).*

En plus de ces validations présentées dans la Sect. 3.2.2, nous avons validé les résultats climatologiques provenant du modèle SOCA-BBP pour l'Atlantique Nord en surface. Nous avons comparé le b_{bp} estimé par SOCA-BBP en **surface** (b_{bp} moyen dans la couche de pénétration $0 - Z_{pd}$ avec $Z_{pd} = Z_e/4.6$) pour chaque mois avec le b_{bp} climatologique provenant de MODIS Aqua. D'après la Figure 4.5, le modèle SOCA-BBP semble surestimer le b_{bp} en surface du mois de mars à octobre. Cependant, l'erreur reste assez faible et le biais est relativement constant temporellement.

Distribution verticale de la biomasse phytoplanctonique totale et des classes de taille phytoplanctoniques : évaluation du modèle SOCA-CHL-NA

Le modèle SOCA-CHL-NA régional restitue la **forme du profil de** $[Chl]$ à partir d'un matchup entre les produits satellitaires de $[Chl]$ et de PAR et un profil hydrologique Argo. Cette forme de profil

est ensuite convertie en 1) $[Chl]$ associée à la **biomasse phytoplanctonique totale** et 2) $[TChl]$ associée à chaque **classe de taille phytoplanctonique** (i.e. $[microChl]$, $[nanoChl]$ et $[picoChl]$ respectivement pour le micro-, nano- et pico-phytoplancton) en utilisant la méthode de calibration **FLAVOR** présentée dans l'article en Sect. 2.2.2 du chapitre 2. Ce sont ces sorties finales (i.e. distributions verticales de $[Chl]$, $[microChl]$, $[nanoChl]$ et $[picoChl]$) qui sont validées pour l'Atlantique Nord en les comparant à des **données *in situ*** (analysées par HPLC) provenant de **campagnes océanographiques**.

TABLE 4.1 – *Abréviations, localisation, période, nombre de stations (avant et après matchup) et PI des deux campagnes océanographiques en Atlantique Nord utilisées dans cette étude*

Projet	Abréviation	Localisation	Période	Nombre de stations	Nombre de stations après matchup	PI
HUDSON 2013-008	LA	Mer du Labrador	Mai 2013	22	14	Claustre H. et Ras J.
GEOVIDE	GEO	Atlantique Nord tempéré et subpolaire et Mer du Labrador	Mai-Juin 2014	33	14	Claustre H. et Ras J.

Les données provenant des campagnes océanographiques pour la période 2002-2015 en Atlantique Nord sont (après matchup) : 1) 14 profils provenant de la campagne HUDSON et 2) 14 profils provenant de la campagne GEOVIDE (voir abréviations, localisations, périodes et PI de ces deux campagnes sur la Table 4.1 et la Figure 4.6). Ces **28 profils de pigments** analysés par HPLC ont été utilisés pour valider notre méthode en Atlantique Nord. SOCA-CHL-NA a été appliqué en utilisant comme entrées : 1) le jour d'acquisition du profil, 2) les données satellitaires correspondant aux matchups des 28 profils *in situ* HPLC avec 3) le profil T/S (CTD) concomitant au profil HPLC. Les profils de $[Chl]$, $[microChl]$, $[nanoChl]$ et $[picoChl]$ estimés par SOCA-CHL-NA ont ensuite été comparés aux profils *in situ* analysés par **HPLC**. Les résultats statistiques de cette comparaison sont présentés dans la Table 4.2. De plus, les résultats provenant du modèle SOCA-CHL-NA sont comparés aux résultats obtenus par la méthode de **Uitz *et al.* (2006)** qui permet d'estimer les distributions verticales de la composition des communautés phytoplanctoniques à partir des données $[Chl]$ couleur de l'eau (détails dans la Sect. 3.3.1 du chapitre3).

D'après ces résultats statistiques, bien que notre modèle semble modéliser la $[Chl]$ et la $[microChl]$

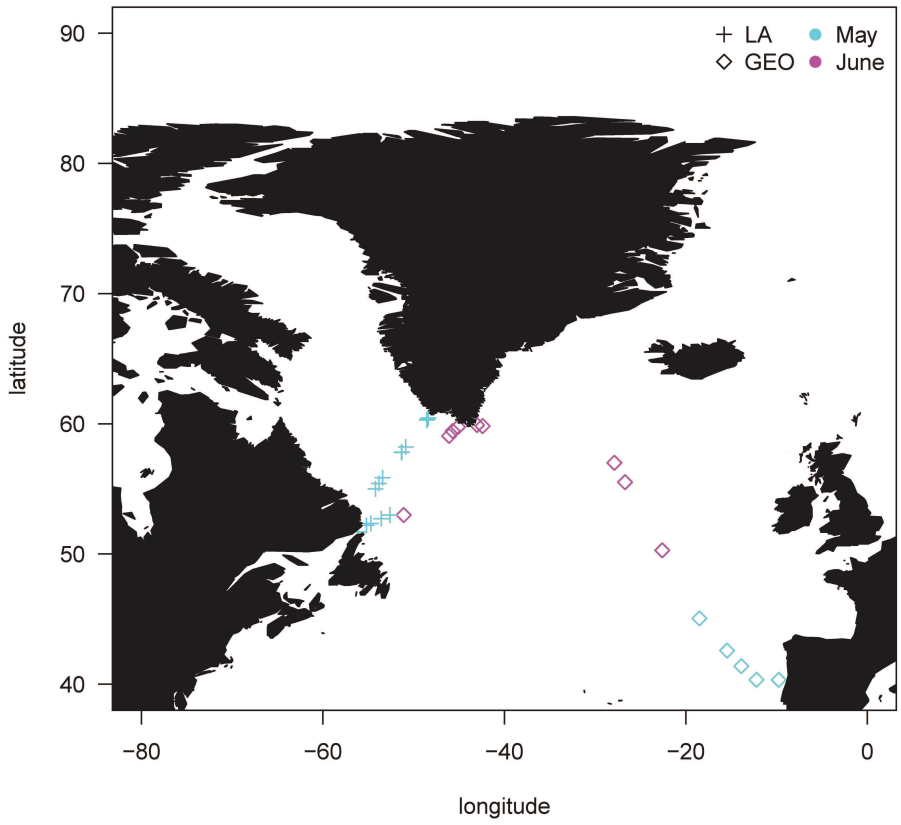


FIGURE 4.6 – *Distribution géographique des 28 profils HPLC utilisés pour valider le modèle SOCA-CHL-NA en Atlantique Nord. La couleur correspond au mois d'échantillonnage et la forme correspond à la campagne océanographique correspondant au profil.*

avec moins de précision que le modèle de Uitz *et al.* (2006) (i.e. MAPD de 77% et 82% pour SOCA-CHL-NA contre 68% et 73% pour le modèle de Uitz *et al.* (2006)), la RMSE calculée avec les résultats provenant de notre modèle est plus faible. Ceci est surement dû à une meilleure restitution des plus fortes valeurs par SOCA-CHL-NA. D'autre part, notre modèle semble mieux estimer la $[nanoChl]$ et la $[picoChl]$ en comparaison au modèle de Uitz *et al.* (2006) (voir valeurs statistiques dans la Table 4.2).

TABLE 4.2 – *Résultats statistiques de la comparaison des valeurs prédites par le modèle SOCA-CHL-NA avec les valeurs observées pendant les campagnes océanographiques LA et GEO. Même comparaison pour le modèle établi par Uitz* et al. *(2006)*

	SOCA-CHL-NA				**Uitz *et al.* (2006)**			
	R^2	a	MAPD(%)	RMSE	R^2	a	MAPD(%)	RMSE
$[TChl]$	0.5	0.5	77	2.58	0.54	0.5	68	2.93
$[microChl]$	0.38	0.59	82	2.64	0.48	0.46	73	3.01
$[nanoChl]$	0.46	0.58	63	0.14	0.27	0.4	149	0.31
$[picoChl]$	0.31	0.53	69	0.08	0.4	0.52	124	0.1

Ces résultats de validation du modèle régional SOCA-CHL-NA sont **satisfaisants**, cependant il est important de noter qu'à cause du **manque de données *in situ*** HPLC, notre modèle SOCA-CHL-NA n'a pu être validé qu'à partir de peu de profils qui ont été collectés pendant une **période très spécifique** (mois de mai et juin).

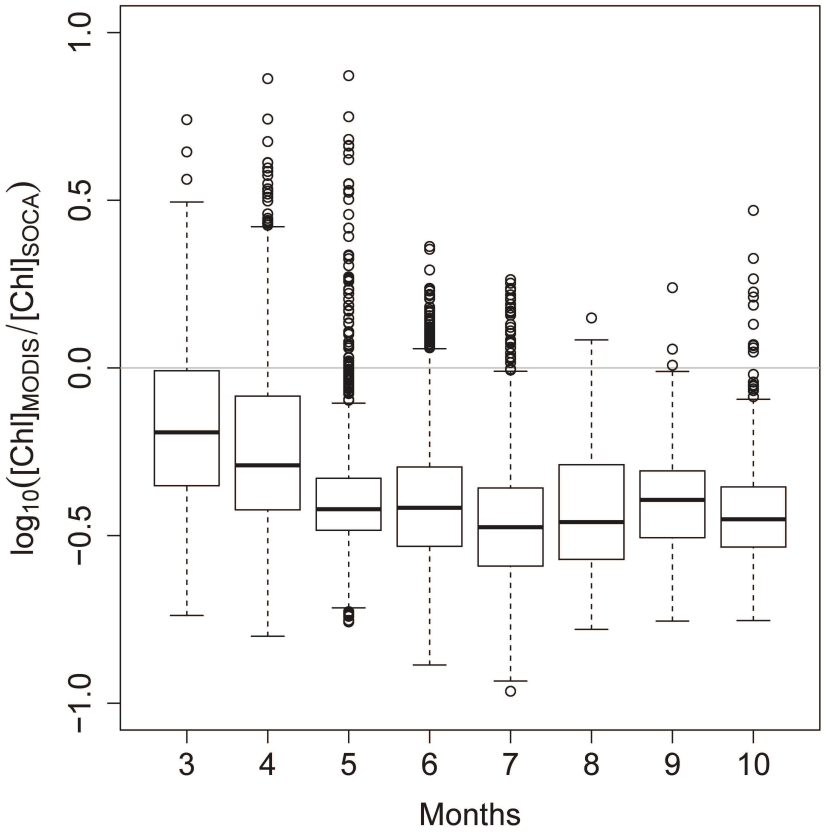


FIGURE 4.7 – *Boîte à moustaches (boxplots) des rapports mensuels transformés en log_{10} entre la concentration en chlorophylle a estimée par SOCA-CHL-NA à la surface ($[Chl]_{SOCA}$) et celle estimée par satellite ($[Chl]_{MODIS}$).*

Afin de valider notre méthode pour les mois de mars à octobre, la $[Chl]$ prédite par SOCA-CHL-NA en surface est comparée à la $[Chl]$ climatologique mensuelle estimée par satellite. Cette deuxième validation est présentée sur la Figure 4.7. La méthode surestime la $[Chl]$ de surface par rapport aux estimations de MODIS Aqua pour tous les mois (similairement à SOCA-BBP). Cependant, le biais entre les deux produits est relativement constant (surtout de mai à octobre). Alors que SOCA-CHL-NA est un **algorithme régional**, l'erreur de prédiction de la $[Chl]$ en surface par rapport aux données satellites est plus élevée que celle correspondante à la prédiction du b_{bp} en surface par SOCA-BBP. En effet, la **variabilité** de la distribution verticale du b_{bp} est surement mieux modélisée par SOCA que celle attribuée à la $[Chl]$ (voir discussion dans la Sect. 3.4 du chapitre 3).

4.3.3 Variabilité saisonnière du coefficient de rétrodiffusion

Les cycles saisonniers climatologiques moyens du b_{bp} estimés par SOCA-BBP (b_{bp_SOCA}) en surface et intégrés dans la couche productive $0 - 1.5.Z_e$ sont présentés respectivement sur la Figure 4.8-a et 4.8-b pour les 5 clusters caractéristiques de l'Atlantique Nord. Dans le but de comparer les cycles saisonniers de b_{bp_SOCA} de surface avec ceux du b_{bp} climatologique satellitaire, b_{bp_MODIS}, les cycles saisonniers de b_{bp_MODIS} moyens sont représentés pour chaque cluster sur la Figure 4.9-b. Les dynamiques temporelles de b_{bp} en surface estimées par BBP-SOCA et issues de MODIS Aqua sont assez **similaires** pour tous les clusters. Cependant, de légères différences peuvent être observées. Le cluster 1 est caractérisé par une augmentation de b_{bp_SOCA} plus prononcée en mai que celle visible à partir de b_{bp_MODIS}. Ensuite, la diminution de b_{bp_SOCA} à partir de mai est plus abrupte que celle de b_{bp_MODIS}. Le cluster 3 est caractérisé par un pic de b_{bp_MODIS} en juin alors que l'augmentation de b_{bp_SOCA} est déjà visible à partir de mai.

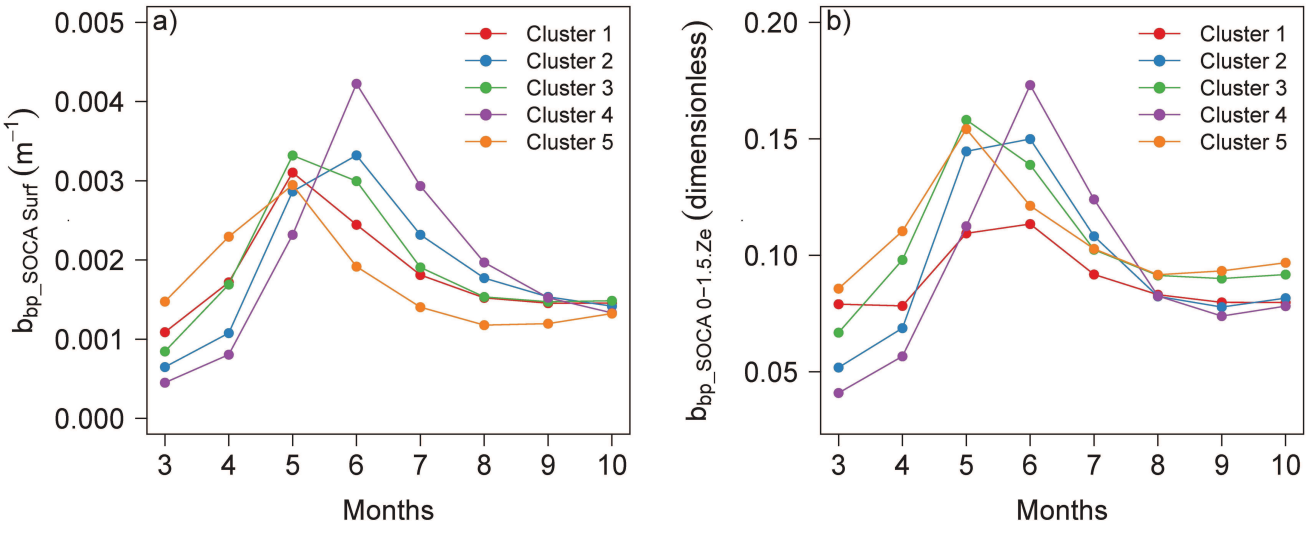


FIGURE 4.8 – *Cycles saisonniers climatologiques moyens du coefficient de rétrodiffusion estimé par SOCA-BBP (b_{bp_SOCA}) a) en surface et b) intégré sur la couche $0 - 1.5.Z_e$ pour les 5 clusters de l'Atlantique Nord.*

A la fois pour les cycles saisonniers moyens de b_{bp_SOCA} et b_{bp_MODIS}, les clusters 1,2,3 et 4 sont caractérisés par des valeurs plus faibles de b_{bp} en fin d'hiver et début de printemps (mars/avril) qu'en fin d'été et début d'automne (aout à octobre). Dans le cluster 5, des valeurs similaires de b_{bp} en fin d'hiver et en fin d'été.

En comparant les **valeurs quantitatives** de b_{bp} pour chaque cycle saisonnier issu de SOCA-BBP

et de MODIS Aqua, il apparaît une légère **surestimation** de b_{bp_SOCA} par rapport à b_{bp_MODIS}. Ceci correspond à la surestimation mise en évidence en présentant la validation du modèle SOCA-BBP en Atlantique Nord en comparant les prédictions en surface aux données satellitaires (voir Sect. 4.3.2 et la Figure 4.5).

La méthode SOCA-BBP permet d'obtenir la **distribution verticale** du b_{bp}. A partir de cette dimension verticale (présentée plus en détails ci-dessous), des cycles saisonniers moyens de b_{bp} **intégré dans la couche productive** $0 - 1.5.Z_e$ ont été calculés pour chaque cluster (voir Figure 4.8-b). L'évolution temporelle de ces cycles est très similaire à celle des cycles de b_{bp_SOCA} de surface pour tous les clusters mis à part le cluster 1. Ce dernier révèle une dynamique beaucoup moins prononcée avec le b_{bp} intégré en comparaison à celle du b_{bp} de surface. De plus, pour le cluster 1, le maximum de b_{bp} de surface est en mai tandis que lorsque le b_{bp} intégré est moyenné le maximum est en juin.

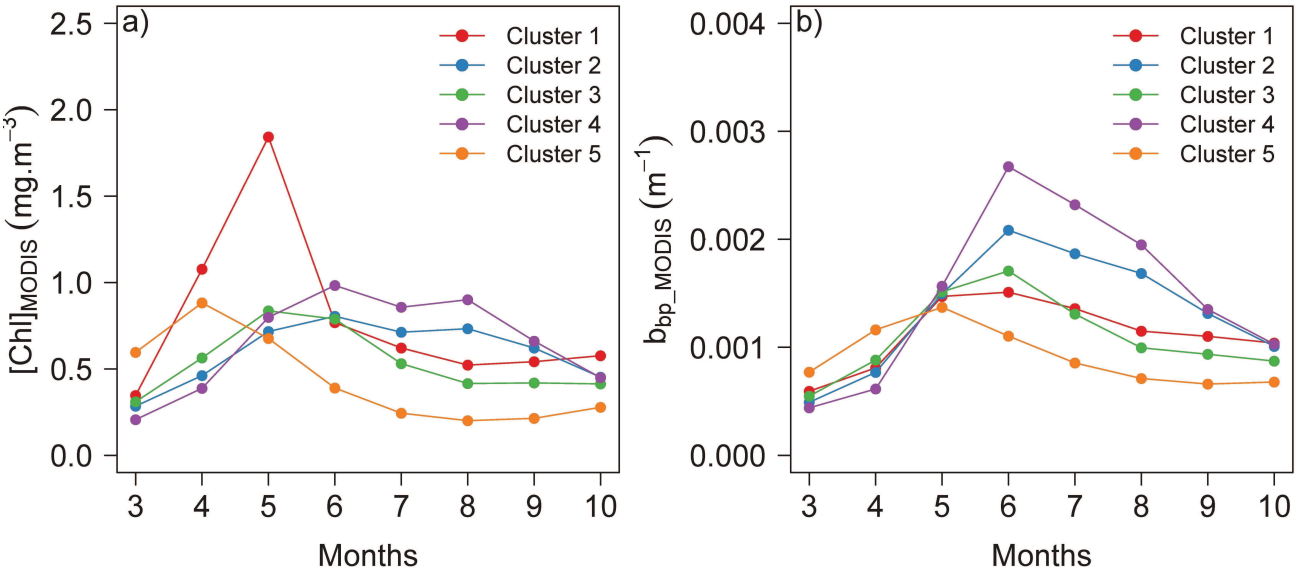


FIGURE 4.9 – *Cycles saisonniers climatologiques moyens a) de la concentration en chlorophylle a estimée par MODIS Aqua ($[Chl]_{MODIS}$) et b) du coefficient de rétrodiffusion de surface estimé par MODIS Aqua (b_{bp_MODIS}) pour les 5 clusters de l'Atlantique Nord.*

Les **profils verticaux moyens mensuels de** b_{bp} sont présentés de la surface à 150 m pour chaque cluster sur la Figure 4.10. Le cluster 1 est caractérisé par des fortes valeurs de b_{bp} du mois d'avril à juillet sur une couche relativement peu profonde (de la surface à ~50 m). Pour les autres clusters, les fortes valeurs de b_{bp} pendant l'été sont visibles sur des couches plus profondes (jusqu'à ~80 m en moyenne). Des valeurs significatives de b_{bp} sont même observées jusqu'à plus de 100 m en mai pour le cluster 1. Les clusters 2 et 3 sont caractérisés par une augmentation abrupte de valeurs élevées

de b_{bp} dans la couche de la surface à ~50-60 m pendant environ deux mois. Il est important de noter que tous les profils ayant de fortes valeurs de b_{bp} dans les couches de surface sont caractérisés par des valeurs de b_{bp} en profondeur encore significative (~0.005 à 0.01 m^{-1}). Le b_{bp} est un estimateur largement reconnu du **POC** (Particulate Organic Carbon) (e.g. Cetinić *et al.*, 2012a, Loisel *et al.*, 2002, Stramski *et al.*, 1999). Ces valeurs de b_{bp} encore relativement élevées en **profondeur** décrivent très probablement l'export des petites particules vers le fond et donc **l'export de carbone** vers les couches profondes de l'océan (e.g. Dall'Olmo et Mork, 2014). Cet export est visible à travers notre approche climatologique du mois de mai au mois de juillet pour tous les clusters (Figure 4.10).

4.3.4 Variabilité saisonnière de la concentration en chlorophylle *a*

Les cycles saisonniers climatologiques moyens de la $[Chl]$ estimée par SOCA-CHL-NA ($[Chl]_{SOCA}$) en surface et intégrée dans la couche productive sont présentés respectivement sur la Figure 4.11-a et 4.11-b pour les 5 clusters caractéristiques de l'Atlantique Nord. Dans le but de comparer les cycles saisonniers de $[Chl]_{SOCA}$ de surface avec ceux de la $[Chl]$ climatologique satellitaire, $[Chl]_{MODIS}$, les dynamiques temporelles de $[Chl]_{MODIS}$ moyennées dans chaque cluster sont représentées sur la Figure 4.9-a. Les cycles saisonniers calculés à partir des données MODIS Aqua semblent assez similaires à ceux provenant de SOCA-CHL-NA. Une différence peut être notée pour le cluster 3 où une diminution de la $[Chl]$ est visible en juin pour la $[Chl]$ de surface estimée par SOCA-CHL-NA tandis que les valeurs sont plus constantes pendant l'été pour les $[Chl]$ dérivées de MODIS Aqua.

En comparant les **valeurs** de $[Chl]$ pour chaque cycle saisonnier issu de SOCA-CHL-NA et de MODIS Aqua, il apparaît une **surestimation** de la $[Chl]_{SOCA}$ moyenne mensuelle par rapport à $[Chl]_{MODIS}$. Ceci correspond à la surestimation qui a été mise en évidence en présentant la validation du modèle SOCA-CHL-NA en surface par rapport aux données satellites pour l'Atlantique Nord (voir Sect. 4.3.2 et la Figure 4.7).

D'après la Figure 4.11-a, les deux clusters 1 et 5 se démarquent particulièrement. Le cluster 1 est caractérisé par une forte augmentation de $[Chl]$ en mai correspondant à des valeurs de $[Chl]$ très élevées en moyenne ($\sim 4mg.m^{-3}$). Le cluster 5, typique des régions tempérées, est caractérisé par une augmentation de $[Chl]$ dès le mois d'avril suivie d'une rapide diminution caractérisée par des faibles valeurs de $[Chl]$ à partir du mois d'aout. Cette diminution est due à la limitation en nutriments en

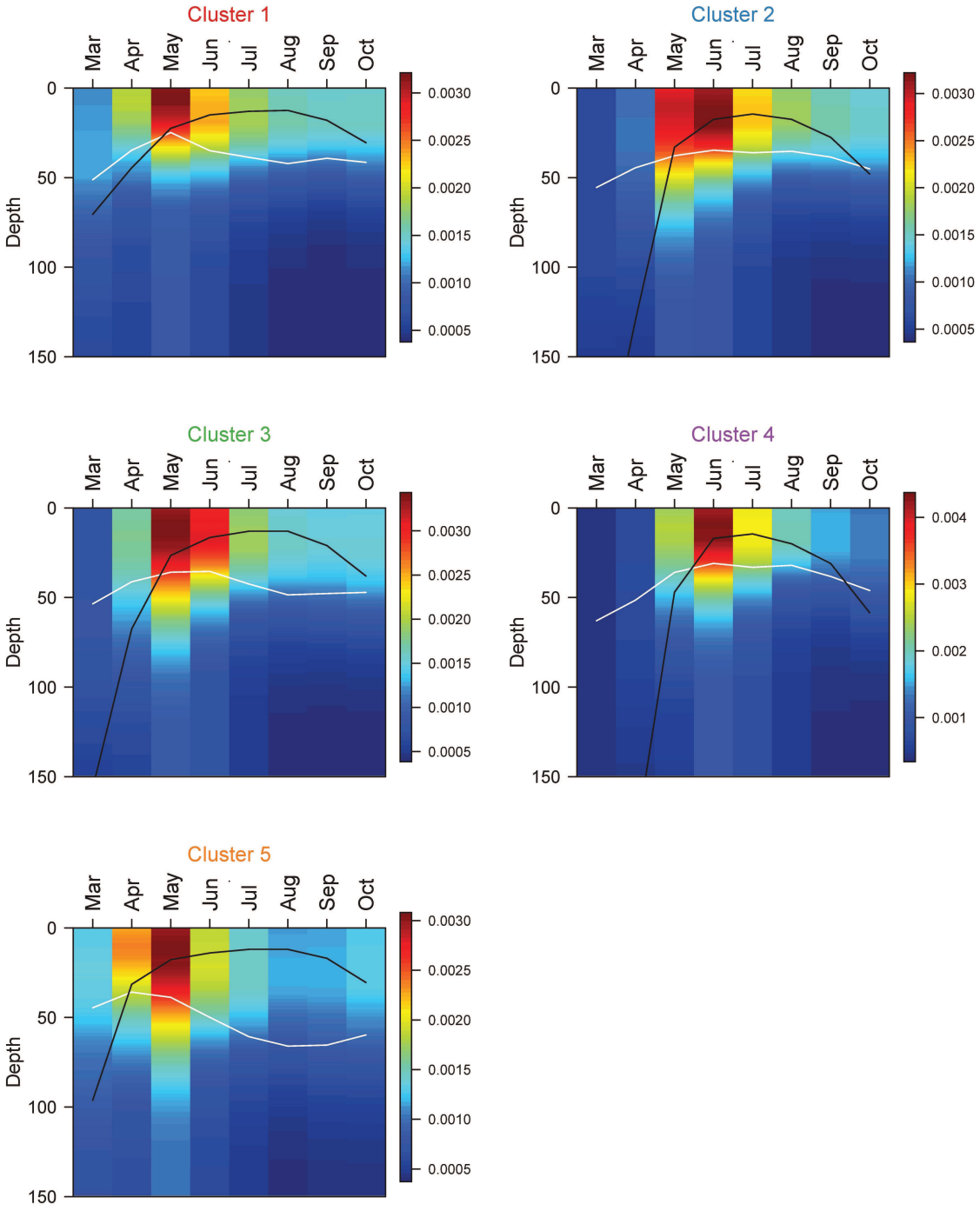


FIGURE 4.10 – *Cycles saisonniers climatologiques moyens de la distribution verticale du coefficient de rétrodiffusion estimé par SOCA-BBP (b_{bp_SOCA} en m^{-1}) pour les 5 clusters de l'Atlantique Nord. La courbe blanche correspond à la profondeur de la couche euphotique (Z_e) moyenne climatologique et la courbe noire correspond à la profondeur de la couche de mélange (Z_m) moyenne climatologique.*

surface, typique de la période estivale en **zone tempérée** (voir Figure 1.2-b du chapitre 1). Les dynamiques temporelles du cluster 1 et du cluster 5 se ressemblent car ce sont des régions ayant la même dynamique hydrologique puisqu'elles sont toutes deux caractérisées par une **stratification** printanière

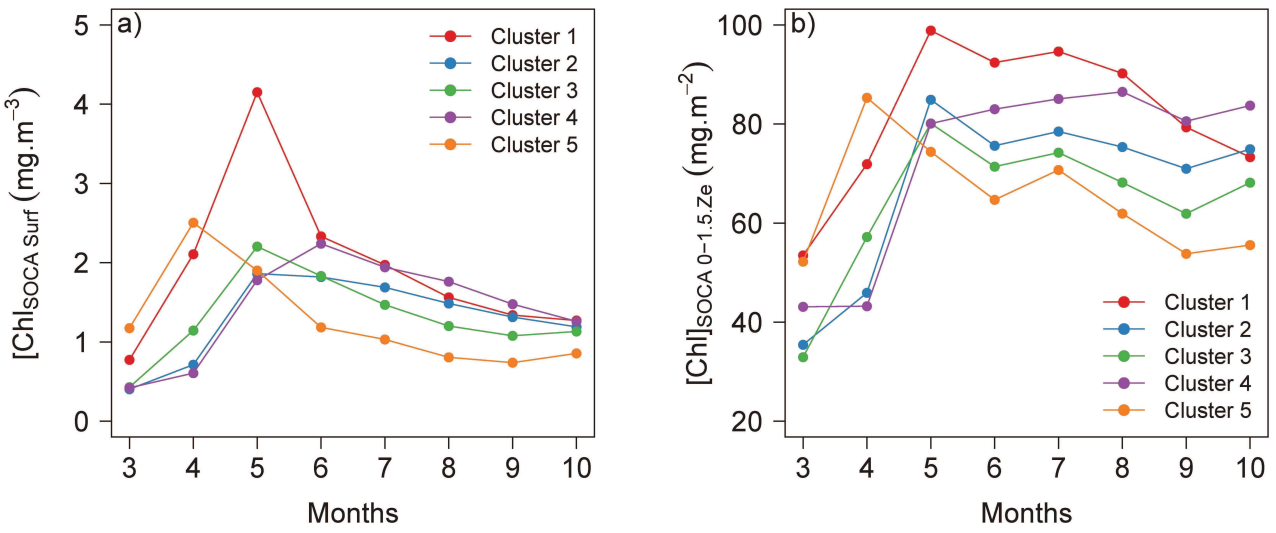


FIGURE 4.11 – *Cycles saisonniers climatologiques moyens de la concentration en chlorophylle a estimée par SOCA-BBP a) en surface et b) intégrée sur la couche* $0 - 1.5Z_e$ *pour les 5 clusters de l'Atlantique Nord.*

relativement précoce de la colonne d'eau. Cette stratification est **thermique** pour le cluster 5 tandis qu'elle est **haline** (due à la fonte des glaces Arctiques) pour le cluster 1. La lumière est disponible plus tôt dans la saison pour le cluster 5 par rapport au Nord de la mer du Labrador (cluster 1), ce qui entraîne une augmentation plus rapide de la $[Chl]$ pour le cluster 5 (mois de mai). Les clusters 2, 3 et 4 sont assez similaires à la fois dans leurs valeurs quantitatives de $[Chl]$ moyennes annuelles mais aussi dans leurs dynamiques. Ces clusters sont typiques des **régions subpolaires** avec une forte augmentation pendant le printemps suivie d'une très légère diminution puisque le phytoplancton n'est pas limité en nutriments pendant l'été. En effet, la nutricline est relativement peu profonde toute l'année dans cette région (voir Figure 1.2-c).

La méthode SOCA-CHL-NA permet d'obtenir en plus de la $[Chl]$ de surface sa **distribution verticale**. A partir de la dimension verticale (présentée plus en détails ci-dessous), des cycles saisonniers moyens de $[Chl]$ **intégrée dans la couche productive** ont été calculés pour chaque cluster (voir Figure 4.11-b). L'évolution saisonnière de ces cycles est assez différente de celle des cycles de $[Chl]$ de surface. En effet, à partir du pic de $[Chl]$ printanier de chaque cluster, la diminution de la $[Chl]$ intégrée est beaucoup plus lente que pour la $[Chl]$ de surface, particulièrement pour les clusters 1, 4 et 5. Le cycle saisonnier du cluster 4 révèle même une $[Chl]$ intégrée qui augmente après le maximum de $[Chl]$ de surface en juin. Cette différence entre les cycles de surface et intégrés est surement due à une **stratification de la colonne d'eau** qui maintient la biomasse dans une couche plus profonde, qui

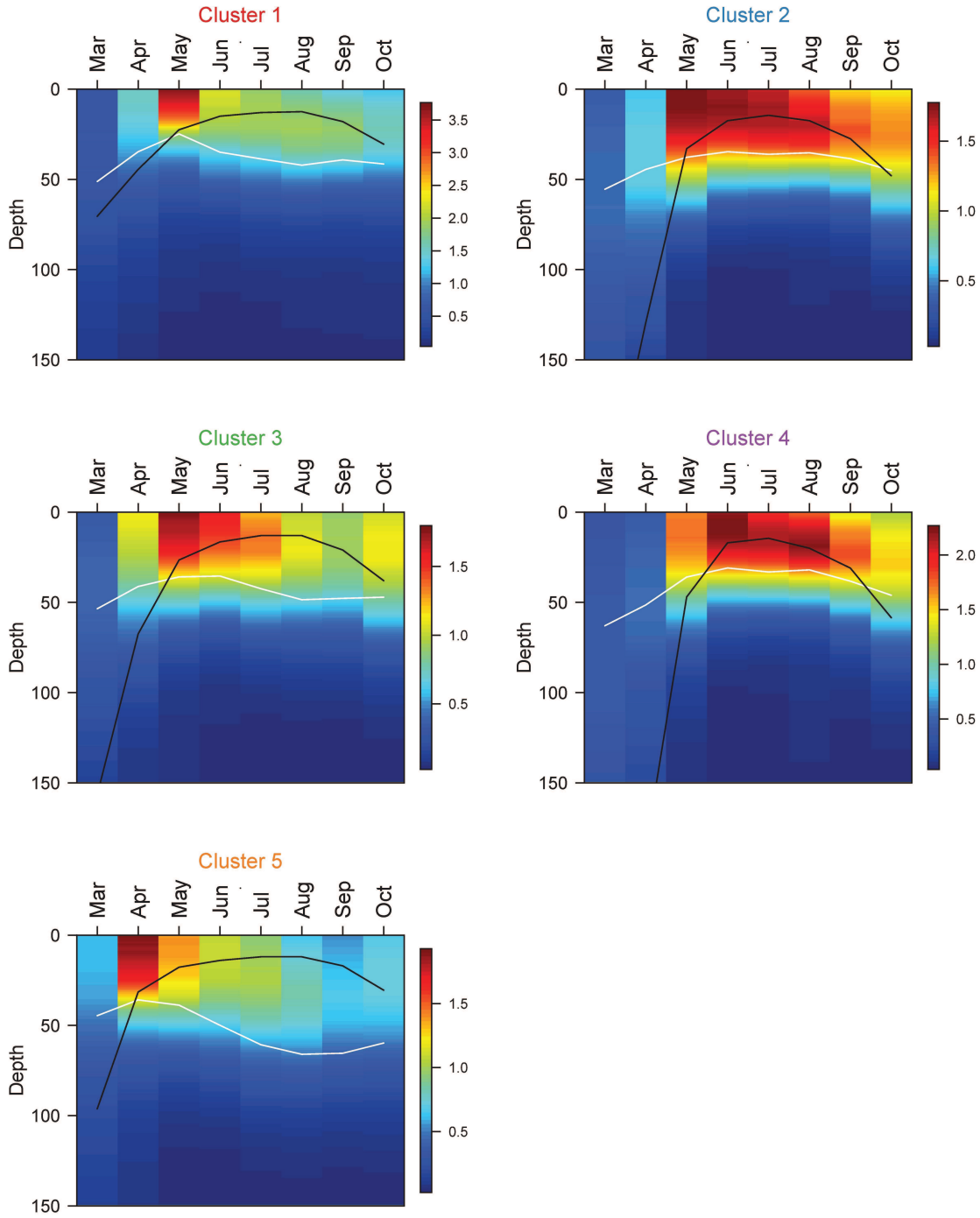


FIGURE 4.12 – *Cycles saisonniers climatologiques moyens de la distribution verticale de la concentration en chlorophylle a estimée par SOCA-BBP ($b_{bp_SOCA}(m^{-1})$) pour les 5 clusters de l'Atlantique Nord. La courbe blanche correspond à la profondeur de la couche euphotique (Z_e) moyenne climatologique et la courbe noire correspond à la profondeur de la couche de mélange (Z_m) moyenne climatologique.*

ne peut être identifiée à partir de données de surface (données satellites).

Afin de vérifier cette hypothèse, les profils **verticaux** moyens de $[Chl]$ sont présentés de la surface à 150 m pour chaque cluster sur la Figure 4.12. Le pic de biomasse du cluster 1 pendant le mois

de mai est confiné dans une couche relativement peu profonde (de la surface à ~25 m). Pour les autres clusters, les valeurs élevées de $[Chl]$ pendant l'été sont visibles sur des couches plus profondes (jusqu'à ~50 m). La dynamique verticale des clusters 1 et 5 se ressemblent particulièrement avec une stratification de la colonne d'eau visible à partir des mois de juillet/aout. Dans les clusters 2, 3 et 4, une légère stratification peut être décelée pour les mois de juillet à septembre. C'est surement cette stratification qui fait que les cycles saisonniers de $[Chl]$ intégrée ne diminuent que très légèrement de juillet à aout pour les clusters 2, 3 et 4.

D'après ces distributions verticales de $[Chl]$ modélisées, le bloom se **déclenche** au moment où la couche de mélange se réduit jusqu'à devenir moins épaisse que la couche euphotique. Ainsi, il semble que pour l'ensemble des clusters et pour l'approche climatologique ici concernée, le mécanisme de déclenchement du bloom est en adéquation avec la théorie de **Sverdrup (1953)**.

4.3.5 Comparaison des cycles saisonniers du coefficient de rétrodiffusion et de la concentration en chlorophylle *a*

Les cycles saisonniers de la $[Chl]$ **intégrée** et du b_{bp} **intégré** dans la couche productive estimés par les méthodes SOCA sont **inter-comparés** (Figure 4.13). Le cluster 5 est caractérisé par un décalage d'un mois entre le maximum de $[Chl]$ intégrée et le maximum de b_{bp} intégré. D'autres études ont permis de déceler un tel **décalage** entre le maximum annuel de $[Chl]$ et le maximum de b_{bp} estimés par satellite à la surface (e.g. Loisel *et al.*, 2002) de l'océan. Loisel *et al.* (2002) soulèvent l'hypothèse que les fortes valeurs de b_{bp} observées en été lorsque la $[Chl]$ a diminué sont due à une accumulation de cellules phytoplanctoniques mortes, ou bien de détritus de zooplancton dans la couche de surface. Les auteurs pensent que ce décalage pourrait également être dû à une diminution du ratio chlorophylle/POC due aux forts éclairements en été.

Les clusters 1, 2, 3 et 4 qui sont caractérisés par des cycles saisonniers relativement **similaires** se démarquent du cluster 5. Un **décalage** entre les cycles de $[Chl]$ intégrée et de b_{bp} intégré apparaît après le pic des deux paramètres bio-optiques (juillet/aout). Les cycles saisonniers climatologiques de **PAR moyenné dans la couche de mélange**, PAR_{ML}, sont présentés pour les 5 clusters de l'Atlantique Nord sur la Figure 4.14. Comme pour les cycles de Chl et b_{bp} intégrés dans la couche productive, le cluster 5 se démarque des quatre autres clusters (i.e. clusters 1, 2 , 3 et 4). Les cycles saisonniers des

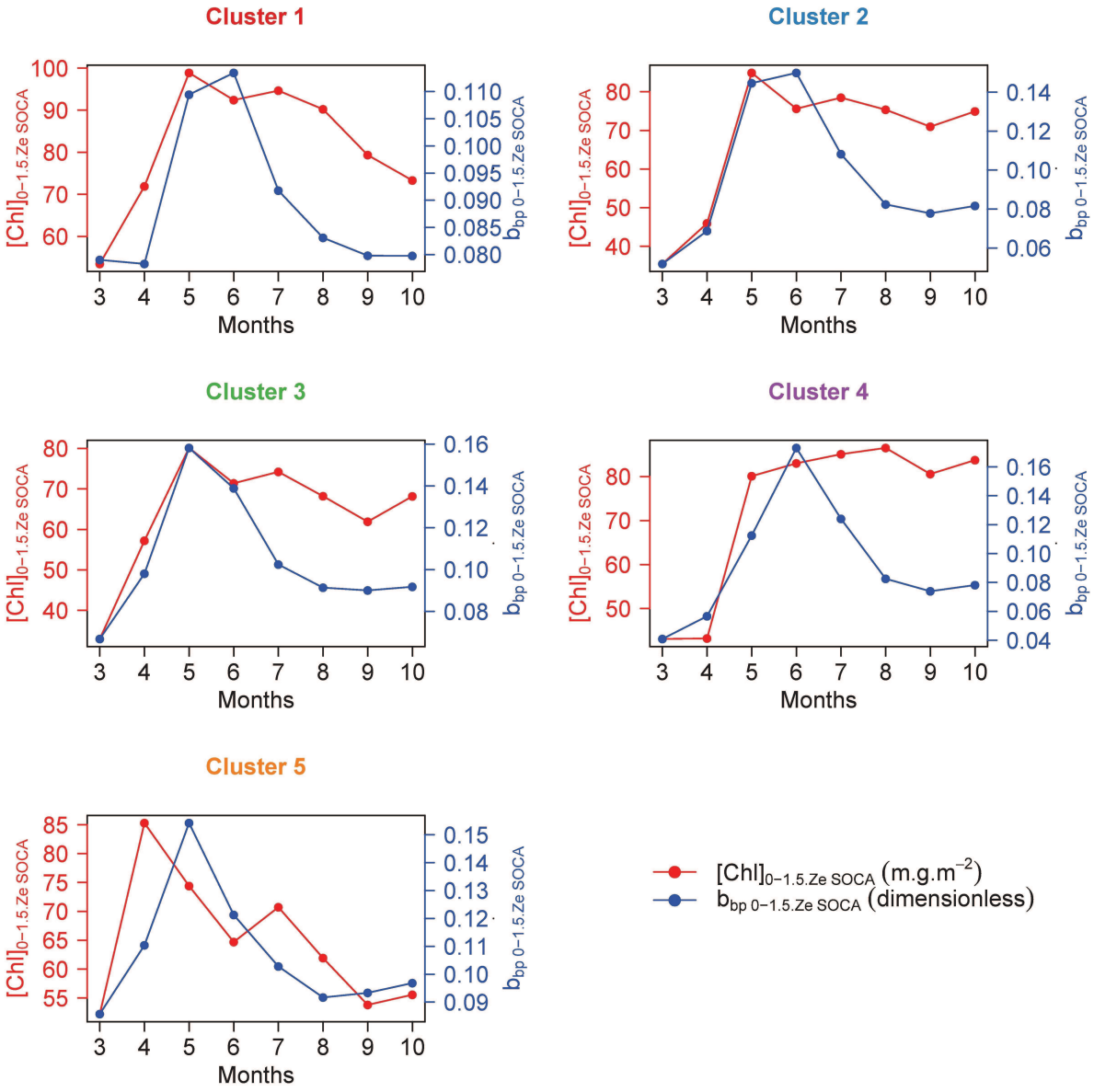


FIGURE 4.13 – *Cycles saisonniers climatologiques moyens superposés de la concentration en chlorophylle a et du coefficient de rétrodiffusion intégrés sur la couche* $0 - 1.5.Z_e$ *pour chaque cluster.*

clusters 1, 2, 3 et 4 sont caractérisés par des faibles valeurs de PAR_{ML} tandis que le cluster 1, pour lequel il n'y a pas de décalage entre la $[Chl]$ et le b_{bp} intégrés dans la couche productive après le bloom printanier, est caractérisé par des valeurs de PAR_{ML} plus élevées. Ainsi, une hypothèse probable pour expliquer le changement dans le rapport **chlorophylle/carbone** à partir des mois de juillet/aout dans les clusters 1, 2, 3 et 4 pourrait éventuellement être la photoadaptation du phytoplancton.

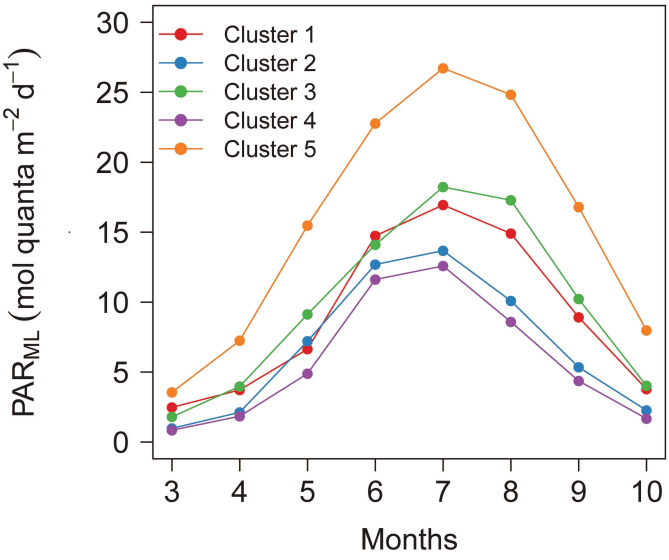


FIGURE 4.14 – *Cycles saisonniers climatologiques moyens du PAR moyenné dans la couche de mélange (PAR_{ML}).*

4.3.6 Variabilité saisonnière de la composition des communautés phytoplanctoniques

En utilisant les profils de la $[Chl]$ associée à chaque **classe de taille phytoplanctonique** (i.e. $[microChl]$, $[nanoChl]$ et $[picoChl]$), les cycles saisonniers moyens des pourcentages de **contribution** de chaque classe à la $[Chl]$ intégrée dans la couche productive ont été calculés (voir Figure 4.15). Les dynamiques saisonnières résultantes sont très similaires pour les clusters 1, 2, 3 et 4. Ces clusters sont caractérisés par des communautés dominées par le nanophytoplancton pendant les mois de mars et avril (contribution entre 40% et 55%), puis par une augmentation de la contribution du microphytoplancton à partir des mois de mai/juin jusqu'à septembre (jusqu'à 80%). Le cluster 5 a une dynamique temporelle propre à lui, avec une augmentation des communautés dominées par le microphytoplancton dès les mois de mars à mai ($\sim$ 55%). Cette augmentation précoce (par rapport aux autres clusters) correspond au pic de biomasse phytoplanctonique ayant lieu plus tôt dans l'année dans cette bio-région (au mois d'avril). Ensuite le microphytoplancton décroit de juin à aout pour laisser place à des communautés dominées par le nano- et le pico-phytoplancton ($\sim$ 40% pour les deux classes de taille). Durant l'été (de juillet à septembre), les communautés dominées par le picophytoplancton sont à leur maximum annuel de contribution. En effet, ce groupe domine les eaux **stratifiées**

qui sont caractéristiques de cette bio-région en été. Finalement, en octobre, les communautés sont de nouveau majoritairement dominées par le microphytoplancton. Ce changement de communautés phytoplanctoniques pourrait éventuellement être dû à un bloom automnal ayant lieu dans les régions tempérées (voir plus loin, Dickey *et al.*, 2001, Findlay *et al.*, 2006).

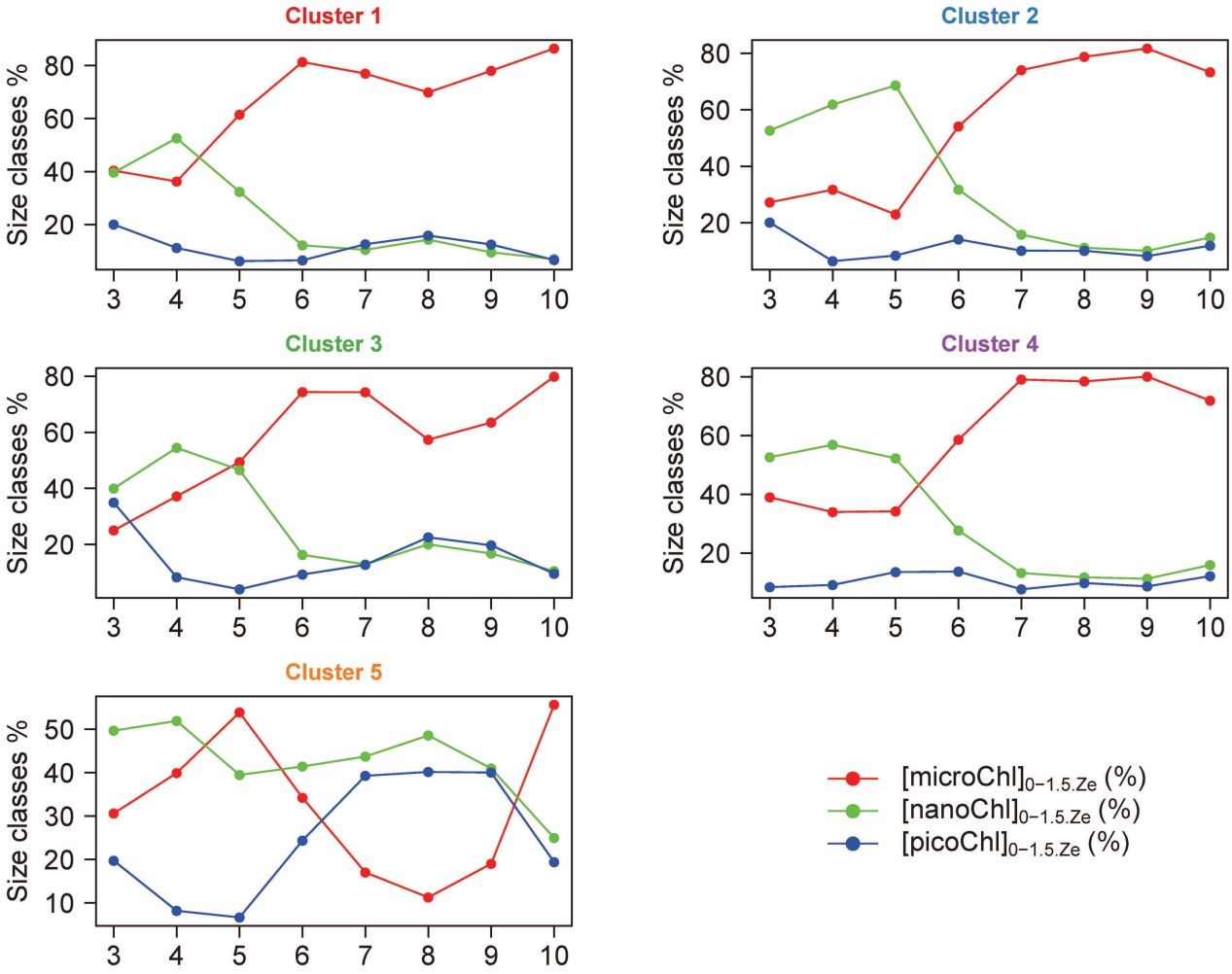


FIGURE 4.15 – *Cycles saisonniers climatologiques moyens des pourcentages de contribution de chaque classe de taille phytoplanctonique à la concentration en chlorophylle a intégrée dans la couche productive* $0 - 1.5.Z_e$ *pour chaque cluster.*

4.3.7 Variabilité saisonnière de la production primaire jusqu'au taux de croissance phytoplanctonique

En utilisant les modèles SOCA, nous avons pu obtenir les distributions verticales mensuelles climatologiques moyennes du b_{bp}, de la $[Chl]$ associée à la biomasse phytoplanctonique totale et associée à sa composition ($[microChl]$, $[nanoChl]$ et $[picoChl]$). Il est possible de dériver plusieurs

paramètres biogéochimiques de ces distributions verticales comme la production primaire, le carbone phytoplanctonique ainsi que le taux de croissance (voir Sect. 4.2.5).

Production primaire

Tout d'abord, le modèle de **production primaire** établi par Morel (1991) a été appliqué dans cette étude en utilisant comme entrées les profils de $[Chl]$ associée aux trois classes de taille phytoplanctoniques déterminés par SOCA-CHL-NA (voir détails dans la Sect. 4.2.4). Ainsi, les productions primaires intégrées dans la couche productive, PP, ont été calculées pour 1) la **biomasse phytoplanctonique totale** et 2) **trois classes de taille phytoplanctoniques** (i.e. micro-, nano- et picophytoplancton). Les dynamiques temporelles climatologiques moyennes de ces propriétés biogéochimiques intégrées verticalement sont présentées sur la Figure 4.16. Ces cycles saisonniers climatologiques montrent que la production primaire totale moyenne résulte essentiellement de la production primaire associée au **microphytoplancton** tout au long de l'année pour les 5 clusters de l'Atlantique Nord. Pour tous les clusters, le pic de production primaire a lieu un mois **plus tard** que le pic de $[Chl]$ de surface. Encore une fois, il est possible de différencier le cluster 5 des autres clusters (i.e clusters 1, 2, 3 et 4) qui ont des dynamiques temporelles très similaires.

Il a été montré précédemment que le cluster 5 est caractéristique des **régions tempérées**. Le cycle moyen temporel de PP_{tot} met en avant un **pic de production primaire** en mai ($1.1\ g.C.m^{-2}.d^{-1}$). Ce bloom printanier est dominé par des larges **diatomées** ($PP_{micro} = 50\% \cdot PP_{tot}$) associées à une légère contribution ($\sim 20\%$) de la classe nanophytoplanctonique. Au milieu de l'été, la **stratification** thermique de la colonne d'eau étant bien en place, les nutriments deviennent limitants, le grazing se développe et la production de diatomées diminue en faveur du nanophytoplancton (Lochte *et al.*, 1993) qui se trouve être souvent des **coccolithophores** (e.g. Iglesias-Rodríguez *et al.*, 2002). Un évènement secondaire de pic de PP_{tot} est visible pour le mois d'octobre ($0.45\ g.C.m^{-2}.d^{-1}$) et est associé à une augmentation de la contribution du microphytoplancton à la production primaire. Dans les régions tempérées, à la fin de l'été, des évènements de mélange épisodiques permettent à la nutricline de remonter et ainsi de favoriser le développement de la biomasse phytoplanctonique (i.e. bloom automnal, Dickey *et al.*, 2001, Findlay *et al.*, 2006). Le second pic de PP_{tot} observé en octobre dans la bio-région 5 pourrait effectivement correspondre à un **bloom automnal** puisque la profondeur de

la couche de mélange Z_m climatologique moyenne s'approfondit à partir de septembre pour le cluster 5 (voir Figure 4.4-a).

Les clusters 1, 2, 3 et 4 sont caractéristiques des **régions subpolaires** (voir Sect. 4.3.1). Les cycles saisonniers moyens de production primaire sont marqués dans ces bio-régions par un pic de production primaire printanier plus élevé que dans les régions tempérées ($PP_{tot} \sim 1.4 g.C.m^{-2}.d^{-1}$). Le **microphytoplancton** contribue largement à ce pic de production primaire visible pour les mois de juin/juillet ($PP_{micro} \simeq 80\% PP_{tot}$). La période de maintien d'une production primaire relativement élevée est plus longue dans les zones sub-polaires que dans les zones tempérées. En effet, la diminution de la PP_{tot} après le pic de juin est beaucoup plus lente pour les clusters 1, 2, 3 et 4 que pour le cluster 1 caractérisé par une diminution abrupte de PP_{tot}. Ces clusters représentant les zones sub-polaires ne semblent pas caractérisés en moyenne par un deuxième pic de production primaire automnale.

Uitz *et al.* (2010) ont publié une étude concernant les évolutions saisonnières et interannuelles de la production primaire associée aux trois classes de taille phytoplanctoniques pour l'océan global. Les auteurs ont étudié l'Atlantique Nord en le séparant en deux grandes zones : la **zone tempérée** allant de 40° à 50° Nord et la **zone sub-polaire** allant de 50° à 70° Nord. Les cycles saisonniers de la production primaire ont été modélisés dans ces deux zones de la même manière que dans l'étude présentée ici (voir Sect. 4.2.4). La différence réside dans les profils verticaux de $[microChl]$, $[nanoChl]$ et $[picoChl]$ utilisés en entrée du modèle de production primaire. Uitz *et al.* (2010) ont estimés ces profils à partir de la méthode de Uitz *et al.* (2006). Finalement, en comparant les cycles saisonniers de la production primaire, les modèles SOCA estiment une PP_{tot} plus élevées dans les deux zones (tempérées et subpolaires) par rapport au modèle de Uitz *et al.* (2010). Uitz *et al.* (2010) modélisent les pics de production primaire printaniers à ~0.9 $g.C.m^{-2}.d^{-1}$ et à ~1.1 $g.C.m^{-2}.d^{-1}$ respectivement pour l'Atlantique Nord tempéré et sub-polaire tandis que les estimations de SOCA sont à ~1.1 $g.C.m^{-2}.d^{-1}$ et ~1.5 $g.C.m^{-2}.d^{-1}$. Cette différence est surement due aux estimations plus élevées du microphytoplancton par SOCA par rapport à celles de Uitz *et al.* (2010) (contributions du microphytoplancton dans les zones tempérées et sub-polaires de l'Atlantique Nord pendant le pic du bloom respectivement estimées à 45% et 50% par Uitz *et al.* (2010) contre 85% et 80% par SOCA). De façon similaire à nos observations, (Uitz *et al.*, 2010) ont mis en évidence une augmentation de la

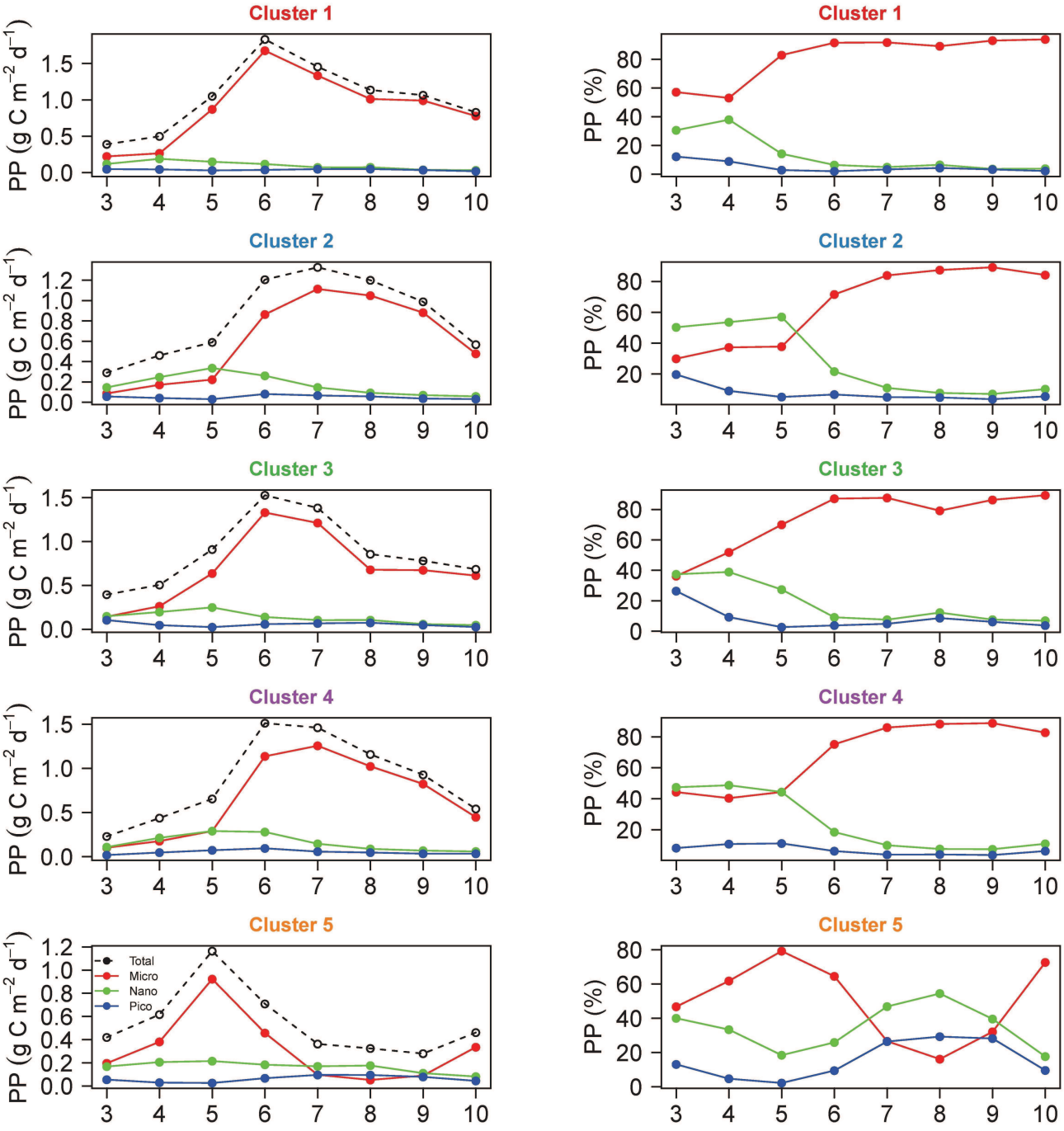


FIGURE 4.16 – *Cycles saisonniers climatologiques moyens de la production primaire intégrée sur la couche productive* $0 - 1.5.Z_e$ *en fonction de la biomasse phytoplanctonique totale et des 3 classes de taille phytoplanctoniques pour les 5 clusters de l'Atlantique Nord. A gauche, les résultats sont présentés en termes de production primaire quantitative* ($g.C.m^{-2}.d^{-1}$) *pour la biomasse phytoplanctonique totale et pour les communautés phytoplanctoniques. A droite, la production primaire est présentée en termes de contribution de chaque classe de taille à la production primaire totale* ($\%$).

production primaire en automne pour la région tempérée de l'Atlantique Nord (de 40° à 50° Nord).

Carbone phytoplanctonique et taux de croissance phytoplanctonique

Comme expliqué dans la Sect. 4.2.5, chaque profil vertical mensuel climatologique moyen a été transformé en profil de **carbone phytoplanctonique** C_{phyto} en utilisant la relation publiée de Graff *et al.* (2015). Cette relation étant linéaire ($C_{phyto} = b_{bp} \cdot 12128 + 0.59$), les dynamiques temporelles du carbone phytoplanctonique ne sont pas discutées puisque ce sont les mêmes que pour le b_{bp}.

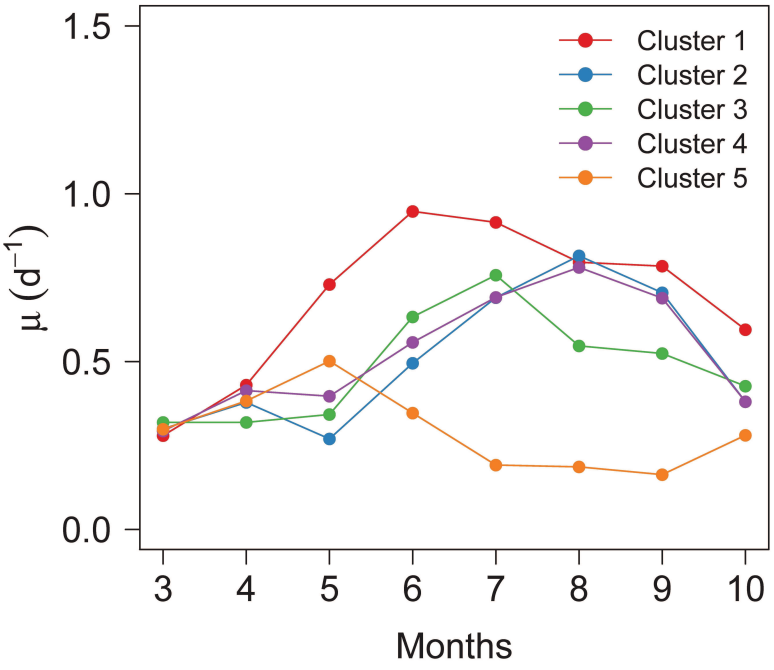


FIGURE 4.17 – *Cycles saisonniers climatologiques moyens du taux de croissance phytoplanctonique (μ) pour chaque cluster de l'Atlantique Nord.*

Des **taux de croissance** phytoplanctonique mensuels climatologiques moyens (μ en d^{-1}) ont été calculés pour chaque cluster de l'Atlantique Nord en divisant la production primaire totale par le carbone phytoplanctonique intégré dans la couche productive $0 - 1.5.Z_e$ (voir détails dans la Sect. 4.2.5). La variabilité des cycles saisonniers de ces taux de croissance est présentée sur la Figure 4.17. La gamme des taux de croissance modélisés dans cette étude est en adéquation avec des estimations déjà publiées (e.g. μ entre 0.5 et 1.5 en Atlantique Nord pendant les mois de juin à aout, Behrenfeld *et al.*, 2005). Comme pour les analyses discutées avant, il apparait sur la Figure 4.17 que les clusters 1 et 5 se démarquent des clusters 2, 3 et 4. En effet, le cluster 1 est caractérisé par des taux de croissance relativement élevés d'avril à octobre. Les taux de croissance moyens les plus élevés dans l'Atlantique Nord (1 d^{-1} au mois de juin) sont estimés pour ce cluster. Le cluster 5 est caractérisé par un pic du taux de croissance en mai suivi de taux de croissance faibles (~0.25 d^{-1} du mois de

juillet à septembre). Les cycles saisonniers des clusters 2, 3 et 4 sont très similaires à la fois par rapport aux valeurs quantitatives estimées de taux de croissance et à leurs évolutions temporelles. Les taux de croissance au sein de ces trois clusters varient au cours de l'année entre ~0.25 et 1 d^{-1}. Leur dynamique temporelle est caractérisée par une lente augmentation du taux de croissance entre mars et juillet/aout, suivie d'une diminution jusqu'à la fin du cycle.

Finalement, les cycles saisonniers climatologiques moyens du rapport carbone phytoplanctonique /chlorophylle, calculés à partir de la $[Chl]$ intégrée dans la couche productive et du C_{phyto} intégré dans la couche productive, sont présentés sur la Figure 4.18. De manière générale, les gammes de valeur estimées correspondent à ce que l'on peut trouver dans la littérature (e.g. Geider *et al.*, 1997). Nous pouvons noter qu'au mois de mars, l'ordre des valeurs du rapport carbone/chlorophylle suit celui du PAR_{ML} (Figure 4.14).

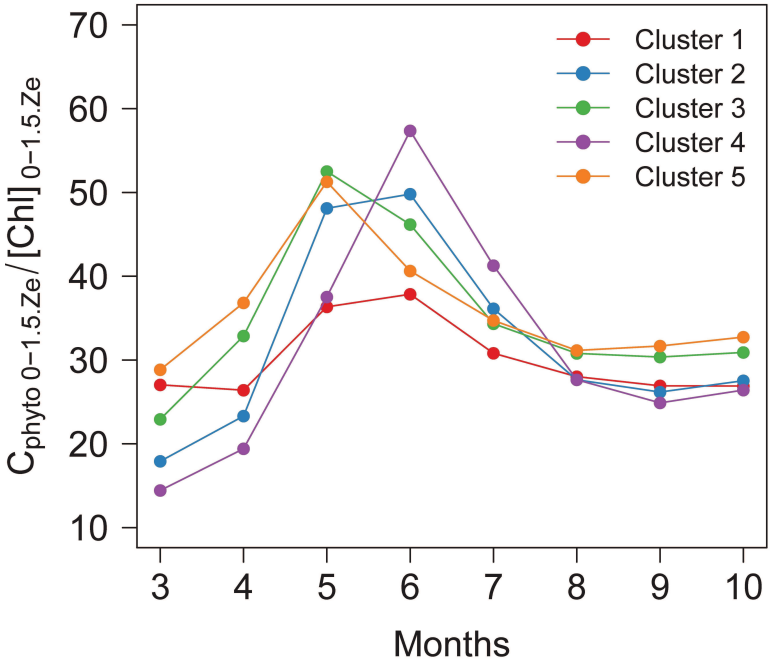


FIGURE 4.18 – *Cycles saisonniers climatologiques moyens du rapport carbone/chlorophylle phytoplanctonique (intégré sur la couche productive) pour chaque cluster de l'Atlantique Nord.*

4.4 Conclusion générale et perspectives

Dans ce chapitre, nous avons appliqué à l'Atlantique Nord des méthodes neuronales pouvant modéliser les distributions verticales 1) du **coefficient de rétrodiffusion particulaire**, b_{bp} et 2) de

la **concentration en chlorophylle *a***, $[Chl]$, associée à la biomasse phytoplanctonique **totale** et aux **trois classes de taille** phytoplanctoniques ($[microChl]$, $[nanoChl]$ et $[picoChl]$) à partir de données satellitaires concommitantes et géo-localisées aux données hydrologiques Argo (matchup).

Un **réseau de neurones régional** a été spécifiquement développé pour cette étude. Cette méthode, nommée SOCA-CHL-NA, permet d'estimer la forme d'un profil de $[Chl]$ à partir d'un matchup satellite/Argo. A partir de cette forme, en utilisant la méthode FLAVOR (présentée dans la Sect. 2.2.2), il est possible d'estimer la $[Chl]$ associée à la **biomasse phytoplanctonique totale** mais aussi aux trois **classes de taille phytoplanctoniques**. Cette méthode régionale développée à partir des données Bio-Argo a ensuite été validée en utilisant des données *in situ* HPLC. Cette validation a montré des résultats relativement satisfaisants. Comparée à la paramétrisation de la distribution verticale des communautés phytoplanctoniques à partir des données couleur de l'eau établie par Uitz *et al.* (2006), même si notre modèle semble moins bien modéliser la contribution du microphytoplancton à la $[Chl]$ totale, notre modèle estime de manière plus précise la contribution du **nanophytoplancton** et du **picophytoplancton** dans la zone de l'Atlantique Nord.

Nous avons montré qu'en utilisant les méthodes SOCA, un grand nombre d'informations intéressantes pouvait être dérivé de ces modèles. Tout d'abord, la **dimension verticale** permet d'avoir une information supplémentaire par rapport aux études basées sur les seules données satellitaires. De plus, les distributions verticales dérivées de nos modèles ont permis grâce à des modèles ou des relations déjà établies de dériver d'autres paramètres bio-géochimiques très utiles dans l'étude de la variabilité de la biomasse phytoplanctonique et de sa dynamique (e.g. PP, C_{phyto}, μ, Morel, 1991, Graff *et al.*, 2015).

De plus, cette étude illustre bien le potentiel de l'utilisation **simultanée** des méthodes SOCA (i.e. SOCA-BBP avec SOCA-CHL-NA). En effet, les estimations **conjointes** des distributions verticales de deux paramètres bio-optiques **complémentaires** que sont le coefficient de rétrodiffusion et la concentration en chlorophylle *a*, est une source d'information nouvelle. Le b_{bp} étant un estimateur du POC (e.g. Cetinić *et al.*, 2012a) et du carbone phytoplanctonique (e.g. Graff *et al.*, 2015), l'utilisation **conjointe** des méthodes SOCA permet d'obtenir des indications sur le rapport chlorophylle/carbone. Certaines études ont déjà montré grâce à des données **simultanées** de b_{bp} et de $[Chl]$ que les changements temporels observés de $[Chl]$ dans certaines régions océaniques pourraient être causés plus par

des modifications physiologiques dans le ratio cellulaire chlorophylle/carbone que par un changement réel de la biomasse phytoplanctonique (Behrenfeld *et al.*, 2005, 2009, Siegel *et al.*, 2013).

Evidemment, cette étude ne repose que sur des **modèles**, ayant chacun leurs **incertitudes** et leurs limites. Le fait d'utiliser une méthode globale pour modéliser la distribution verticale du b_{bp} avec un modèle régional pour la$[Chl]$ et la composition des communautés phytoplanctoniques est surement un point à améliorer. Le développement d'un réseau de neurones régional pour estimer la distribution verticale du b_{bp} spécifiquement en Atlantique Nord pourrait surement permettre d'estimer de manière encore plus précise le b_{bp} et ainsi améliorer l'étude présentée ici.

Finalement, cette approche **climatologique** a permis d'étudier la variabilité **saisonnière** de la distribution verticale de la biomasse phytoplanctonique en Atlantique Nord. Etant donné la densité croissante des profils collectés par les flotteurs profileurs Argo, il serait éventuellement possible d'étudier la **variabilité interannuelle** de la distribution verticale de la biomasse phytoplanctonique en évaluant les déviations par rapport aux tendances moyennes révélées dans cette étude.

CHAPITRE 5

Conclusions générales

Sommaire

5.1 Rappels des objectifs principaux 150

5.2 Résultats majeurs 150

5.3 Intérêts et limites des approches présentées 153

5.4 Travaux futurs et perspectives 154

5.1 Rappels des objectifs principaux

Tout au long de ce travail de thèse, nous nous sommes intéressés à l'étude et la paramétrisation de la **distribution verticale** de la biomasse phytoplanctonique ainsi que de la composition des **communautés phytoplanctoniques** décrites en terme de trois grandes classes de taille (i.e. micro-, nano- et pico-phytoplancton) à l'échelle de l'océan global. Ce travail avait pour objectif principal de mettre à disposition de la communauté océanographique de **nouveaux outils** permettant d'obtenir une vision **tridimensionnelle** de la biomasse phytoplanctonique et de sa composition dans l'océan global. Notre travail, pour une large partie basé sur des développements de **méthodes**, a largement bénéficié des bases de données très denses de profils **bio-optiques** qui se constituent actuellement. Dans ce contexte, nous nous étions fixés **deux objectifs principaux** :

- Utiliser le potentiel des profils de **fluorescence**, estimateur bio-optique largement utilisé de la concentration en chlorophylle *a*, pour développer une base de données de la distribution de la biomasse phytoplanctonique ainsi que de la composition des communautés phytoplanctoniques, indispensable à une représentation **3D** de ces variables à **l'échelle globale**.
- Exploiter l'information ainsi que la haute résolution d'acquisition (spatiale et temporelle) des mesures provenant des capteurs satellitaires **couleur de l'océan** et des **plateformes autonomes** *in situ* en développant une méthode **globale** d'estimation de la distribution verticale de deux **estimateurs bio-optiques** de la biomasse phytoplanctonique (la concentration en chlorophylle *a* et le coefficient de rétrodiffusion particulaire).

5.2 Résultats majeurs

Calibration de la fluorescence en concentration en chlorophylle *a* et classes de taille phytoplanctoniques associées

Comme mentionné plus haut, la **calibration** de la fluorescence de la chlorophylle (notée fluorescence par la suite) en concentration en chlorophylle *a*, $[Chl]$, est nécessaire puisque le **rapport** fluorescence/$[Chl]$ varie en fonction du capteur utilisé, des communautés phytoplanctoniques présentes et de leur état physiologique (dépendant des conditions environnementales). Bien que plusieurs méthodes de calibration aient été développées au cours des dernières années, aucune d'entre elles ne

permet de rendre réellement **intéropérables** tous les profils de fluorescence collectés jusqu'alors pour les utiliser dans leur **globalité**. Nous avons donc développé une méthode neuronale (FLAVOR) permettant de calibrer un profil de fluorescence en utilisant comme données d'entrée la seule **forme** de celui-ci, associée à sa géo-localisation et sa date d'acquisition. Ainsi, FLAVOR permet pour la première fois de « réconcilier » les profils « historiques » avec les profils de « nouvelle génération » provenant des plateformes autonomes *in situ*. Il apparaît donc possible désormais de créer une base de données globale de fluorescence calibrée de manière **homogène** en $[Chl]$ (avec une méthode de calibration unique) sur un continuum temporel allant des années 70 à nos jours et qui pourra également les nouvelles données collectées dans le futur. FLAVOR permet également d'estimer à partir de la forme du profil de fluorescence la distribution verticale de la composition des communautés phytoplanctoniques selon trois **classes de taille** (micro-, nano- et pico-phytoplancton). Grâce à cet outil, il devient possible de quantifier la fraction de biomasse phytoplanctonique associée à chacune des trois classes à partir d'un profil de fluorescence. L'erreur globale de notre méthode est respectivement de 40%, 46%, 35% et 40% pour l'estimation de la $[Chl]$ associée à la biomasse phytoplanctonique totale, au micro-, au nano- et au pico-phytoplancton.

Proposition d'une première base de données globale de concentration en chlorophylle *a* et classes de taille phytoplanctoniques associées à partir de profils de fluorescence

Une base de données **globale** regroupant tous les profils de fluorescence provenant de différentes sources disponibles en ligne et non publiées (~49 000 profils) a été assemblée dans le but de la « soumettre » à la méthode **FLAVOR**. Après l'application d'un **contrôle qualité** spécifiquement développé, la base de données de fluorescence globale a été calibrée en termes de $[Chl]$: 1) associée à la **biomasse phytoplanctonique totale** et 2) associée aux **classes de taille phytoplanctoniques**. Cette base de données globale contenant une nouvelle information sur la distribution verticale de la biomasse phytoplanctonique a été mise à disposition de la communauté océanographique (accessible via l'archive en ligne PANGAEA). Afin de montrer le potentiel d'une telle base de données, nous avons créé les premières climatologies mensuelles et **tridimensionnelles** de la biomasse phytoplanctonique et de sa composition pour quatre provinces écologiques définies par Longhurst (2010).

Paramétrisation de la distribution verticale des propriétés bio-optiques à partir de données satellitaires de la couleur de l'eau

Uitz *et al.* (2006) et Morel et Berthon (1989) ont montré qu'il était possible de relier la concentration en chlorophylle *a* ($[Chl]$) de surface estimée par **satellite** à la **distribution verticale** de la $[Chl]$. Cependant, ces relations globales ont été établies en utilisant relativement **peu de données *in situ*** (i.e. données HPLC). Ces études méritaient d'être **renforcées** en s'appuyant sur la densité croissante des données **bio-optiques** collectées *in situ* par les **flotteurs Bio-Argo**. Ainsi, nous avons développé deux réseaux de neurones capables d'estimer la distribution verticale de deux **propriétés bio-optiques** considérées comme des estimateurs de la biomasse phytoplanctonique. Ces méthodes requièrent comme données d'entrée des données satellites **couleur de l'eau** co-localisées avec un profil hydrologique mesuré par un flotteur profileur **Argo**. Ces réseaux de neurones, SOCA-BBP et SOCA-CHL, estimant respectivement la distribution verticale du coefficient de rétrodiffusion, b_{bp}, et de la $[Chl]$, ont été entrainés à partir d'une base de données de ~5 000 profils hydrologiques et bio-optiques collectés par des **flotteurs profileurs Bio-Argo** de façon concomitante à l'acquisition de mesures satellitaires (i.e. $[Chl]$, b_{bp} et PAR). Les résultats de validation de ces deux méthodes obtenus à partir de bases de données indépendantes sont satisfaisants (i.e. erreurs globales respectivement de 22% et 50% pour l'estimation des distributions verticales de b_{bp} et de $[Chl]$).

Etude climatologique de la variabilité saisonnière en Atlantique Nord

L'étude de la **variabilité saisonnière** en Atlantique Nord a été menée à partir des méthodes développées dans le chapitre 3 (SOCA-BBP et SOCA-CHL). Les méthodes **SOCA** ont pris tout leur sens à travers cette **application régionale**. Les données ont d'abord été regroupées selon des bio-régions identifiées à partir de cycles saisonniers climatologiques de $[Chl]$ de surface estimées par satellite. Sur cette base, nous avons pu mettre en évidence pour cinq grandes provinces écologiques de l'Atlantique Nord des cycles mensuels moyens de la $[Chl]$, de la composition des communautés phytoplanctoniques, du b_{bp} ainsi que des produits pouvant être dérivés de ces grandeurs bio-optiques en utilisant des modèles ou des relations déjà publiées (i.e. la production primaire, la concentration en carbone phytoplanctonique et le taux de croissance). Cette application a permis de mettre en avant le potentiel d'une utilisation **conjointe** des méthodes développées. En effet, la variabilité saisonnière

d'un ensemble de variables biogéochimiques clés et complémentaires a pu être étudiée grâce à SOCA-BBP et SOCA-CHL (i.e. b_{bp}, $[Chl]$, C_{phyto} [1], PP [2], $C_{phyto}/[Chl]$ et μ [3]).

5.3 Intérêts et limites des approches présentées

Aspects méthodologiques

Le développement et l'utilisation de réseaux de neurones ont permis de mettre en évidence leur **potentiel** pour modéliser la distribution verticale de la biomasse phytoplanctonique à partir de différentes données d'entrée caractérisées par leurs variabilités (i.e. profils *in situ* de fluorescence, données satellites fusionnées avec des données hydrologiques Argo). Ces méthodes se sont avérées **robustes** et ont permis d'estimer des réponses biologiques en fonction de différentes variables d'entrées alors que les relations sous-jacentes étaient **complexes** et non décrites de manière mécanistique.

Cependant, les méthodes neuronales étant basée sur un **apprentissage**, elles restent très dépendantes de la **représentativité** de la base de données utilisée lors de cet apprentissage. Les bases de données utilisées pour entrainer les réseaux de neurones développés dans cette thèse se veulent les plus représentatives possibles des conditions hydrologiques, biogéochimiques et trophiques rencontrées dans **l'océan ouvert**. La distribution verticale de la biomasse phytoplanctonique dans certaines zones **particulières** qui ne sont pas représentées dans ces bases de données d'apprentissage (e.g. upwellings) n'est donc surement pas bien modélisée par ce type de méthodes.

Quantité et qualité des données

Bien que les bases de données bio-optiques collectées par les flotteurs Bio-Argo deviennent de plus en plus **denses**, certaines zones restent encore **sous échantillonnées** (e.g. Océan Indien, Océan Austral). Cependant, l'effort de la communauté océanographique est constant pour essayer de progressivement atténuer ce biais spatial d'échantillonnage. Par exemple, depuis quelques temps l'échantillonnage a été intensifié dans l'Océan Austral grâce à des déploiements significatifs dans le cadre de projets dédiés (e.g. SOCCOM, SOCLIM).

Nous avons vu que la calibration de la **fluorescence** en $[Chl]$ est très complexe et soumise à un

1. carbone phytoplanctonique
2. production primaire
3. taux de croissance

certain nombre d'incertitudes (chapitre 1 et chapitre 2). La base de données de b_{bp} a donc l'avantage d'être beaucoup plus homogène à celle de Chl puisque le b_{bp} n'a pas besoin de calibration a posteriori. En effet, les calibrations du constructeur sont considérées comme valables a priori et seules les données acquises par les flotteurs Bio-Argo ont été utilisées. C'est la raison pour laquelle les données de b_{bp} ont été plus faciles à traiter et les méthodes développées à partir de ce paramètre bio-optique apparaissent plus robustes (erreur de SOCA-BBP inférieure à celle de SOCA-CHL). Ainsi, les profils de b_{bp} pourraient offrir une **alternative intéressante** pour estimer la distribution verticale de la **biomasse phytoplanctonique** dans l'océan.

Echelles globale et régionale

L'échelle **globale** a été majoritairement considérée dans cette étude. Ce choix a priori présente ses propres limites et intérêts. En effet, les méthodes que nous avons proposées décrivent des tendances générales pour l'océan global, qu'il s'agisse de la calibration de la fluorescence en $[Chl]$ et des communautés phytoplanctoniques ou bien de la distribution verticale des propriétés bio-optiques à partir des données couleur de l'eau et Argo. Cette approche permet de mettre en avant les **grandes tendances** de la variabilité de la biomasse phytoplanctonique à l'échelle de l'océan global. En revanche, pour des applications à des **échelles spatio-temporelles** moins grandes et caractérisées par des réponses présentant des déviations par rapport aux tendances générales, les outils développés sont certainement moins adaptés et en tous les cas à utiliser en connaissance de cause.

Dans le cas d'études dédiées à une **région spécifique**, il conviendrait certainement de développer une **approche régionale** permettant d'intégrer les éventuelles particularités de la distribution verticale de la biomasse et de la composition des assemblages phytoplanctoniques de la région considérée. Dans un premier temps, de telles études pourraient être envisagées pour des régions où l'échantillonnage Bio-Argo est particulièrement dense, par exemple l'Atlantique Nord (comme nous l'avons proposé dans le chapitre 4) ou la Mer Méditerranée.

5.4 Travaux futurs et perspectives

Développements envisagés

Dans la continuité du travail présenté ici, plusieurs pistes **d'amélioration** sont envisagées. Tout

d'abord nous avons présenté la méthode SOCA-CHL "globale" qui ne permet d'estimer que la $[Chl]$ associée à la biomasse phytoplanctonique totale. Cependant, nous avons montré qu'il était possible de développer un réseau de neurones régional pouvant prédire non seulement la $[Chl]$ associée à la biomasse totale, mais aussi la $[Chl]$ associée à trois **classes de taille** de phytoplancton. Un premier travail consisterait donc à étendre le potentiel de SOCA-CHL à l'estimation de la structure des **communautés phytoplanctoniques** dans l'océan global.

D'autre part, l'étude prospective présentée dans le chapitre 4 a permis d'illustrer le potentiel des deux approches SOCA proposées et de leur utilisation couplée à l'échelle du **bassin** (Atlantique Nord). Il serait donc intéressant de développer une **méthode régionale** permettant d'estimer le b_{bp}, la $[Chl]$ et la composition des communautés phytoplanctoniques pour une autre région d'intérêt biogéochimique majeur : la Mer Méditerranée. La Méditerranée est souvent considérée comme un "océan miniature" (e.g. The MerMex group, 2011). Au Nord-Ouest du bassin, la Mer Ligure présente un cycle saisonnier marqué, caractérisé par un mélange hivernal intense suivi d'une floraison phytoplanctonique au printemps et d'une période estivale oligotrophe. A l'Est, la Mer Levantine présente une oligotrophie quasi-permanente, associée à de faibles $[Chl]$ et à des maxima profonds. La Méditerranée présenterait également un assemblage phytoplanctonique "atypique" comparé à d'autres régions de statut trophique comparable (Vidussi *et al.*, 2001, Navarro *et al.*, 2014, Gitelson *et al.*, 1996). Enfin, cette région se trouve très bien échantillonnée par le réseau d'observation Bio-Argo, ce qui conférerait à une méthode SOCA dédiée une grande robustesse.

Caractère évolutif des méthodes développées

Les résultats présentés dans ce travail de thèse reposent sur l'analyse de bases de données se voulant aussi représentatives que possible de **l'océan global** (i.e. base de données de profils concomitants de fluorescence et de pigments déterminés par HPLC pour FLAVOR, et de profils bio-optiques Bio-Argo pour SOCA). Ces bases de données doivent donc être en **évolution permanente**, mises à jour régulièrement et enrichies grâce aux données collectées lors de nouvelles campagnes en mer ou acquises par les flotteurs profileurs Bio-Argo. Cet enrichissement permettra incontestablement d'améliorer la **représentativité** des bases de données d'apprentissage et de rendre ainsi les méthodes neuronales développées plus **robustes**. Il est donc important de souligner le caractère **évolutif** des

méthodes proposées, avec la possibilité de mises à jour régulières.

Potentiel des données Bio-Argo

Cette thèse a permis de mettre en avant le potentiel des données bio-optiques collectées par les flotteurs profileurs Bio-Argo dans des études menées à **l'échelle globale**. Ce potentiel avait notamment déjà été mentionné par Johnson *et al.* (2009). Cependant, un nombre croissant d'études utilisant des données de flotteurs Bio-Argo portent sur des processus particuliers ayant souvent leurs propres échelles caractéristiques (e.g. Boss *et al.*, 2008, Boss et Behrenfeld, 2010, Xing *et al.*, 2014, Mignot *et al.*, 2014). Les méthodes SOCA présentées ici ont pu être développées grâce la **densité** de profils bio-optiques collectés par les flotteurs Bio-Argo dans l'océan ouvert global. Il est important de noter que ces méthodes pourront également permettre dans le futur d'optimiser les **stratégies d'échantillonnage** des propriétés bio-optiques par les flotteurs Bio-Argo à l'échelle globale. En effet, les zones où l'estimation des indicateurs bio-optiques par les méthodes globales est imprécise sont surement **sous-échantillonnées** par la flotte Bio-Argo.

Perspectives de l'approche climatologique

Les méthodes proposées dans ce travail ont permis de paramétriser la distribution verticale de la biomasse phytoplanctonique et de la structure des communautés aux échelles globale et régionale. Une des applications de ces paramétrisations est la création de **climatologies 3D** de la biomasse phytoplanctonique et des classes de taille phytoplanctoniques associées avec une résolution verticale relativement **fine**. A terme, ces nouvelles climatologies seront utiles à la validation et l'initialisation des **modèles biogéochimiques** intégrant explicitement la distribution verticale de différents groupes phytoplanctoniques (e.g. Aumont *et al.*, 2003, Le Quere *et al.*, 2005, Bopp, 2005). Il est en effet avéré que ces modèles souffrent actuellement d'un déficit de données de validation, bien que des efforts de développement de bases de données de référence soient en cours (e.g. MAREDAT, Peloquin *et al.*, 2013). Par ailleurs, les climatologies ainsi développées pourront également servir de **référence** pour l'identification et la quantification de la variabilité biogéochimique au second ordre. En outre, les climatologies de $[Chl]$ et de b_{bp} issues des outils proposés ici pourront permettre de guider le **contrôle de qualité** des données acquises par les capteurs bio-optiques embarquées sur les plateformes autonomes

(e.g. gamme de valeurs attendue pour une région donnée à une période donnée).

Ainsi, les résultats présentés dans cette thèse mettent en avant le très fort **potentiel** des bases de données *in situ*, en particulier **Bio-Argo**, et de leur "fusion" aux **observations satellitaires** de la couleur de l'océan. De jour en jour, le réseau d'observation Bio-Argo s'intensifie et promet l'acquisition de bases de données bio-optiques, globale et régionale, avec une couverture spatio-temporelle remarquable. Cette thèse propose des **outils novateurs**, ouvrant des **perspectives** intéressantes pour l'exploitation de ces bases de données en biogéochimie marine.

Bibliographie

Abbott M.R., Richerson P.J. et Powell T.M. (1982). In situ response of phytoplankton fluorescence to rapid variations in light. *Limnology and Oceanography*, **27** (2), 218–225, ISSN 00243590, doi :10.4319/lo.1982.27.2.0218.

Agusti S. (2004). Viability and niche segregation of Prochlorococcus and Synechococcus cells across the Central Atlantic Ocean. *Aquatic microbial ecology*, **36** (1), 53–59, ISSN 0948-3055.

Aiken J., Pradhan Y., Barlow R., Lavender S., Poulton A., Holligan P. et 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*, **56** (15), 899–917, ISSN 09670645, doi : 10.1016/j.dsr2.2008.09.017.

Alvain S., Moulin C., Dandonneau Y. et Bréon F. (2005). Remote sensing of phytoplankton groups in case 1 waters from global SeaWiFS imagery. *Deep Sea Research Part I : Oceanographic Research Papers*, **52** (11), 1989–2004, ISSN 09670637, doi :10.1016/j.dsr.2005.06.015.

Antoine D., André J.M. et Morel A. (1996). Oceanic primary production : 2. Estimation at global scale from satellite (Coastal Zone Color Scanner) chlorophyll. *Global Biogeochemical Cycles*, **10** (1), 57–69, ISSN 08866236, doi :10.1029/95GB02832.

Armstrong R.A., Lee C., Hedges J.I., Honjo S. et Wakeham S.G. (2001). A new, mechanistic model for organic carbon fluxes in the ocean based on the quantitative association of POC with ballast minerals. *Deep Sea Research Part II : Topical Studies in Oceanography*, **49** (1-3), 219–236, ISSN 09670645, doi :10.1016/S0967-0645(01)00101-1.

Arnone R., Fargion G., Wang M., Martinolich P., Davis C., Trees C., Ladner S., Lawson A., Zibordi G., Lee Z., Ondrusek M. et Ahmed S. (2012). Ocean Color products from Visible Infared Imager Radiometer Suite

(VIIRS), dans *2012 IEEE International Geoscience and Remote Sensing Symposium*, p. 287–290, IEEE, ISBN 978-1-4673-1159-5, ISSN 2153-6996, doi :10.1109/IGARSS.2012.6351581.

Aumont O., Maier-Reimer E., Blain S. et Monfray P. (2003). An ecosystem model of the global ocean including Fe, Si, P colimitations. *Global Biogeochemical Cycles*, **17** (2), 1060, ISSN 08866236, doi : 10.1029/2001GB001745.

Balch W.M., Drapeau D.T., Fritz J.J., Bowler B.C. et Nolan J. (2001). Optical backscattering in the Arabian Sea—continuous underway measurements of particulate inorganic and organic carbon. *Deep Sea Research Part I : Oceanographic Research Papers*, **48** (11), 2423–2452, ISSN 09670637, doi :10.1016/S0967-0637(01)00025-5.

Beaulieu C., Henson S.a., Sarmiento J.L., Dunne J.P., Doney S.C., Rykaczewski R.R. et Bopp L. (2013). Factors challenging our ability to detect long-term trends in ocean chlorophyll. *Biogeosciences*, **10** (4), 2711–2724, ISSN 1726-4189, doi :10.5194/bg-10-2711-2013.

Behrenfeld M.J. (2010). Abandoning Sverdrup's Critical Depth Hypothesis on phytoplankton blooms. *Ecology*, **91** (4), 977–89, ISSN 0012-9658.

Behrenfeld M.J. et Boss E. (2003). The beam attenuation to chlorophyll ratio : An optical index of phytoplankton physiology in the surface ocean ? *Deep-Sea Research Part I : Oceanographic Research Papers*, **50** (12), 1537–1549, ISSN 09670637, doi :10.1016/j.dsr.2003.09.002.

Behrenfeld M.J. et Boss E. (2006). Beam attenuation and chlorophyll concentration as alternative optical indices of phytoplankton biomass. *Journal of Marine Research*, **64** (3), 431–451.

Behrenfeld M.J., Boss E., Siegel D.A. et Shea D.M. (2005). Carbon-based ocean productivity and phytoplankton physiology from space. *Global Biogeochemical Cycles*, **19** (1), n/a–n/a, ISSN 08866236, doi : 10.1029/2004GB002299.

Behrenfeld M.J., Westberry T.K., Boss E.S., O'Malley R.T., Siegel D.A., Wiggert J.D., Franz B.A., McClain C.R., Feldman G.C., Doney S.C., Moore J.K., Dall'Olmo G., Milligan A.J., Lima I. et Mahowald N. (2009). Satellite-detected fluorescence reveals global physiology of ocean phytoplankton. *Biogeosciences*, **6** (5), ISSN 1726-4189.

Ben Mustapha Z., Alvain S., Jamet C., Loisel H. et 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*, ISSN 00344257, doi :10.1016/j.rse.2013.08.046.

Bishop C.M. (1995). *Neural Networks for Pattern Recognition*, Oxford University Press, Inc., ISBN 0198538642.

Bishop J.K. (2009). Autonomous observations of the ocean biological carbon pump. *Lawrence Berkeley National Laboratory*.

Bishop J.K.B., Davis R.E. et Sherman J.T. (2002). Robotic observations of dust storm enhancement of carbon biomass in the North Pacific. *Science (New York, N.Y.)*, **298** (5594), 817–21, ISSN 1095-9203, doi :10.1126/science.1074961.

Bishop J.K.B. et Wood T.J. (2009). Year-round observations of carbon biomass and flux variability in the Southern Ocean. *Global Biogeochemical Cycles*, **23** (2), GB2019, ISSN 08866236, doi :10.1029/2008GB003206.

Blackford J., Allen J. et Gilbert F. (2004). Ecosystem dynamics at six contrasting sites : a generic modelling study. *Journal of Marine Systems*, **52** (1-4), 191–215, ISSN 09247963, doi :10.1016/j.jmarsys.2004.02.004.

Bopp L. (2005). Response of diatoms distribution to global warming and potential implications : A global model study. *Geophysical Research Letters*, **32** (19), L19606, ISSN 0094-8276, doi :10.1029/2005GL023653.

Boss E. et Behrenfeld M. (2010). In situ evaluation of the initiation of the North Atlantic phytoplankton bloom. *Geophysical Research Letters*, **37** (18), n/a–n/a, ISSN 00948276, doi :10.1029/2010GL044174.

Boss E. et Pegau W.S. (2001). Relationship of Light Scattering at an Angle in the Backward Direction to the Backscattering Coefficient. *Applied Optics*, **40** (30), 5503, ISSN 0003-6935, doi :10.1364/AO.40.005503.

Boss E., Stramski D., Bergmann T., Pegau W.S. et Lewis M. (2004). Why should we measure the optical backscattering coefficient ? *Oceanography*, **17** (2), 44–49.

Boss E., Swift D., Taylor L., Brickley P., Zaneveld R., Riser S., Perry M. et Strutton P. (2008). Observations of pigment and particle distributions in the western North Atlantic from an autonomous float and ocean color satellite.

Boyce D.G., Dowd M., Lewis M.R. et Worm B. (2014). Estimating global chlorophyll changes over the past century. *Progress in Oceanography*, (January), ISSN 00796611, doi :10.1016/j.pocean.2014.01.004.

Boyce D.G., Lewis M.R. et Worm B. (2010). Global phytoplankton decline over the past century. *Nature*, **466** (7306), 591–596, ISSN 1476-4687, doi :10.1038/nature09268.

Boyer T.P., Antonov J.I., Baranova O.K., Garcia H.E., Johnson D.R., Locarnini R.A., Mishonov A.V., Seidov D., Smolyar I.V., Zweng M.M. et Oceanographic N. (2009). *NOAA Atlas NESDIS 66*.

Brasseur P., Gruber N., Barciela R., Brander K., Doron M., El Moussaoui A., Hobday A., Huret M., Kremeur A.S., Lehodey P., Matear R., Moulin C., Murtugudde R., Senina I. et Svendsen E. (2009). Integrating Biogeochemistry and Ecology Into Ocean Data Assimilation Systems. *Oceanography*, **22** (3), 206–215, ISSN 10428275, doi :10.5670/oceanog.2009.80.

Brewin R.J., Sathyendranath S., Hirata T., Lavender S.J., Barciela R.M. et Hardman-Mountford N.J. (2010). A three-component model of phytoplankton size class for the Atlantic Ocean. *Ecological Modelling*, **221** (11), 1472–1483, ISSN 03043800, doi :10.1016/j.ecolmodel.2010.02.014.

Brewin R.J., Sathyendranath S., Lange P.K. et Tilstone G. (2014). Comparison of two methods to derive the size-structure of natural populations of phytoplankton. *Deep Sea Research Part I : Oceanographic Research Papers*, **85**, 72–79, ISSN 09670637, doi :10.1016/j.dsr.2013.11.007.

Bricaud A., Babin M., Morel A. et Claustre H. (1995). Variability in the chlorophyll-specific absorption coefficients of natural phytoplankton : Analysis and parameterization. *Journal of Geophysical Research*, **100** (C7), 13321, ISSN 0148-0227, doi :10.1029/95JC00463.

Bricaud A., Bosc E. et Antoine D. (2002). Algal biomass and sea surface temperature in the Mediterranean Basin. *Remote Sensing of Environment*, **81** (2-3), 163–178, ISSN 00344257, doi :10.1016/S0034-4257(01)00335-2.

Bricaud A., Claustre H., Ras J. et Oubelkheir K. (2004). Natural variability of phytoplanktonic absorption in oceanic waters : Influence of the size structure of algal populations. *Journal of Geophysical Research*, **109** (C11), C11010, ISSN 0148-0227, doi :10.1029/2004JC002419.

Bricaud A., Mejia C., Blondeau-Patissier D., Claustre H., Crepon M. et Thiria S. (2007). Retrieval of pigment concentrations and size structure of algal populations from their absorption spectra using multilayered perceptrons. *Applied optics*, **46** (8), 1251–60, ISSN 0003-6935.

Briggs N., Perry M.J., Cetinić I., Lee C., D'Asaro E., Gray A.M. et Rehm E. (2011). High-resolution observations of aggregate flux during a sub-polar North Atlantic spring bloom. *Deep Sea Research Part I : Oceanographic Research Papers*, **58** (10), 1031–1039, ISSN 09670637, doi :10.1016/j.dsr.2011.07.007.

Brown C., Huot Y., WerdellL P., Gentili B. et Claustre H. (2008). The origin and global distribution of second order variability in satellite ocean color and its potential applications to algorithm development. *Remote Sensing of Environment*, **112** (12), 4186–4203, ISSN 00344257, doi :10.1016/j.rse.2008.06.008.

Bruland K.W., Rue E.L. et Smith G.J. (2001). Iron and macronutrients in California coastal upwelling regimes : Implications for diatom blooms. *Limnology and Oceanography*, **46** (7), 1661–1674, ISSN 00243590, doi : 10.4319/lo.2001.46.7.1661.

Bruland K.W., Rue E.L., Smith G.J. et DiTullio G.R. (2005). Iron, macronutrients and diatom blooms in the Peru upwelling regime : brown and blue waters of Peru. *Marine Chemistry*, **93** (2-4), 81–103, ISSN 03044203, doi :10.1016/j.marchem.2004.06.011.

Buitenhuis E.T., Li W.K.W., Vaulot D., Lomas M.W., Landry M.R., Partensky F., Karl D.M., Ulloa O., Campbell L., Jacquet S., Lantoine F., Chavez F., Macias D., Gosselin M. et McManus G.B. (2012). Picophytoplankton biomass distribution in the global ocean. *Earth System Science Data*, **4** (1), 37–46, ISSN 1866-3516, doi :10.5194/essd-4-37-2012.

Buitenhuis E.T., Vogt M., Moriarty R., Bednaršek N., Doney S.C., Leblanc K., Quéré C.L., Luo Y.W., O'Brien C., O'Brien T. et Others (2013). MAREDAT : towards a world atlas of MARine Ecosystem DATa. *Earth System Science Data*, **5** (2), 227–239.

Campbell J.W. (1995). The lognormal distribution as a model for bio-optical variability in the sea. *Journal of Geophysical Research*, **100** (C7), 13,231–13,254.

Cetinić I., Perry M.J., Briggs N.T., Kallin E., D'Asaro E.A. et Lee C.M. (2012a). Particulate organic carbon and inherent optical properties during 2008 North Atlantic Bloom Experiment. *Journal of Geophysical Research*, **117** (C6), C06028, ISSN 0148-0227, doi :10.1029/2011JC007771.

Cetinić I., Perry M.J., Briggs N.T., Kallin E., D'Asaro E.A. et Lee C.M. (2012b). Particulate organic carbon and inherent optical properties during 2008 North Atlantic Bloom Experiment. *Journal of Geophysical Research*, **117** (C6), C06028, ISSN 0148-0227, doi :10.1029/2011JC007771.

Chisholm S.W. (1992). Phytoplankton size, dans *Primary productivity and biogeochemical cycles in the sea*, p. 213–237, Springer.

Chiswell S. (2011). Annual cycles and spring blooms in phytoplankton : don't abandon Sverdrup completely. *Marine Ecology Progress Series*, **443**, 39–50, ISSN 0171-8630, doi :10.3354/meps09453.

Ciotti A.M., Lewis M.R. et Cullen J.J. (2002). Assessment of the relationships between dominant cell size in natural phytoplankton communities and the spectral shape of the absorption coefficient. *Limnology and Oceanography*, **47** (2), 404–417, ISSN 00243590, doi :10.4319/lo.2002.47.2.0404.

Claustre H. (1994). The trophic status of various oceanic provinces as revealed by phytoplankton pigment signatures. *Limnology and Oceanography*, **39** (5), 1206–1210.

Claustre H., Antoine D., Boehme L., Boss E., DOrtenzio F., DAndon O.F., Guinet C., Gruber N., Handegard N.O., Hood M. et Others (2010a). Guidelines towards an integrated ocean observation system for ecosystems and biogeochemical cycles. *Proceedings of the OceanObs 09 : Sustained Ocean Observations and Information for Society Conference (Vol.1)*, doi :10.5270/OceanObs09.pp.14.

Claustre H., Babin M., Merien D., Ras J., Prieur L., Dallot S., Prasil O., Dousova H. et Moutin T. (2005). Toward a taxon-specific parameterization of bio-optical models of primary production : A case study in the North Atlantic. *Journal of geophysical research*, **110** (C7), C07S12.1–C07S12.17, ISSN 0148-0227.

Claustre H., Bishop J., Boss E., Bernard S., Johnson K., Lotiker A., Ulloa O., Perry M.J., Uitz J., Curie M., Villefranche D., Lazaret C., Division E.S., Berkeley L., Road O.C., Observation E., Africa S., Fermi V., Brest C.D., Valley O. et Jolla L. (2010b). Bio-optical profiling floats as new observational tools for biogeochemical and ecosystem studies : potential synergies with ocean color remote sensing., dans J. Hall, D. Harrison et D. Stammer, rédacteurs, *Proceedings of the OceanObs 09 : Sustained Ocean Observations and Information for Society Conference (Vol.2)*, 1, ESA Publ., Venice, Italy, doi :10.5270/OceanObs09.cwp.17.

Claustre H., Fell F., Oubelkheir K., Prieur L., Sciandra A., Gentili B. et Babin M. (2000). Continuous monitoring of surface optical properties across a geostrophic front : Biogeochemical inferences. *Limnology and oceanography*, **45** (2), 309–321, ISSN 0024-3590.

Claustre H., Hooker S.B., Van Heukelem L., Berthon J.F., Barlow R., Ras J., Sessions H., Targa C., Thomas C.S., van der Linde D. et Marty J.C. (2004). An intercomparison of HPLC phytoplankton pigment methods

using in situ samples : application to remote sensing and database activities. *Marine Chemistry*, **85** (1-2), 41–61, ISSN 03044203, doi :10.1016/j.marchem.2003.09.002.

Claustre H., Kerhervé P., Marty J.C., Prieur L., Videau C. et Hecq J.H. (1994). Phytoplankton dynamics associated with a geostrophic front : Ecological and biogeochemical implications. *Journal of Marine Research*, **52** (4), 711–742, ISSN 00222402, doi :10.1357/0022240943077000.

Claustre H. et Marty J.C. (1995). Specific phytoplankton biomasses and their relation to primary production in the tropical North Atlantic. *Deep Sea Research Part I : Oceanographic Research Papers*, **42** (8), 1475–1493, ISSN 09670637, doi :10.1016/0967-0637(95)00053-9.

Claustre H., Moline M.A. et Prézelin B.B. (1997). Sources of variability in the column photosynthetic cross section for Antarctic coastal waters. *Journal of Geophysical Research*, **102** (C11), 25047, ISSN 0148-0227, doi :10.1029/96JC02439.

Claustre H., Morel A., Babin M., Cailliau C., Marie D., Marty J.C., Tailliez D. et Vaulot D. (1999). Variability in particle attenuation and chlorophyll fluorescence in the tropical Pacific : Scales, patterns, and biogeochemical implications. *Journal of Geophysical Research*, **104** (C2), 3401, ISSN 0148-0227, doi :10.1029/98JC01334.

Collier J.L. (2000). FLOW CYTOMETRY AND THE SINGLE CELL IN PHYCOLOGY. *Journal of Phycology*, **36** (4), 628–644, ISSN 0022-3646, doi :10.1046/j.1529-8817.2000.99215.x.

Conkright M.E., Locarnini R.A., Garcia H.E., O'Brien T.D., Boyer T.P., Stephens C. et Antonov J.I. (2002). *World Ocean Atlas 2001 : Objective analyses, data statistics, and figures : CD-ROM documentation*, US Department of Commerce, National Oceanic and Atmospheric Administration, National Oceanographic Data Center, Ocean Climate Laboratory.

Crombet Y., Leblanc K., Quéguiner B., Moutin T., Rimmelin P., Ras J., Claustre H., Leblond N., Oriol L. et Pujo-Pay M. (2011). Deep silicon maxima in the stratified oligotrophic Mediterranean Sea. *Biogeosciences*, **8** (2), 459–475, ISSN 1726-4189, doi :10.5194/bg-8-459-2011.

Cullen J. (1982). The deep chlorophyll maximum : comparing vertical profiles of chlorophyll a. *Canadian journal of fisheries and aquatic sciences*, ISSN 0706-652X.

Cullen J., Franks P., Karl D. et Longhurst A.R. (2002). Physical influences on marine ecosystem dynamics, dans J.J. McCarthy et B.J. Rothschild, rédacteurs, *The sea, ideas and observations on progress in the study of the seas*, chapitre Tome 12, p. 297–336, john wiley édition.

Cullen J.J. (2015). Subsurface chlorophyll maximum layers : enduring enigma or mystery solved ? *Annual review of marine science*, **7**, 207–39, ISSN 1941-1405, doi :10.1146/annurev-marine-010213-135111.

Cullen J.J. et Lewis M.R. (1995). Biological processes and optical measurements near the sea surface : Some issues relevant to remote sensing. *Journal of Geophysical Research*, **100** (C7), 13255, ISSN 0148-0227, doi :10.1029/95JC00454.

Cunningham A. (1996). Variability of in-vivo chlorophyll fluorescence and its implication for instrument development in bio-optical oceanography.

Cushing D.H. (1959). The seasonal variation in oceanic production as a problem in population dynamics. *Journal du Conseil*, **24** (3), 455–464.

Cybenko G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems*, **2** (4), 303–314, ISSN 0932-4194, doi :10.1007/BF02551274.

Dall'Olmo G. et Mork K.A. (2014). Carbon export by small particles in the Norwegian Sea. *Geophysical Research Letters*, **41** (8), 2921–2927, ISSN 00948276, doi :10.1002/2014GL059244.

de Boyer Montégut C., Madec G., Fischer A.S., Lazar A. et Iudicone D. (2004). Mixed layer depth over the global ocean : An examination of profile data and a profile-based climatology. *Journal of Geophysical Research*, **109** (C12), C12003, ISSN 0148-0227, doi :10.1029/2004JC002378.

Devred E., Sathyendranath S. et Platt T. (2007). Delineation of ecological provinces using ocean colour radiometry. *Marine Ecology Progress Series*, **346**, 1–13.

Devred E., Sathyendranath S., Stuart V. et Platt T. (2011). A three component classification of phytoplankton absorption spectra : Application to ocean-color data. *Remote Sensing of Environment*, **115** (9), 2255–2266, ISSN 00344257, doi :10.1016/j.rse.2011.04.025.

Dickey T., Zedler S., Yu X., Doney S., Frye D., Jannasch H., Manov D., Sigurdson D., McNeil J., Dobeck L., Gilboy T., Bravo C., Siegel D. et Nelson N. (2001). Physical and biogeochemical variability from hours to years at the Bermuda Testbed Mooring site : June 1994–March 1998. *Deep Sea Research Part II : Topical Studies in Oceanography*, **48** (8-9), 2105–2140, ISSN 09670645, doi :10.1016/S0967-0645(00)00173-9.

D'Ortenzio F. et Ribera d'Alcalà M. (2008). On the trophic regimes of the Mediterranean Sea : a satellite analysis. *Biogeosciences Discussions*, **5** (4), 2959–2983.

D'Ortenzio F., Thierry V., Eldin G., Claustre H., Testor P., Coatanoan C., Guinet C., Poteau A., Prieur L., Lefevre D., Bourrin F., Carval T., Goutx M., Thouron D., Lacombe M., Lherminier P., Loisiel H., Mortier L. et Antoine D. (2010). Oceanic Autonomous Platforms For Biogeochemical Studies : Instrumentation and Measure (PABIM), dans *White Book*, December 2009.

Duforêt-Gaurier L., Loisel H., Dessailly D., Nordkvist K. et Alvain S. (2010). Estimates of particulate organic carbon over the euphotic depth from in situ measurements. Application to satellite data over the global ocean. *Deep Sea Research Part I : Oceanographic Research Papers*, **57** (3), 351–367, ISSN 09670637, doi : 10.1016/j.dsr.2009.12.007.

Dugdale R.C. et Goering J.J. (1967). Uptake of new and regenerated forms of nitrogen in primary productivity. *Limnology and Oceanography*, **12** (2), 196–206, ISSN 00243590, doi :10.4319/lo.1967.12.2.0196.

Dutkiewicz S., Follows M., Marshall J. et Gregg W.W. (2001). Interannual variability of phytoplankton abundances in the North Atlantic. *Deep Sea Research Part II : Topical Studies in Oceanography*, **48** (10), 2323–2344, ISSN 09670645, doi :10.1016/S0967-0645(00)00178-8.

Edwards M. et Richardson A.J. (2004). Impact of climate change on marine pelagic phenology and trophic mismatch. *Nature*, **430** (7002), 881–4, ISSN 1476-4687, doi :10.1038/nature02808.

Eppley R.W. et Peterson B.J. (1979). Particulate organic matter flux and planktonic new production in the deep ocean. *Nature*, **282** (5740), 677–680, ISSN 0028-0836, doi :10.1038/282677a0.

Falkowski P., Kiefer D.A., Division O.S. et Angeles L. (1985). Chlorophyll a fluorescence in phytoplankton : relationship to photosynthesis and biomass. *Journal of Plankton Research*, **7** (5), 715–731.

Falkowski P.G., Barber T.R. et Smetacek V. (1998). Biogeochemical Controls and Feedbacks on Ocean Primary Production. *Science*, **281** (5374), 200–206, doi :10.1126/science.281.5374.200.

Feldman G., Kuring N., Ng C., Esaias W., McClain C., Elrod J., Maynard N., Endres D., Evans R., Brown J., Walsh S., Carle M. et Podesta G. (1989). Ocean color : Availability of the global data set. *Eos, Transactions American Geophysical Union*, **70** (23), 634, ISSN 0096-3941, doi :10.1029/89EO00184.

Fennel K. et Boss E. (2003). Subsurface maxima of phytoplankton and chlorophyll : Steady-state solutions from a simple model. *Limnology and Oceanography*, **48** (4), 1521–1534, ISSN 00243590, doi :10.4319/lo.2003.48.4.1521.

Field C.B. (1998). Primary Production of the Biosphere : Integrating Terrestrial and Oceanic Components. *Science*, **281** (5374), 237–240, doi :10.1126/science.281.5374.237.

Findlay H., Yool A., Nodale M. et Pitchford J. (2006). Modelling of autumn plankton bloom dynamics.

Franz B.A., Werdell P.J., Meister G., Bailey S.W., EpleeJr. R.E., Feldman G.C., Kwiatkowskaa E., McClain C.R., Patt F.S. et Thomas D. (2005). <title>The continuity of ocean color measurements from SeaWiFS to MODIS</title>, dans J.J. Butler, rédacteur, *Optics & Photonics 2005*, p. 58820W–58820W–13, International Society for Optics and Photonics, doi :10.1117/12.620069.

Friedrich T. et Oschlies a. (2009). Neural network-based estimates of North Atlantic surface pCO 2 from satellite data : A methodological study. *Journal of Geophysical Research*, **114** (C3), C03020, ISSN 0148-0227, doi :10.1029/2007JC004646.

Funahashi K.I. (1989). On the approximate realization of continuous mappings by neural networks. *Neural Networks*, **2** (3), 183–192, ISSN 08936080, doi :10.1016/0893-6080(89)90003-8.

Gardner W., Mishonov A. et Richardson M. (2006). Global POC concentrations from in-situ and satellite data. *Deep Sea Research Part II : Topical Studies in Oceanography*, **53** (5-7), 718–740, ISSN 09670645, doi : 10.1016/j.dsr2.2006.01.029.

Geider R., MacIntyre H. et Kana T. (1997). Dynamic model of phytoplankton growth and acclimation : responses of the balanced growth rate and the chlorophyll a : carbon ratio to light, nutrient-limitation and temperature. *R.J. Geider, H.L. MacIntyre, T.M. Kana. 1997. Dynamic model of phytoplankton growth and acclimation : responses of the balanced growth rate and the chlorophyll a : carbon ratio to light, nutrient-limitation and temperature. Oceanographic Literature Review 4*, **9** (44), 974, ISSN 0967-0653.

Geider R.J., Platt T. et Raven J.A. (1986). Size dependence of growth and photosynthesis in diatoms : a synthesis. *Mar. Ecol. Prog. Ser*, **30** (9).

Gitelson A., Karnieli A., Goldman N., Yacobi Y. et Mayo M. (1996). Chlorophyll estimation in the Southeastern Mediterranean using CZCS images : adaptation of an algorithm and its validation. *Journal of Marine Systems*, **9** (3-4), 283–290, ISSN 09247963, doi :10.1016/S0924-7963(95)00047-X.

Glover D.M., Jenkins W.J. et Doney S.C. (2011). Modeling Methods for Marine Science. p. 499.

Goldman J.C. (1993). Potential role of large oceanic diatoms in new primary production. *Deep Sea Research Part I : Oceanographic Research Papers*, **40** (1), 159–168, ISSN 09670637, doi :10.1016/0967-0637(93)90059-C.

Gordon H.R. et McCluney W.R. (1975). Estimation of the depth of sunlight penetration in the sea for remote sensing. *Applied optics*, **14** (2), 413–6, ISSN 0003-6935, doi :10.1364/AO.14.000413.

Graff J.R., Westberry T.K., Milligan A.J., Brown M.B., Dall'Olmo G., van Dongen-Vogels V., Reifel K.M. et Behrenfeld M.J. (2015). Analytical phytoplankton carbon measurements spanning diverse ecosystems. *Deep Sea Research Part I : Oceanographic Research Papers*, ISSN 09670637, doi :10.1016/j.dsr.2015.04.006.

Gregg W.W. (2001). Tracking the SeaWiFS record with a coupled physical/biogeochemical/radiative model of the global oceans. *Deep Sea Research Part II : Topical Studies in Oceanography*, **49** (1-3), 81–105, ISSN 09670645, doi :10.1016/S0967-0645(01)00095-9.

Gregg W.W. et Conkright M.E. (2002). Decadal changes in global ocean chlorophyll. *Geophysical Research Letters*, **29** (15), 20–1–20–4, ISSN 00948276, doi :10.1029/2002GL014689.

Gross L., Thiria S., Frouin R. et Mitchell B.G. (2000). Artificial neural networks for modeling the transfer function between marine reflectance and phytoplankton pigment concentration. *Journal of Geophysical Research*, **105** (C2), 3483, ISSN 0148-0227, doi :10.1029/1999JC900278.

Gruber N. (2002). Interannual Variability in the North Atlantic Ocean Carbon Sink. *Science*, **298** (5602), 2374–2378, ISSN 00368075, doi :10.1126/science.1077077.

Guidi L., Stemmann L., Jackson G.A., Ibanez F., Claustre H., Legendre L., Picheral M. et Gorsky G. (2009). Effects of phytoplankton community on production, size, and export of large aggregates : A world-ocean analysis. *Limnology and Oceanography*, **54** (6), 1951–1963, ISSN 00243590, doi :10.4319/lo.2009.54.6.1951.

Guinet C., Xing X., Walker E., Monestiez P., Marchand S., Picard B., Jaud T., Authier M., 'Cotté C., Dragon A.C., Diamond E., Antoine D., Lovell P., Blain S., D'Ortenzio F. et Claustre H. (2012). Calibration procedures and first data set of Southern Ocean chlorophyll a profiles collected by elephant seal equipped with a newly developed CTD-fluorescence tags. *Earth System Science Data Discussions*, **5** (2), 853–891, ISSN 1866-3591, doi :10.5194/essdd-5-853-2012.

Hashimoto S. et Shiomoto A. (2002). Light utilization efficiency of size-fractionated phytoplankton in the subarctic Pacific, spring and summer 1999 : high efficiency of large-sized diatom. *Journal of plankton research*, **24** (1), 83–87.

Haykin S. (1998). Neural Networks : A Comprehensive Foundation.

Henson S., Sarmiento J., Dunne J., Bopp L., Lima I., Doney S., John J. et Beaulieu C. (2010). Detection of anthropogenic climate change in satellite records of ocean chlorophyll and productivity. *Biogeosciences*, **7** (2).

Henson S.A., Robinson I., Allen J.T. et Waniek J.J. (2006). Effect of meteorological conditions on interannual variability in timing and magnitude of the spring bloom in the Irminger Basin, North Atlantic. *Deep Sea Research Part I : Oceanographic Research Papers*, **53** (10), 1601–1615, ISSN 09670637, doi :10.1016/j.dsr.2006.07.009.

Holm-Hansen O., Lorenzen C.J., Holmes R.W. et Strickland J.D.H. (1965). Fluorometric Determination of Chlorophyll. *ICES Journal of Marine Science*, **30** (1), 3–15, ISSN 1054-3139, doi :10.1093/icesjms/30.1.3.

Hood R.R., Laws E.A., Armstrong R.A., Bates N.R., Brown C.W., Carlson C.A., Chai F., Doney S.C., Falkowski P.G., Feely R.A., Friedrichs M.A., Landry M.R., Keith Moore J., Nelson D.M., Richardson T.L., Salihoglu B., Schartau M., Toole D.A. et Wiggert J.D. (2006). Pelagic functional group modeling : Progress, challenges and prospects. *Deep Sea Research Part II : Topical Studies in Oceanography*, **53** (5-7), 459–512, ISSN 09670645, doi :10.1016/j.dsr2.2006.01.025.

Hooker S.B. et Esaias W.E. (1993). An overview of the SeaWiFS Project. *Eos, Transactions American Geophysical Union*, **74** (21), 241, ISSN 0096-3941, doi :10.1029/93EO00945.

Hooker S.B., Van Heukelem L., Thomas C.S., Claustre H., Ras J., Schlüter L., Perl J., Trees C., Stuart V., Head E. et Others (2009). The third sea-wifs HPLC analysis Round-Robin experiment (SeaHARRE-3). *NASA Tech Memo*, **215849**, 2009.

Hopfield J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, **79** (8), 2554–2558, ISSN 0027-8424, doi :10.1073/pnas.79.8.2554.

Hopfield J.J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences*, **81** (10), 3088–3092, ISSN 0027-8424, doi :10.1073/pnas.81.10.3088.

Hornik K., Stinchcombe M. et White H. (1989). Multilayer Feedforward Networks are Universal Approximators. *Neural Networks*, **2**, 359–366.

Huot Y., Babin M. et Bruyant F. (2013). Photosynthetic parameters in the Beaufort Sea in relation to the phytoplankton community structure. *Biogeosciences*, **10** (5), 3445–3454, ISSN 1726-4189, doi :10.5194/bg-10-3445-2013.

Huot Y., Babin M., Bruyant F., Grob C., Twardowski M.S. et Claustre H. (2007). Relationship between photosynthetic parameters and different proxies of phytoplankton biomass in the subtropical ocean. *Biogeosciences*, **4** (5), 853–868.

Huot Y., Morel A., Twardowski M.S., Stramski D. et Reynolds R.A. (2008). Particle optical backscattering along a chlorophyll gradient in the upper layer of the eastern South Pacific Ocean. *Biogeosciences*, **5** (2), 495–507.

Iglesias-Rodríguez M.D., Brown C.W., Doney S.C., Kleypas J., Kolber D., Kolber Z., Hayes P.K. et Falkowski P.G. (2002). Representing key phytoplankton functional groups in ocean carbon cycle models : Coccolithophorids. *Global Biogeochemical Cycles*, **16** (4), 47–1–47–20, ISSN 08866236, doi :10.1029/2001GB001454.

Jamet C., Loisel H. et Dessailly D. (2012). Retrieval of the spectral diffuse attenuation coefficient K d (λ) in open and coastal ocean waters using a neural network inversion. *Journal of Geophysical Research*, **117** (C10), C10023, ISSN 0148-0227, doi :10.1029/2012JC008076.

Jeffrey S., Mantoura R. et Wright S. (1997). Phytoplankton pigments in oceanography : guidelines to modern methods. *Monographs on Oceanographic Methodology*.

Johnson K., Berelson W., Boss E., Chase Z., Claustre H., Emerson S., Gruber N., Kortzinger A., Perry M. et Riser S. (2009). Observing Biogeochemical Cycles at Global Scales With Profiling Floats and Gliders Prospects for a Global Array. *Oceanography*, **22** (3), 216 – 225.

Johnson K.S., Coletti L.J., Jannasch H.W., Sakamoto C.M., Swift D.D. et Riser S.C. (2013). Long-Term Nitrate Measurements in the Ocean Using the in situ Ultraviolet Spectrophotometer : Sensor Integration into the

APEX Profiling Float. *Journal of Atmospheric and Oceanic Technology*, **30** (8), 1854–1866, ISSN 0739-0572, doi :10.1175/JTECH-D-12-00221.1.

Kiefer D.A. (1973). Chlorophyll a fluorescence in marine centric diatoms : Responses of chloroplasts to light and nutrient stress. *Marine Biology*, **23** (1), 39–46, ISSN 0025-3162, doi :10.1007/BF00394110.

Kishi M.J., Kashiwai M., Ware D.M., Megrey B.A., Eslinger D.L., Werner F.E., Noguchi-Aita M., Azumaya T., Fujii M., Hashimoto S., Huang D., Iizumi H., Ishida Y., Kang S., Kantakov G.A., Kim H.c., Komatsu K., Navrotsky V.V., Smith S.L., Tadokoro K., Tsuda A., Yamamura O., Yamanaka Y., Yokouchi K., Yoshie N., Zhang J., Zuenko Y.I. et Zvalinsky V.I. (2007). NEMURO—a lower trophic level model for the North Pacific marine ecosystem. *Ecological Modelling*, **202** (1-2), 12–25, ISSN 03043800, doi :10.1016/j.ecolmodel.2006.08.021.

Kohonen T. (1990). The self-organizing map. *Proceedings of the IEEE*, **78** (9), 1464–1480, ISSN 00189219, doi :10.1109/5.58325.

Kokhanovsky A.A. (2012). *Light Scattering Reviews 7 : Radiative Transfer and Optical Properties of Atmosphere and Underlying Surface*, ISBN 3642219071.

Kolber Z. et Falkowski P.G. (1993). Use of active fluorescence to estimate phytoplankton photosynthesis in situ. *Limnology and Oceanography*, **38** (8), 1646–1665, ISSN 00243590, doi :10.4319/lo.1993.38.8.1646.

Körtzinger A., Schimanski J., Send U. et Wallace D. (2004). The ocean takes a deep breath. *Science (New York, N.Y.)*, **306** (5700), 1337, ISSN 1095-9203, doi :10.1126/science.1102557.

Kostadinov T.S., Siegel D.A. et Maritorena S. (2009). Retrieval of the particle size distribution from satellite ocean color observations. *Journal of Geophysical Research*, **114** (C9), C09015, ISSN 0148-0227, doi :10.1029/2009JC005303.

Kostadinov T.S., Siegel D.A. et Maritorena S. (2010). Global variability of phytoplankton functional types from space : assessment via the particle size distribution. *Biogeosciences*, **7** (10), 3239–3257, ISSN 1726-4189, doi :10.5194/bg-7-3239-2010.

Krasnopolsky V.M. (2009). Neural Network Applications to Solve Forward and Inverse Problems in Atmospheric and Oceanic Satellite Remote Sensing. p. 191–205.

Lacour L., Claustre H., Prieur L. et D'Ortenzio F. (2015). Phytoplankton biomass cycles in the North Atlantic subpolar gyre : A similar mechanism for two different blooms in the Labrador Sea. *Geophysical Research Letters*, **42** (13), 5403–5410, ISSN 00948276, doi :10.1002/2015GL064540.

Lavigne H., D'Ortenzio F., Claustre H. et Poteau A. (2012). Towards a merged satellite and in situ fluorescence ocean chlorophyll product. *Biogeosciences*, **9** (6), 2111–2125, ISSN 1726-4189, doi :10.5194/bg-9-2111-2012.

Lavigne H., D'Ortenzio F., Ribera D'Alcalà M., Claustre H., Sauzède R. et Gacic M. (2015). On the vertical distribution of the chlorophyll a concentration in the Mediterranean Sea : a basin-scale and seasonal approach. *Biogeosciences*, **12** (16), 5021–5039, ISSN 1726-4189, doi :10.5194/bg-12-5021-2015.

Laws E.A., Falkowski P.G., Smith W.O., Ducklow H. et McCarthy J.J. (2000). Temperature effects on export production in the open ocean. *Global Biogeochemical Cycles*, **14** (4), 1231–1246, ISSN 08866236, doi : 10.1029/1999GB001229.

Le Quere C., Harrison S., Prentice I., Buitenhuis E., Aumont O., Bopp L., Claustre H., de Cunha L., Geider R., Giraud X., Klaas C., Kohfield K., Legendre L., Manizza M., Platt T., Rivkin R., Sathyendranath S., Uitz J., Watson A. et Wolf-Gladrow D. (2005). Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models.

Le Quéré C., Raupach M.R., Canadell J.G., Marland et al. G., Le Quéré et al. C., Marland G., Bopp L., Ciais P., Conway T.J., Doney S.C., Feely R.A., Foster P., Friedlingstein P., Gurney K., Houghton R.A., House J.I., Huntingford C., Levy P.E., Lomas M.R., Majkut J., Metzl N., Ometto J.P., Peters G.P., Prentice I.C., Randerson J.T., Running S.W., Sarmiento J.L., Schuster U., Sitch S., Takahashi T., Viovy N., van der Werf G.R. et Woodward F.I. (2009). Trends in the sources and sinks of carbon dioxide. *Nature Geoscience*, **2** (12), 831–836, ISSN 1752-0894, doi :10.1038/ngeo689.

Leblanc K., Aristegui J., Armand L., Assmy P., Beker B., Bode A., Breton E., Cornet V., Gibson J., Gosselin M.P., Kopczynska E., Marshall H., Peloquin J., Piontkovski S., Poulton A., Queguiner B., Schiebel R., Shipe R., Stefels J., van Leeuwe M., Varela M., Widdicombe C. et Yallop M. (2012). A global diatom database – abundance, biovolume and biomass in the world ocean.

Lee Z., Carder K.L. et Arnone R.A. (2002). Deriving Inherent Optical Properties from Water Color : a Multiband Quasi-Analytical Algorithm for Optically Deep Waters. *Applied Optics*, **41** (27), 5755, ISSN 0003-6935, doi :10.1364/AO.41.005755.

Lee Z. et Huot Y. (2014). On the non-closure of particle backscattering coefficient in oligotrophic oceans. *Optics express*, **22** (23), 29223–33, ISSN 1094-4087, doi :10.1364/OE.22.029223.

Lee Z., Lubac B., Werdell J. et Arnone R. (2009). An update of the quasi-analytical algorithm (QAA_v5). *International Ocean Color Group Software Report*.

Legendre L. et Le Fevre J. (1991). From individual plankton cells to pelagic marine ecosystems and to global biogeochemical cycles. *NATO ASI series : Series G : Ecological sciences*.

Lek S. et Guégan J. (1999). Artificial neural networks as a tool in ecological modelling, an introduction. *Ecological Modelling*, **120** (2-3), 65–73, ISSN 03043800, doi :10.1016/S0304-3800(99)00092-7.

Letelier R.M., Karl D.M., Abbott M.R. et Bidigare R.R. (2004). Light driven seasonal patterns of chlorophyll and nitrate in the lower euphotic zone of the North Pacific Subtropical Gyre. *Limnology and Oceanography*, **49** (2), 508–519, ISSN 00243590, doi :10.4319/lo.2004.49.2.0508.

Levitus S., Antonov J.I., Baranova O.K., Boyer T.P., Coleman C.L., Garcia H.E., Grodsky A.I., Johnson D.R., Locarnini R.A., Mishonov A.V. et Others (2013). The World Ocean Database. *Data Science Journal*, **12** (0), WDS229—-WDS234.

Lewis M.R., Cullen J.J. et Platt T. (1983). Phytoplankton and thermal structure in the upper ocean : Consequences of nonuniformity in chlorophyll profile. *Journal of Geophysical Research*, **88** (C4), 2565, ISSN 0148-0227, doi :10.1029/JC088iC04p02565.

Lippmann R. (1987). An introduction to computing with neural nets. *IEEE ASSP Magazine*, **4** (2), 4–22, ISSN 0740-7467, doi :10.1109/MASSP.1987.1165576.

Liu H., Probert I., Uitz J., Claustre H., Aris-Brosou S., Frada M., Not F. et de Vargas C. (2009). Extreme diversity in noncalcifying haptophytes explains a major pigment paradox in open oceans. *Proceedings of the National Academy of Sciences of the United States of America*, **106** (31), 12803–8, ISSN 1091-6490, doi :10.1073/pnas.0905841106.

Lochte K., Ducklow H., Fasham M. et 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*, **40** (1-2), 91–114, ISSN 09670645, doi :10.1016/0967-0645(93)90008-B.

Loisel H., Bosc E., Stramski D., Oubelkheir K. et Deschamps P.Y. (2001). Seasonal variability of the backscattering coefficient in the Mediterranean Sea based on satellite SeaWiFS imagery. *Geophysical Research Letters*, **28** (22), 4203–4206, ISSN 00948276, doi :10.1029/2001GL013863.

Loisel H., Mériaux X., Berthon J.F. et Poteau A. (2007). Investigation of the optical backscattering to scattering ratio of marine particles in relation to their biogeochemical composition in the eastern English Channel and southern North Sea. *Limnology and Oceanography*, **52** (2), 739–752, ISSN 00243590, doi :10.4319/lo.2007.52.2.0739.

Loisel H., Nicolas J.M., Deschamps P.Y. et Frouin R. (2002). Seasonal and inter-annual variability of particulate organic matter in the global ocean. *Geophysical Research Letters*, **29** (24), 2196, ISSN 0094-8276, doi : 10.1029/2002GL015948.

Loisel H., Nicolas J.M., Sciandra A., Stramski D. et Poteau A. (2006). Spectral dependency of optical backscattering by marine particles from satellite remote sensing of the global ocean. *Journal of Geophysical Research*, **111** (C9), C09024, ISSN 0148-0227, doi :10.1029/2005JC003367.

Longhurst A.R. (2010). *Ecological Geography of the Sea*, ISBN 0080465579.

Lorenzen C.J. (1966). A method for the continuous measurement of in vivo chlorophyll concentration. *Deep Sea Research and Oceanographic Abstracts*, **13** (2), 223–227, ISSN 00117471, doi :10.1016/0011-7471(66)91102-8.

Lorenzen C.J. (1967). Vertical distribution of chlorophyll and phaeo-pigments : Baja California. *Deep Sea Research and Oceanographic Abstracts*, **14** (6), 735–745, ISSN 00117471, doi :10.1016/S0011-7471(67)80010-X.

MacIntyre H.L., Kana T.M., Anning T. et Geider R.J. (2002). Photoacclimatation of photosynthesis irradiance response curves and photosynthetic pigments in microalgae and cyanobacteria. *Journal of Phycology*, **38** (1), 17–38, ISSN 0022-3646, doi :10.1046/j.1529-8817.2002.00094.x.

Malone T. (1980). Algal size, dans I. Morris, rédacteur, *The physiological ecology of phytoplankton*, p. 433–463, university édition.

Mantoura R. et Llewellyn C. (1983). The rapid determination of algal chlorophyll and carotenoid pigments and their breakdown products in natural waters by reverse-phase high-performance liquid chromatography. *Analytica Chimica Acta*, **151**, 297–314, ISSN 00032670, doi :10.1016/S0003-2670(00)80092-6.

Margalef R. (1967). Some concepts relative to the organization of plankton. *Oceanography and Marine Biology : An Annual Review*.

Marinov I., Doney S.C. et Lima I.D. (2010). Response of ocean phytoplankton community structure to climate change over the 21st century : partitioning the effects of nutrients, temperature and light. *Biogeosciences*, **7** (12), 3941–3959, ISSN 1726-4189, doi :10.5194/bg-7-3941-2010.

Martinez E., Antoine D., D'Ortenzio F. et Gentili B. (2009). Climate-driven basin-scale decadal oscillations of oceanic phytoplankton. *Science (New York, N.Y.)*, **326** (5957), 1253–6, ISSN 1095-9203, doi :10.1126/science.1177012.

Martinez-Vicente V., Dall'Olmo G., Tarran G., Boss E. et Sathyendranath S. (2013). Optical backscattering is correlated with phytoplankton carbon across the Atlantic Ocean. *Geophysical Research Letters*, **40** (6), 1154–1158, ISSN 00948276, doi :10.1002/grl.50252.

Marty J.C., Chiavérini J., Pizay M.D. et 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*, **49** (11), 1965–1985, ISSN 09670645, doi :10.1016/S0967-0645(02)00022-X.

Marzban C. (2009). Basic Statistics and Basic AI : Neural Networks, dans S.E. Haupt, A. Pasini et C. Marzban, rédacteurs, *Artificial Intelligence Methods in the Environmental Sciences*, chapitre 2, p. 15–47, ISBN 978-1-4020-9117-9.

McClain C.R. (2009). A decade of satellite ocean color observations. *Annual review of marine science*, **1**, 19–42, ISSN 1941-1405, doi :10.1146/annurev.marine.010908.163650.

McClain C.R., Feldman G.C. et Hooker S.B. (2004). An overview of the SeaWiFS project and strategies for producing a climate research quality global ocean bio-optical time series. *Deep Sea Research Part II : Topical Studies in Oceanography*, **51** (1-3), 5–42, ISSN 09670645, doi :10.1016/j.dsr2.2003.11.001.

McCulloch W.S. et Pitts W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, **5** (4), 115–133, ISSN 0007-4985, doi :10.1007/BF02478259.

Michaels A.F. et Silver M.W. (1988). Primary production, sinking fluxes and the microbial food web. *Deep Sea Research Part A. Oceanographic Research Papers*, **35** (4), 473–490, ISSN 01980149, doi :10.1016/0198-0149(88)90126-4.

Mignot A., Claustre H., D'Ortenzio F., Xing X., Poteau A. et Ras J. (2011). From the shape of the vertical profile of in vivo fluorescence to Chlorophyll-a concentration. *Biogeosciences*, **8** (2), 3697–3737, ISSN 1810-6285, doi :10.5194/bgd-8-3697-2011.

Mignot A., Claustre H., Uitz J., Poteau A., D'Ortenzio F. et Xing X. (2014). Understanding the seasonal dynamics of phytoplankton biomass and the deep chlorophyll maximum in oligotrophic environments : A Bio-Argo float investigation. *Global Biogeochemical Cycles*, **28** (8), 856–876, ISSN 08866236, doi : 10.1002/2013GB004781.

Mignot A., Ferrari R. et Mork K.A. (2015). Spring bloom onset in the Nordic Seas. *Biogeosciences Discussions*, **12** (16), 13631–13673, ISSN 1810-6285, doi :10.5194/bgd-12-13631-2015.

Moloney C.L. et Field J.G. (1991). The size-based dynamics of plankton food webs. I. A simulation model of carbon and nitrogen flows. *Journal of Plankton Research*, **13** (5), 1003–1038, ISSN 0142-7873, doi : 10.1093/plankt/13.5.1003.

Morel A. (1991). Light and marine photosynthesis : a spectral model with geochemical and climatological implications. *Progress in Oceanography*, **26** (3), 263–306, ISSN 00796611, doi :10.1016/0079-6611(91) 90004-6.

Morel A. et Ahn Y.H. (1991). Optics of heterotrophic nanoflagellates and ciliates : A tentative assessment of their scattering role in oceanic waters compared to those of bacterial and algal cells. *Journal of Marine Research*, **49** (1), 177–202, ISSN 00222402, doi :10.1357/002224091784968639.

Morel A. et Antoine D. (1994). Heating rate within the upper ocean in relation to its bio-optical state. *Journal of physical oceanography*, **24** (7), 1652–1665, ISSN 0022-3670.

Morel A. et Berthon J.F. (1989). Surface Pigments, Algal Biomass Profiles, and Potential Production of the Euphotic Layer : Relationships Reinvestigated in View of Remote-Sensing Applications. *Limnology and Oceanography*, **34** (8), 1545–1562.

Morel A. et Bricaud A. (1981). Theoretical results concerning light absorption in a discrete medium, and application to specific absorption of phytoplankton. *Deep Sea Research Part A. Oceanographic Research Papers*, **28** (11), 1375–1393, ISSN 01980149, doi :10.1016/0198-0149(81)90039-X.

Morel A., Gentili B., Chami M. et Ras J. (2006). Bio-optical properties of high chlorophyll Case 1 waters and of yellow-substance-dominated Case 2 waters. *Deep Sea Research Part I : Oceanographic Research Papers*, **53** (9), 1439–1459, ISSN 09670637, doi :10.1016/j.dsr.2006.07.007.

Morel A. et Maritorena S. (2001). Bio-optical properties of oceanic waters : A reappraisal. *Journal of Geophysical Research*, **106** (C4), 7163, ISSN 0148-0227, doi :10.1029/2000JC000319.

Muller P. (2001). Non-Photochemical Quenching. A Response to Excess Light Energy. *PLANT PHYSIOLOGY*, **125** (4), 1558–1566, ISSN 00320889, doi :10.1104/pp.125.4.1558.

Navarro G., Alvain S., Vantrepotte V. et Huertas I. (2014). Identification of dominant phytoplankton functional types in the Mediterranean Sea based on a regionalized remote sensing approach. *Remote Sensing of Environment*, **152**, 557–575, ISSN 00344257, doi :10.1016/j.rse.2014.06.029.

Niang a., Badran F., Moulin C., Crépon M. et 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*, **100** (1), 82–94, ISSN 00344257, doi :10.1016/j.rse.2005.10.005.

Niewiadomska K., Claustre H., Prieur L. et D'Ortenzio F. (2008). Submesoscale physical-biogeochemical coupling across the Ligurian current (northwestern Mediterranean) using a bio-optical glider. *Limnology and Oceanography*, **53** (5_part_2), 2210–2225, ISSN 00243590, doi :10.4319/lo.2008.53.5_part_2.2210.

O'Brien C.J., Peloquin J., Vogt M., Heinle M., Gruber N., Ajani P., Andruleit H., Aristegui J., Beaufort L., Estrada M., Karentz D., Kopczynska E., Lee R., Pritchard T. et Widdicombe C. (2013). Global marine plankton functional type biomass distributions : coccolithophores. *Earth System Science Data*, **5**, 259–276.

Olson R.J., Zettler E.R. et Anderson O.K. (1989). Discrimination of eukaryotic phytoplankton cell types from light scatter and autofluorescence properties measured by flow cytometry. *Cytometry*, **10** (5), 636–43, ISSN 0196-4763, doi :10.1002/cyto.990100520.

O'Reilly J.E., Maritorena S., Mitchell B.G., Siegel D.A., Carder K.L., Garver S.A., Kahru M. et McClain C. (1998). Ocean color chlorophyll algorithms for SeaWiFS. *Journal of Geophysical Research*, **103** (C11), 24937, ISSN 0148-0227, doi :10.1029/98JC02160.

Palacz A.P., John M.A.S., Brewin R.J.W., Hirata T. et Gregg W.W. (2013). Distribution of phytoplankton functional types in high-nitrate, low-chlorophyll waters in a new diagnostic ecological indicator model. *Biogeosciences*, **10** (11), 7553–7574, ISSN 1726-4189, doi :10.5194/bg-10-7553-2013.

Partensky F., Hess W.R. et Vaulot D. (1999). Prochlorococcus, a marine photosynthetic prokaryote of global significance. *Microbiology and molecular biology reviews : MMBR*, **63** (1), 106–27, ISSN 1092-2172.

Peloquin J., Swan C., Gruber N., Vogt M., Claustre H., Ras J., Uitz J., Barlow R., Behrenfeld M., Bidigare R., Dierssen H., Ditullio G., Fernandez E., Gallienne C., Gibb S., Goericke R., Harding L., Head E., Holligan P., Hooker S., Karl D., Landry M., Letelier R., Llewellyn C.A., Lomas M., Lucas M., Mannino A., Marty J.C., Mitchell B.G., Muller-Karger F., Nelson N., O'Brien C., Prezelin B., Repeta D., Jr. Smith W.O., Smythe-Wright D., Stumpf R., Subramaniam A., Suzuki K., Trees C., Vernet M., Wasmund N. et Wright S. (2013). The MAREDAT global database of high performance liquid chromatography marine pigment measurements. *Earth System Science Data*, **5** (1), 109–123, ISSN 1866-3516, doi :10.5194/essd-5-109-2013.

Platt T. et Denman K. (1977). Organisation in the pelagic ecosystem. *Helgoländer Wissenschaftliche Meeresuntersuchungen*, **30** (1-4), 575–581, ISSN 0017-9957, doi :10.1007/BF02207862.

Platt T. et Denman K. (1978). The structure of pelagic marine ecosystems., dans *International Council for the Exploration of the Sea, Charlottenlund (Denmark). Marine Ecosystems and Fisheries Oceanography. Edinburgh (UK). 13 Sep 1976.*

Platt T., Sathyendranath S., Caverhill C.M. et Lewis M.R. (1988). Ocean primary production and available light : further algorithms for remote sensing. *Deep Sea Research Part A. Oceanographic Research Papers*, **35** (6), 855–879, ISSN 01980149, doi :10.1016/0198-0149(88)90064-7.

Poggio T. et Girosi F. (1990). Networks for approximation and learning. *Proceedings of the IEEE*, **78** (9), 1481–1497, ISSN 00189219, doi :10.1109/5.58326.

Post A. (2006). The genus Prochlorococcus, dans E. Stackebrandt, D. Jones et W. Holzapfel, rédacteurs, *The Prokaryotes*, p. 1099–1110, springer édition.

Raitsos D.E., Lavender S.J., Maravelias C.D., Haralabous J., Richardson A.J. et Reid P.C. (2008). Identifying four phytoplankton functional types from space : An ecological approach. *Limnology and Oceanography Oceanogr*, **53** (2), 605–613.

Ras J., Claustre H. et Uitz J. (2008). Spatial variability of phytoplankton pigment distributions in the Subtropical South Pacific Ocean : comparison between in situ and predicted data. *Biogeosciences*, **5** (2), 353–369.

Rast M., Bezy J.L. et Bruzzi S. (1999). The ESA Medium Resolution Imaging Spectrometer MERIS a review of the instrument and its mission. *International Journal of Remote Sensing*, **20** (9), 1681–1702, ISSN 0143-1161, doi :10.1080/014311699212416.

Reynolds C.S. (2006). *The Ecology of Phytoplankton*, ISBN 1139454897.

Richardson A., Silulwane N., Mitchell-Innes B. et Shillington F. (2003). A dynamic quantitative approach for predicting the shape of phytoplankton profiles in the ocean. *Progress in Oceanography*, **59** (2-3), 301–319, ISSN 00796611, doi :10.1016/j.pocean.2003.07.003.

Riser S.C. et Johnson K.S. (2008). Net production of oxygen in the subtropical ocean. *Nature*, **451** (7176), 323–5, ISSN 1476-4687, doi :10.1038/nature06441.

Rochford P.A., Kara A.B., Wallcraft A.J. et Arnone R.A. (2001). Importance of solar subsurface heating in ocean general circulation models. *Journal of Geophysical Research*, **106** (C12), 30923, ISSN 0148-0227, doi :10.1029/2000JC000355.

Roemmich D., Belbéoc M., Freeland H., Garzoli S., Gould J., Grant F., Ignaszewski M., King B., Klein B., Le Traon P.Y., Mork K.A., Brechner Owens W., Pouliquen S., Ravichandran M., Riser S., Sterl A., Suga T., Suk M.S., Sutton P. et Thierry V. (2009a). Argo : The Challenge of Continuing 10 Years of Progress. *Oceanography*, **22** (3), 46–55, ISSN 1042-8275.

Roemmich D., Boebel O., Freeland H., King B. et LeTraon P.Y. (1999). *On the design and implementation of Argo - An initial plan for a global array of profiling floats*, c/o 2605 Bureau of Meteorology, Melbourne, Australia, the godae édition.

Roemmich D. et Gilson J. (2009). The 2004–2008 mean and annual cycle of temperature, salinity, and steric height in the global ocean from the Argo Program. *Progress in Oceanography*, **82** (2), 81–100, ISSN 00796611, doi :10.1016/j.pocean.2009.03.004.

Roemmich D., Johnson G., Riser S., Davis R., Gilson J., Owens W.B., Garzoli S., Schmid C. et Ignaszewski M. (2009b). The Argo Program : Observing the Global Oceans with Profiling Floats. *Oceanography*, **22** (2), 34–43, ISSN 10428275, doi :10.5670/oceanog.2009.36.

Roy S., Llewellyn C.A., Egeland E.S. et Johnsen G. (2011). *Phytoplankton pigments : characterization, chemotaxonomy and applications in oceanography*, Cambridge University Press.

Rumelhart D.E., Hinton G.E. et Williams R.J. (1988). Learning representations by back-propagating errors. *Cognitive modeling*, **5**.

Rumelhart D.E., McClelland J.L., Group P.D.P.R. et Others (1986). Parallel distributed processing, Vols 1 and 2. *Cambridge, MA : The MIT Press*.

Salomonson V.V., Barnes W., Montgomery H. et Ostrow H. (1987). MODIS- Advanced facility instrument for studies of the earth as a system((Moderate Resolution Imaging Spectrometer)), dans *IGARSS'87- International Geoscience and Remote Sensing Symposium, Ann Arbor, MI*, p. 361–365.

Sanders R., Henson S.A., Koski M., De La Rocha C.L., Painter S.C., Poulton A.J., Riley J., Salihoglu B., Visser A., Yool A., Bellerby R. et Martin A.P. (2014). The Biological Carbon Pump in the North Atlantic. *Progress in Oceanography*, **129**, 200–218, ISSN 00796611, doi :10.1016/j.pocean.2014.05.005.

Sathyendranath S., Lazzara L. et Prieur L. (1987). Variations in the spectral values of specific absorption of phytoplankton. *Limnology and Oceanography*, **32** (2), 403–415, ISSN 00243590, doi :10.4319/lo.1987.32.2.0403.

Sathyendranath S., Platt T., Caverhill C.M., Warnock R.E. et Lewis M.R. (1989). Remote sensing of oceanic primary production : computations using a spectral model. *Deep Sea Research Part A. Oceanographic Research Papers*, **36** (3), 431–453, ISSN 01980149, doi :10.1016/0198-0149(89)90046-0.

Sathyendranath S., Watts L., Devred E., Platt T., Caverhill C. et Maass H. (2004). Discrimination of diatoms from other phytoplankton using ocean-colour data. *Marine ecology. Progress series*, **272**, 59–68, ISSN 0171-8630.

Sauzède R., Claustre H., Jamet C., Uitz J., Ras J., Mignot A. et D'Ortenzio F. (2015a). Retrieving the vertical distribution of chlorophyll a concentration and phytoplankton community composition from in situ fluorescence profiles : A method based on a neural network with potential for global-scale applications. *Journal of Geophysical Research : Oceans*, **120** (1), 451–470, ISSN 21699275, doi :10.1002/2014JC010355.

Sauzède R., Claustre H., Uitz J., Jamet C., Dall'Olmo G., D'Ortenzio F., Gentili B., Poteau A. et Schmechtig C. (2015b). A neural network for merging ocean color and Argo data to extend surface bio-optical properties to depth : particulate backscattering coefficient retrieval (Submitted). *Journal of Geophysical Research : Oceans*.

Sauzède R., Claustre H., Uitz J., Jamet C., D'Ortenzio F., Gentili B., Schmechtig C. et Poteau A. (2015c). A neural network-based method for merging ocean color and Argo data to extend surface bio-optical properties to depth : chlorophyll a concetration retrieval (In prep.).

Sauzède R., Lavigne H., Claustre H., Uitz J., Schmechtig C., D'Ortenzio F., Guinet C. et Pesant S. (2015d). Vertical distribution of chlorophyll a concentration and phytoplankton community composition from in situ fluorescence profiles : a first database for the global ocean. *Earth System Science Data Discussions*, **8** (1), 365–399, ISSN 1866-3591, doi :10.5194/essdd-8-365-2015.

Savtchenko A., Ouzounov D., Ahmad S., Acker J., Leptoukh G., Koziana J. et Nickless D. (2004). Terra and Aqua MODIS products available from NASA GES DAAC. *Advances in Space Research*, **34** (4), 710–714, ISSN 02731177, doi :10.1016/j.asr.2004.03.012.

Schmitz W.J. et McCartney M.S. (1993). On the North Atlantic Circulation. *Reviews of Geophysics*, **31** (1), 29, ISSN 8755-1209, doi :10.1029/92RG02583.

Sieburth J.M., Smetacek V. et Lenz J. (1978). Pelagic ecosystem structure : Heterotrophic compartments of the plankton and their relationship to plankton size fractions. *Limnology and Oceanography*, **23** (6), 1256–1263, ISSN 00243590, doi :10.4319/lo.1978.23.6.1256.

Siegel D., Behrenfeld M., Maritorena S., McClain C., Antoine D., Bailey S., Bontempi P., Boss E., Dierssen H., Doney S., Eplee R., Evans R., Feldman G., Fields E., Franz B., Kuring N., Mengelt C., Nelson N., Patt F., Robinson W., Sarmiento J., Swan C., Werdell P., Westberry T., Wilding J. et Yoder J. (2013). Regional to global assessments of phytoplankton dynamics from the SeaWiFS mission. *Remote Sensing of Environment*, **135**, 77–91, ISSN 00344257, doi :10.1016/j.rse.2013.03.025.

Siegel D.A. (2002a). The North Atlantic Spring Phytoplankton Bloom and Sverdrup's Critical Depth Hypothesis. *Science*, **296** (5568), 730–733, ISSN 00368075, doi :10.1126/science.1069174.

Siegel D.A. (2002b). The North Atlantic Spring Phytoplankton Bloom and Sverdrup's Critical Depth Hypothesis. *Science*, **296** (5568), 730–733, ISSN 00368075, doi :10.1126/science.1069174.

Siegel D.A., Maritorena S., Nelson N. et Behrenfeld M.J. (2005). Independence and interdependencies among global ocean color properties : Reassessing the bio-optical assumption. *Journal of Geophysical Research*, **110** (C7), C07011, ISSN 0148-0227, doi :10.1029/2004JC002527.

Smayda T.J. (1980). Phytoplankton species succession. *Studies in ecology (USA). v. 7.*

Sontag E.D. (1992). Feedback stabilization using two-hidden-layer nets. *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*, **3** (6), 981–90, ISSN 1045-9227, doi :10.1109/72.165599.

Stramska M. (2009). Particulate organic carbon in the global ocean derived from SeaWiFS ocean color. *Deep Sea Research Part I : Oceanographic Research Papers*, **56** (9), 1459–1470, ISSN 09670637, doi :10.1016/j.dsr.2009.04.009.

Stramski D., Reynolds R.A., Babin M., Kaczmarek S., Lewis M.R., Röttgers R., Sciandra A., Stramska M., Twardowski M.S., Franz B.A. et Claustre H. (2008). Relationships between the surface concentration of particulate organic carbon and optical properties in the eastern South Pacific and eastern Atlantic Oceans. *Biogeosciences*, **5** (1), 171–201.

Stramski D., Reynolds R.A., Kahru M. et Mitchell B.G. (1999). Estimation of Particulate Organic Carbon in the Ocean from Satellite Remote Sensing. *Science*, **285** (5425), 239–242, ISSN 00368075, doi :10.1126/science.285.5425.239.

Sullivan J.M. et Twardowski M.S. (2009). Angular shape of the oceanic particulate volume scattering function in the backward direction. *Applied optics*, **48** (35), 6811–9, ISSN 1539-4522, doi :10.1364/AO.48.006811.

Sunda W.G. et Huntsman S.A. (1997). Interrelated influence of iron, light and cell size on marine phytoplankton growth.

Sverdrup H.U. (1953). On conditions for the vernal blooming of phytoplankton. *Journal du Conseil*, **18** (3), 287–295.

Szeto M., Werdell P.J., Moore T.S. et Campbell J.W. (2011). Are the world's oceans optically different ? *Journal of Geophysical Research*, **116**, C00H04, ISSN 0148-0227, doi :10.1029/2011JC007230.

Takahashi T., Sutherland S.C., Wanninkhof R., Sweeney C., Feely R.A., Chipman D.W., Hales B., Friederich G., Chavez F., Sabine C., Watson A., Bakker D.C., Schuster U., Metzl N., Yoshikawa-Inoue H., Ishii M., Midorikawa T., Nojiri Y., Körtzinger A., Steinhoff T., Hoppema M., Olafsson J., Arnarson T.S., Tilbrook B., Johannessen T., Olsen A., Bellerby R., Wong C., Delille B., Bates N. et de Baar H.J. (2009). Climatological mean and decadal change in surface ocean pCO2, and net sea–air CO2 flux over the global oceans. *Deep Sea Research Part II : Topical Studies in Oceanography*, **56** (8-10), 554–577, ISSN 09670645, doi :10.1016/j.dsr2.2008.12.009.

Taylor J.R. et Ferrari R. (2011). Shutdown of turbulent convection as a new criterion for the onset of spring phytoplankton blooms. *Limnology and Oceanography*, **56** (6), 2293–2307, ISSN 00243590, doi :10.4319/lo.2011.56.6.2293.

Telszewski M., Chazottes A., Schuster U., Watson A., Moulin C., Bakker D., Gonzalez-Davila M., Johannessen T., Kortzinger A., Luger H., Olsen A., Omar A., Padin X., Rios A., Steinhoff T., Santana-Casiano M., Wallace D. et Wanninkhof R. (2009). Estimating the monthly pCO2 distribution in the north Atlantic using a self-organizing neural network.

The MerMex group (2011). Marine ecosystems' responses to climatic and anthropogenic forcings in the Mediterranean. *Progress in Oceanography*, **91** (2), 97–166, ISSN 00796611, doi :10.1016/j.pocean.2011.02.003.

Tomas C. (1997). *Identifying Marine Phytoplankton*, Elsevier, ISBN 978-0-12-693018-4, doi :10.1016/B978-012693018-4/50000-8.

Uitz J., Claustre H., Gentili B. et Stramski D. (2010). Phytoplankton class-specific primary production in the world's oceans : Seasonal and interannual variability from satellite observations. *Global Biogeochemical Cycles*, **24** (3), n/a–n/a, ISSN 08866236, doi :10.1029/2009GB003680.

Uitz J., Claustre H., Griffiths F.B., Ras J., Garcia N. et Sandroni V. (2009). A phytoplankton class-specific primary production model applied to the Kerguelen Islands region (Southern Ocean). *Deep Sea Research Part I : Oceanographic Research Papers*, **56** (4), 541–560, ISSN 09670637, doi :10.1016/j.dsr.2008.11.006.

Uitz J., Claustre H., Morel A. et Hooker S.B. (2006). Vertical distribution of phytoplankton communities in open ocean : An assessment based on surface chlorophyll. *Journal of Geophysical Research*, **111** (C8), C08005, ISSN 0148-0227, doi :10.1029/2005JC003207.

Uitz J., Huot Y., Bruyant F., Babin M. et Claustre H. (2008). Relating phytoplankton photophysiological properties to community structure on large scales. *Limnology and Oceanography*, **53** (2), 614–630, ISSN 00243590, doi :10.4319/lo.2008.53.2.0614.

Vidussi F., Claustre H., Bustillos-Guzmàn J., Cailliau C. et Marty J.C. (1996). Determination of chlorophylls and carotenoids of marine phytoplankton : separation of chlorophyll a from divinylchlorophyll a and zeaxanthin from lutein. *Journal of Plankton Research*, **18** (12), 2377–2382, ISSN 0142-7873, doi : 10.1093/plankt/18.12.2377.

Vidussi F., Claustre H., Manca B.B., Luchetta A. et Marty J.C. (2001). Phytoplankton pigment distribution in relation to upper thermocline circulation in the eastern Mediterranean Sea during winter. *Journal of Geophysical Research*, **106** (C9), 19939, ISSN 0148-0227, doi :10.1029/1999JC000308.

Volk T. et Hoffert M. (1985). *The Carbon Cycle and Atmospheric CO2 : Natural Variations Archean to Present*, tome 32 de *Geophysical Monograph Series*, American Geophysical Union, Washington, D. C., ISBN 0-87590-060-7, doi :10.1029/GM032.

Ward B.A., Dutkiewicz S., Jahn O. et Follows M.J. (2012). A size-structured food-web model for the global ocean. *Limnology and Oceanography*, **57** (6), 1877–1891, ISSN 00243590, doi :10.4319/lo.2012.57.6.1877.

Westberry T., Behrenfeld M.J., Siegel D.A. et Boss E. (2008). Carbon-based primary productivity modeling with vertically resolved photoacclimation. *Global Biogeochemical Cycles*, **22** (2), n/a–n/a, ISSN 08866236, doi :10.1029/2007GB003078.

White H. (1990). Connectionist nonparametric regression : Multilayer feedforward networks can learn arbitrary mappings. *Neural Networks*, **3** (5), 535–549, ISSN 08936080, doi :10.1016/0893-6080(90)90004-5.

Whitmire A.L., Letelier R.M., Villagrán V. et Ulloa O. (2009). Autonomous observations of in vivo fluorescence and particle backscatteringin an oceanic oxygen minimum zone. *Optics express*, **17** (24), 21992–2004, ISSN 1094-4087, doi :10.1364/OE.17.021992.

Wilson C. (2005). Global climatological relationships between satellite biological and physical observations and upper ocean properties. *Journal of Geophysical Research*, **110** (C10), C10001, ISSN 0148-0227, doi : 10.1029/2004JC002724.

Wong A., Keeley R., Carval T. et the Argo Data Management Team (2014). Argo quality control manual. *Technic Report*.

Xing X., Claustre H., Blain S., D'Ortenzio F., Antoine D., Ras J. et Guinet C. (2012a). Quenching correction for in vivo chlorophyll fluorescence acquired by autonomous platforms : a case study with instrumented elephant seals in the Kerguelen region (Southern Ocean). *Limnology and Oceanography : Methods*, **10**, 483–495, doi :10.4319/lom.2012.10.483.

Xing X., Claustre H., Uitz J., Mignot A., Poteau A. et Wang H. (2014). Seasonal variations of bio-optical properties and their interrelationships observed by Bio-Argo floats in the subpolar North Atlantic. *Journal of Geophysical Research : Oceans*, **119** (10), 7372–7388, ISSN 21699275, doi :10.1002/2014JC010189.

Xing X., Morel A., Claustre H., Antoine D., D'Ortenzio F., Poteau A. et Mignot A. (2011). Combined processing and mutual interpretation of radiometry and fluorimetry from autonomous profiling Bio-Argo floats : Chlorophyll a retrieval. *Journal of Geophysical Research*, **116** (C6), C06020, ISSN 0148-0227, doi : 10.1029/2010JC006899.

Xing X., Morel A., Claustre H., D'Ortenzio F. et Poteau A. (2012b). Combined processing and mutual interpretation of radiometry and fluorometry from autonomous profiling Bio-Argo floats : 2. Colored dissolved organic matter absorption retrieval. *Journal of Geophysical Research*, **117** (C4), C04022, ISSN 0148-0227, doi :10.1029/2011JC007632.

Yentsch C.S. (1960). The influence of phytoplankton pigments on the colour of sea water. *Deep Sea Research (1953)*, **7** (1), 1–9, ISSN 01466313, doi :10.1016/0146-6313(60)90002-2.

Zhang X., Hu L. et He M.X. (2009). Scattering by pure seawater : Effect of salinity. *Optics Express*, **17** (7), 5698, ISSN 1094-4087, doi :10.1364/OE.17.005698.

ANNEXE A

On the vertical distribution of the chlorophyll-a concentration in the Mediterranean Sea : a basin scale and seasonal approach

Biogeosciences, 12, 5021–5039, 2015
www.biogeosciences.net/12/5021/2015/
doi:10.5194/bg-12-5021-2015
© Author(s) 2015. CC Attribution 3.0 License.

Biogeosciences Open Access

On the vertical distribution of the chlorophyll *a* concentration in the Mediterranean Sea: a basin-scale and seasonal approach

H. Lavigne[1], F. D'Ortenzio[2,3], M. Ribera D'Alcalà[4], H. Claustre[2,3], R. Sauzède[2,3], and M. Gacic[1]

[1]Istituto Nazionale di Oceanografia e di Geofisica Sperimentale – OGS, Dip. di Oceanografia, Borgo Grotta Gigante 42/c, 34010 Sgonico (Trieste), Italy
[2]CNRS, UMR 7093, Laboratoire d'Océanographie de Villefranche, 06230 Villefranche-sur-Mer, France
[3]Université Pierre et Marie Curie, Paris 6, UMR7093, Laboratoire d'Océanographie de Villefranche, 06230 Villefranche-sur-Mer, France
[4]Laboratorio di Oceanografia Biologica, Stazione Zoologica "A. Dohrn", Villa Comunale, Napoli, Italy

Correspondence to: H. Lavigne (hlavigne@ogs.trieste.it)

Received: 16 January 2015 – Published in Biogeosciences Discuss.: 6 March 2015
Revised: 21 July 2015 – Accepted: 21 July 2015 – Published: 26 August 2015

Abstract. The distribution of the chlorophyll *a* concentration ([Chl *a*]) in the Mediterranean Sea, mainly obtained from satellite surface observations or from scattered in situ experiments, is updated by analyzing a database of fluorescence profiles converted into [Chl *a*]. The database, which includes 6790 fluorescence profiles from various origins, was processed with a specific quality control procedure. To ensure homogeneity between the different data sources, 65 % of fluorescence profiles have been intercalibrated on the basis of their concomitant satellite [Chl *a*] estimation. The climatological pattern of [Chl *a*] vertical profiles in four key sites of the Mediterranean Sea has been analyzed. Climatological results confirm previous findings over the range of existing [Chl *a*] values and throughout the principal Mediterranean trophic regimes. They also provide new insights into the seasonal variability in the shape of the vertical [Chl *a*] profile, inaccessible through remote-sensing observations. An analysis based on the recognition of the general shape of the fluorescence profile was also performed. Although the shape of [Chl *a*] vertical distribution characterized by a deep chlorophyll maximum (DCM) is ubiquitous during summer, different forms are observed during winter, thus suggesting that factors affecting the vertical distribution of the biomass are complex and highly variable. The [Chl *a*] spatial distribution in the Mediterranean Sea mimics, on smaller scales, what is observed in the global ocean. As already evidenced by analyzing satellite surface observations, midlatitude- and subtropical-like phytoplankton dynamics coexist in the Mediterranean Sea. Moreover, the Mediterranean DCM variability appears to be characterized by patterns already observed on the global scale.

1 Introduction

1.1 Surface chlorophyll distribution

Chlorophyll *a* concentration (hereafter [Chl *a*]) is the main proxy of phytoplankton biomass (Strickland, 1965; Cullen, 1982), representing a key oceanic biogeochemical variable. However, in the Mediterranean Sea, as in the global ocean, comprehensive knowledge of [Chl *a*] spatiotemporal variability has been prevented due to a lack of in situ observations (Conkright et al., 2002; Manca et al., 2004). The understanding of the [Chl *a*] distribution is essentially restricted to the surface, as it is based on remote-sensing observations. In the Mediterranean Sea, ocean color sensors, such as CZCS (Coastal Zone Color Scanner; Feldman et al., 1989) or SeaWiFS (Sea-viewing Wide Field-of-view Sensor; McClain et al., 1998), provide observations with high temporal and spatial resolution over the whole basin (Morel and André, 1991; Antoine et al., 1995; Bosc et al., 2004).

As in situ observations have demonstrated (Dolan et al., 1999, 2002; Ignatiades et al., 2009), satellite data confirm the oligotrophic nature of the basin (Dugdale and Wilkerson, 1988) as well as the east–west gradient in oligotro-

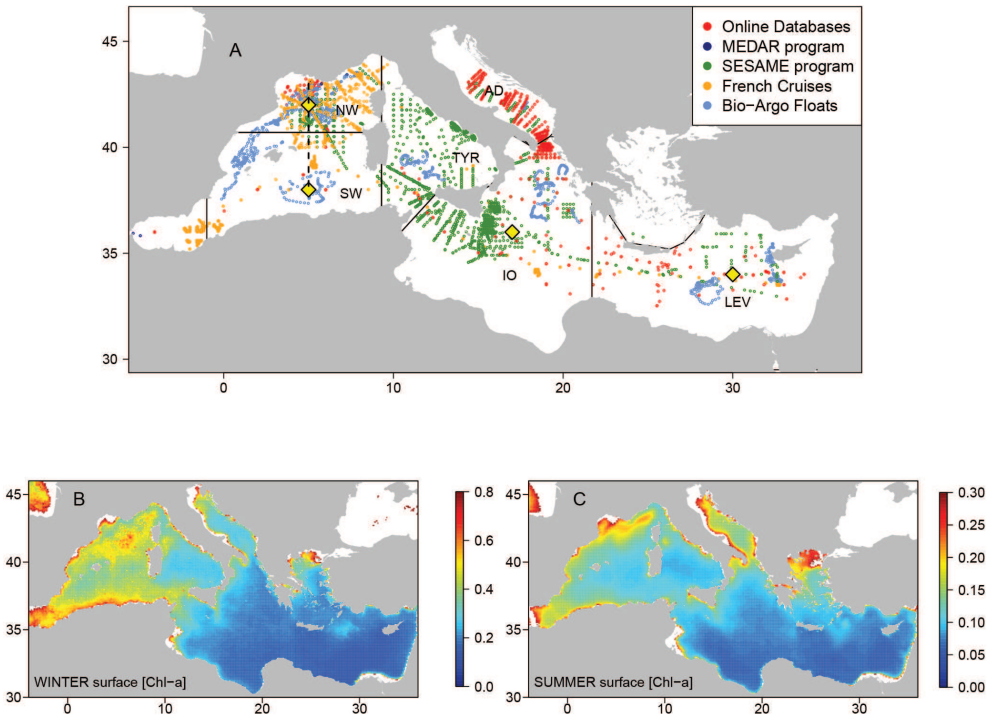


Figure 1. Panel **(a)**: spatial distribution of fluorescence profiles available in the database. Colors indicate the source of data. Black lines delineate large Mediterranean regions: they are referred to as NW for "northwest", SW for "southwest", TYR for "Tyrrhenian", AD for "Adriatic", IO for "Ionian" and LEV for "Levantine". Yellow diamonds refer to the center of region for which a climatology of [Chl *a*] vertical profile has been computed (see Fig. 3), and the dashed black line shows the center of the northwest transect (see Fig. 4). Panels **(b)** and **(c)**: SeaWiFS climatology of surface [Chl *a*] for winter **(b)** and summer **(c)**. Note that color scales are not the same.

phy (see Fig. 1b and c). Excepting the Liguro–Provençal region, where a large spring bloom takes place, and some localized spots, most of the basin exhibits very low values ($< 0.2\,\mathrm{mg\,m^{-2}}$) of satellite surface [Chl *a*]. Surface [Chl *a*] decreases eastward (Bosc et al., 2004; Barale et al., 2008), displaying a sharp gradient between the west and east basins (mean [Chl *a*] is about $0.4\,\mathrm{mg\,m^{-3}}$ in the west basin and $0.05\,\mathrm{mg\,m^{-3}}$ in the east basin; Bosc et al., 2004; Fig. 1b and c). Superimposed on this general pattern, ocean color data also provide insights into the occurrence and the influence of meso- and sub-mesoscale structures on [Chl *a*] (Taupier-Letage et al., 2003; Navarro et al., 2011, D'Ortenzio et al., 2014).

Satellite observations have also been the primary source of information for the characterization of the [Chl *a*] seasonal and interannual variability (D'Ortenzio and Ribera d'Alcalà, 2009; Volpe et al., 2012; Lavigne et al., 2013). On a global scale, ocean color satellite observations indicate that surface [Chl *a*] annual cycles display different patterns moving from a tropical to a temperate or a polar environment (Yoder et al., 1993), generally following latitudinal gradients. Boundaries between large ecological regions have been determined from satellite observations, in the global ocean (Longhurst, 2006) but also on regional scales (Devred et al., 2007; D'Ortenzio and Ribera d'Alcalà, 2009; Platt et al., 2010). Indeed, focusing on ocean color observations, D'Ortenzio and Ribera d'Alcalà (2009) confirmed the presence, in the Mediterranean Sea, of surface [Chl *a*] annual cycles, displaying similarities with subtropical or with temperate regions. The authors demonstrated that a subtropical-like [Chl *a*] seasonality (highest [Chl *a*] during winter and lowest during summer) encompasses most of the basin, whereas a temperate-like seasonality, marked by a high peak of surface [Chl *a*] in spring (in March and/or April), is recurrently observed in the northwestern basin and occasionally in other Mediterranean regions. Further analysis (Lavigne et al., 2013) showed that the coexistence of different regimes in the Mediterranean Sea is mainly due to the high variability in the interplay between physical forcing, which affects the mixed layer depth (hereafter MLD), and chemical forcing (i.e., nutrient availability).

1.2 The vertical [Chl *a*] distribution

Contrary to the horizontal distribution of [Chl *a*], which, despite the uncertainties due to the impact of bio-optical processes (see below), is regularly assessed within the basin as low cloud coverage allows high-frequency measurements, vertical distributions of [Chl *a*] are much less documented due to in situ undersampling and to the intrinsic limits of

color remote sensing in the retrieval of information from subsurface layers.

So far, the largest part of the information derives from studies conducted at specific sites (e.g., Dolan et al., 2002; Christaki et al., 2001; Estrada et al., 1993; Casotti et al., 2003; Marty et al., 2002; Psarra et al., 2000; Krom et al., 1992), generalizations based on large-scale cruises (Moutin and Raimbault, 2002; Crombet et al., 2011), and synthetic analyses (e.g., Siokou-Frangou et al., 2012) or reconstructions derived from modeling studies (e.g., Macias et al., 2014; Crise et al., 1999). These studies showed that deep chlorophyll maxima (hereafter DCMs) are ubiquitous over the Mediterranean from spring to autumn (Crise et al., 1999; Moutin and Raimbault 2002; Siokou-Frangou et al., 2010). They display a longitudinal deepening from west to east (see Crise et al., 1999 for a review), with their depth ranging from 30 m in the westernmost area (Dolan et al., 2002) to 70 m in the south Adriatic and more than 100 m in the Levantine Sea (Christaki et al., 2001). During winter, DCM generally disappear in the whole basin and the so-called "mixed" shape (Morel and Berthon, 1989; Uitz et al., 2006), characterized by a constant [Chl *a*] from the surface to the bottom of the MLD, is often observed (Krom et al., 1992; Marty et al., 2002; Mignot et al., 2014). Alternatively, a [Chl *a*] vertical shape marked by a high subsurface maximum close to the surface (less than 10 m) has also been documented for the northwestern basin, during the spring bloom period (Marty et al., 2002; Manca et al., 2004). In spite of these focused studies and the compilation of Chl *a* climatology provided by the MEDAR (Mediterranean Data Archaeology and Rescue)/MEDATLAS project (Maillard et al., 2005), the spatial distribution of [Chl *a*] vertical profiles and their yearly patterns are still poorly documented in the basin. Satellite [Chl *a*] values may provide additional information using the approach introduced for global assessments of depth-integrated Chl *a* values (e.g., Morel and Berthon, 1989). In many instances, (e.g., Bosc et al., 2004) their use was implicit and no specific analysis on the vertical distribution per se was carried out.

As discussed in a recent review by Cullen (2015), there is no unique DCM and its dynamics result from the interactions between external forcing, e.g., the penetration of light into water, the intensity of vertical mixing, and subsurface nutrient distribution and biotic processes, e.g., photoacclimation, grazing and phytoplankton composition. To assess which and how many DCMs exist in the Mediterranean Sea because of its known geographical and dynamical gradients, a starting point is to produce a quantitative characterization of the shapes of these DCMs and of their seasonal evolution, which is one of the main aims of this contribution. In addition, a good appreciation of seasonal changes in vertical [Chl *a*] distribution, the other objective of this study, is a first step towards a better understanding of mechanisms controlling seasonal phytoplankton development. It is also essential to better interpret changes in surface [Chl *a*] as detected by satellite sensors. This study will help with the biogeographical interpretation of surface [Chl *a*] patterns, paving the way for focused area studies based on in situ sampling or autonomous vehicles.

1.3 Fluorescence

In situ [Chl *a*] is obtained from filtered water samples, from which the pigment content was extracted and analyzed. The most accurate results are today obtained by high-performance liquid chromatography (HPLC, Gieskes and Kraay, 1983). Its associated protocols are most often expensive, time-consuming and depend on direct sampling with bottles. It hence provides discrete values on a vertical scale with a limited horizontal and temporal resolution. To overcome the above limitations, fluorescence observations can be used. The estimation of [Chl *a*] from the fluorescence technique (Lorenzen, 1966) is based on the chlorophyll *a* property of absorbing blue light and reemitting it, as fluorescence, in the red part of the spectrum. The quantity of fluorescence emitted by a water sample is proportional to [Chl *a*], which can then be easily derived by measuring emitted radiation at red wavelengths. The fluorescence technique therefore represents a noninvasive method to observe continuous vertical profiles of [Chl *a*]. Today, fluorimeters commonly equip CTD (conductivity–temperature–depth) sondes and can even be built into autonomous profilers. Indeed, an increasing number of profiling floats and gliders are equipped with a fluorimeter (Johnson et al., 2009), while fluorescence is becoming the main source of data for [Chl *a*] vertical profiles. To date, more than 67 900 fluorescence profiles are available in the World Ocean Database 2013 (Boyer et al., 2013).

However, fluorescence is only a proxy for [Chl *a*], meaning that the fluorescence signal needs to be calibrated for a [Chl *a*] estimation. Calibration coefficients (α and β, see Eq. 1) provided by manufacturers are only indicative of the response of the sensor to a given Chl *a* concentration in an extract or in an algal suspension and cannot be applied to all in situ conditions. The fluorescence-to-[Chl *a*] ratio is highly variable, since it changes with the taxonomic assemblage or environmental conditions (Kiefer, 1973), or it may be affected by dissolved materials (Rottgers and Koch, 2012). For instance, under low-light conditions, the chlorophyll content per cell can increase while the fluorescence-to-[Chl *a*] ratio decreases due to the packaging effect (Sosik et al., 1989). In response to supraoptimal light irradiation, phytoplankton triggers photo-protection mechanisms, inducing a drastic decrease in the fluorescence-to-[Chl *a*] ratio (Kolber and Falkowski, 1993; Müller et al., 2001); this mechanism is called non-photochemical quenching (NPQ). The main result of the NPQ effect is a decrease in fluorescence at the surface, even for constant [Chl *a*] (Cullen and Lewis, 1995; Xing et al., 2012).

$$[\text{Chl } a] = \alpha \times (\text{FLUO} - \beta) \quad (1)$$

Better estimates are obtained by determining the empirical coefficients (i.e., α and β) that fit fluorescence with in situ data for each profile (Morel and Maritorena, 2001) or for each cruise (Sharples et al., 2001; Strass, 1990; Cetinic et al., 2009). However, this calibration method based on the existence of simultaneous in situ samples is not always applicable. Alternative calibration methods, independent of concomitant HPLC observations, have therefore recently been developed (Boss et al., 2008; Xing et al., 2011; Mignot et al., 2011; Lavigne et al., 2012). They are based on additional information such as irradiance profiles (Xing et al., 2011), ocean color observations (Boss et al., 2008; Lavigne et al., 2012) or the shape of the fluorescence profile (Mignot et al., 2011). Although these new calibration methods do not reach the accuracy of HPLC-based calibration, they offer an acceptable alternative to extract reliable estimates of [Chl *a*] vertical profiles from a large quantity of fluorescence profiles.

1.4 Outlines

This study aims at improving knowledge on the spatiotemporal variability in the vertical distribution of the [Chl *a*] in the Mediterranean Sea, focusing particularly on [Chl *a*] seasonality. For this, all the available proxies of [Chl *a*] are merged to build a new database. Special attention is paid to the shape of the [Chl *a*] profiles: different patterns can point to different processes controlling the phytoplankton distribution. The spatial and seasonal variability in the DCM, which is one of the most common features in Mediterranean [Chl *a*] vertical profiles, is also specifically investigated. The objective of this paper is the description of the variability of [Chl *a*] vertical profiles, as they result from the interactions between many factors that can be complex as well as poorly documented. This variability is only discussed with regard to Mediterranean hydrology and light fields.

In the following section, the fluorescence database is presented, including the quality control and calibration procedures that were applied. In the results section, the seasonal and spatial variability in climatological [Chl *a*] vertical profiles, derived from fluorescence-based reconstructed [Chl *a*] profiles, is presented. Climatological results are completed by the analysis of the shape of the [Chl *a*] profiles. Contrary to the climatology of [Chl *a*] vertical profiles, the shape analysis is based on normalized [Chl *a*] profiles and does not account for the [Chl *a*] values. The seasonal variability in occurrences in principal [Chl *a*] vertical shapes is also investigated here. In the fourth section, certain methodological points related to the production of climatological patterns are addressed. Results are also compared with previous remote-sensing-based observations. Finally, the diversity in Mediterranean [Chl *a*] patterns is highlighted in a comparison with the global ocean.

2 Data and methods

2.1 Data set of fluorescence chlorophyll profiles

More than 6000 chlorophyll fluorescence profiles and their corresponding temperature and salinity profiles from the Mediterranean Sea in areas where bathymetry exceeds 100m depth were collected from various data sources (Table 1). These comprise online databases (986 profiles), French cruises (2670 profiles), the MEDAR (228 profiles), the SESAME (Southern European Seas: Assessing and Modelling Ecosystem changes) programs database (1815 profiles) and, finally, fluorescence profiles derived from Bio-Argo floats (1091 profiles). The profiles cover the whole Mediterranean Basin, although some areas are better represented than others (Fig. 1). Many profiles are available in the northwestern Mediterranean Sea, whereas the southwestern Mediterranean Sea and the Levantine Sea are poorly represented. Available profiles range between 1994 and 2014, all seasons being equally represented (winter 30 % of data, spring 21 %, summer 25 % and autumn 24 %). Although only 16 % of the database are Bio-Argo profiles, they represent half of the available profiles for the 2008–2014 period.

2.2 Data processing and calibration

Prior to calibration, a quality control procedure was applied to fluorescence profiles. This comprises a test of uniqueness (to eliminate repetitions of the same profile), the identification of spikes (see D'Ortenzio et al., 2010) and the identification of the signs of fluorometer failure (a portion of profile with exactly the same value or jumps in the fluorescence profile). After this quality control step, 593 profiles were removed from the database. Then, incomplete profiles (i.e., profiles for which the acquisition was not deep enough to display the whole fluorescence shape) were also removed. Profiles with a surface fluorescence value lower than the bottom value were removed from the database (202 profiles removed). In addition, the profiles obtained during the three "long-duration" stations of the BOUM (Biogeochemistry from the Oligotrophic to the Ultra-oligotrophic Mediterranean) cruise (Moutin et al., 2012) were removed from the data set because they had been sampled at a very high temporal frequency within an anticyclonic eddy (Moutin and Prieur, 2012). These 404 profiles, which are therefore not independent, would have overrepresented specific environments in the data set.

The remaining fluorescence profiles (5571 profiles) were calibrated using satellite ocean color matchups as surface references (Lavigne et al., 2012). This method has been validated in the Mediterranean Sea by comparing satellite calibrated profiles and in situ HPLC [Chl *a*] data. In the Mediterranean Sea, the calibrated profiles are unbiased and present a median error of 41 %, which is reduced to 34 % when compared to climatological averages. In summary (see Lavigne

Table 1. Description of sources for fluorescence profiles. In this table, only fluorescence profiles obtained in Mediterranean regions where bathymetry is greater than 100 m are counted. Coastal regions have been neglected.

	Data source	Number of profiles
Online databases	PANGAEA (http://www.pangaea.de/)	93
	SISMER (http://www.ifremer.fr/sismer/index_FR.htm)	110
	WOD09 (http://www.nodc.noaa.gov/)	94
	OGS database (http://nodc.ogs.trieste.it/cocoon/data/dataset)	689
	SUB-TOTAL	986
French cruises	PROSOPE (Claustre et al., 2004)	96
	DYNAPROC (Andersen and Prieur, 2000)	251
	BOUM (Moutin et al., 2012)	573
	ALMOFRONT (Claustre at al., 2000)	1046
	DYFAMED (Marty et al., 2002)	191
	MOOSE-GE (http://hermes.dt.insu.cnrs.fr/moose)	285
	DEWEX (Durrieu de Madron et al., 2011)	228
	SUB-TOTAL	2670
SESAME Program (http://www.sesame-ip.eu/)		1815
MEDAR Program (MEDAR Group., 2002)		228
Bio-Argo (Xing et al.. 2011; http://www.oao.obs-vlfr.fr/web/index.php)		1091
TOTAL		6790

et al., 2012, for a comprehensive description and validation of the procedure), the method consists of (step 1) a correction for the NPQ effect, (step 2) the adjustment to a zero value of the fluorescence profile at depth and (step 3) the application of a calibration coefficient obtained from ocean color satellite matchups. The last step has only been applied to the fluorescence profiles available for the 1998–2014 period (i.e., the time during which SeaWiFS or MODIS (Moderate Resolution Imaging Spectroradiometer) Aqua data were available and could be used to calculate the matchups).

Step 1 provides a systematic correction of the NPQ effect by extrapolating the maximum fluorescence value observed in the mixed layer up to the surface (Xing et al., 2012). Although Biermann et al. (2014) proposed an improvement of the method for profiles with a euphotic depth above the MLD, we preferred to use a unique data processing procedure to avoid the introduction of an artificial bias due to a heterogenic data treatment. The MLD was evaluated from potential density profiles using a density criterion of $0.03\,kg\,m^{-3}$ (de Boyer Montegut et al., 2004; D'Ortenzio et al., 2005). This method proved to be an efficient NPQ correction in most conditions (Xing et al., 2012; Lavigne et al., 2012), although it presented limitations for shallow MLDs and stratified water columns. By applying the equation proposed by Sackmann et al. (2008) on monthly averaged light fields, the impact of NPQ was observed to be significant only above 60 m, thus leading to a 2-fold underestimation of surface [Chl *a*]. Considering this result, the weak efficiency of the NPQ correction method in stratified conditions should not have major consequences for the present study. Only the analysis of the surface-to-integrated-content chlorophyll ratio (see Table 3) should be considered with caution.

Step 2 corrects the systematic instrumental offset, which impacts the whole profile, although it can only be detected at depth. Except for very specific cases, [Chl *a*] is considered to reach a zero value at depths where there is no more light availability. If this is not the case, a correction factor (i.e., β in Eq. 1) is subtracted from the whole fluorescence profile, considering that the median of the 10 deepest observations is equal to 0. Profiles in which the MLD was deeper than the deepest fluorescence observation were not processed but were not removed from the database either (1.1 % of data set). After step 1 and step 2, 5571 profiles were successfully corrected and stored in the so-called "1994–2014 database". These fluorescence profiles were used later for the shape analysis (see Sect. 2.3 and Sect. 3.2).

In step 3, fluorescence profiles collected after 1998 were converted into [Chl *a*] units using a transformation based on ocean color satellite observations (Lavigne et al., 2012). Eight-day Level 3 standard mapped images of SeaWiFS and MODIS Aqua surface chlorophyll at a 9 km resolution were obtained from the NASA web site (http://oceancolor.gsfc.nasa.gov/) for the 1998–2014 period (1998–2007 for SeaWiFS and 2008–2014 for MODIS Aqua). The use of NASA [Chl *a*] standard products allows for a good consistency between SeaWiFS and MODIS data sets, thus avoiding the introduction of any bias between the two time series (Franz et al., 2005). For each fluorescence profile, the satellite image matching the profile date was selected. The corresponding surface [Chl *a*] values over a $0.1° \times 0.1°$ box centered on the

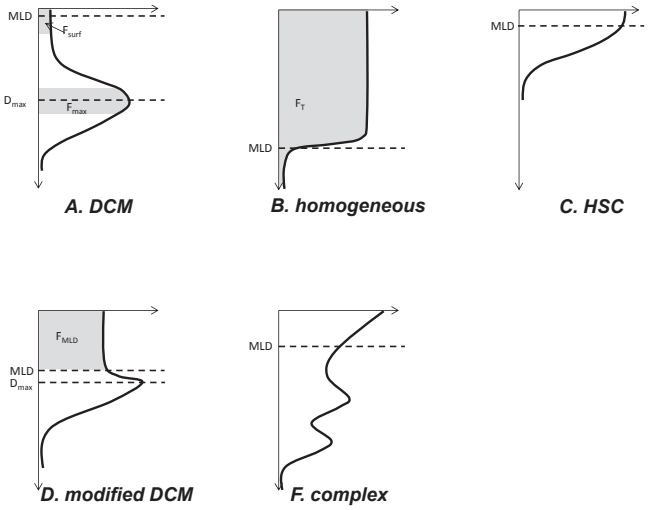


Figure 2. The five standard shapes for [Chl *a*] vertical profiles identified in our data set. See Sect. 2.3 of the text for more details about these shapes and for a description of the algorithm used to identify them. Black solid lines represent the normalized [Chl *a*] vertical profile. Metrics used for the determination of the profile standard shape (i.e., MLD, D_{max}, F_{surf}, F_{max}, F_T; see Sect. 2.3 for definitions) are represented on standard profiles. Although all of these metrics have been computed on each fluorescence profile, they could not be represented on the same profile for practical reasons.

geographical position of the profile were extracted and averaged. The integrated chlorophyll content over $1.5Z_e$ (where Z_e is the euphotic depth) is then estimated from satellite [Chl *a*] using empirical relationships (Uitz et al., 2006) and Z_e is calculated from the chlorophyll-integrated content using the equations of Morel and Berthon (1989). A multiplicative coefficient (α coefficient in Eq. 1) is applied to the fluorescence profile, imposing the condition that the integrated fluorescence content matches the integrated chlorophyll content derived from satellite. In the end, 3867 fluorescence profiles were successfully transformed into [Chl *a*]. These [Chl *a*] profiles formed the "1998–2014 database" and, like the fluorescence profiles of the "1994–2014 database", they are available upon request from the first author.

2.3 Determination of the shape of fluorescence profiles

On the basis of a visual analysis of the whole database, five general types of fluorescence vertical shapes were identified. These five categories, which represent the most frequent shapes of vertical distribution observed in the Mediterranean, also reflect their conditioning by physical–biological processes. These categories are referred to as *DCM*, *homogeneous*, *HSC* (for high surface chlorophyll), *complex* and *modified DCM* on the basis of their general characteristics (Fig. 2). The *DCM* and *homogeneous* shapes have been commonly used to describe [Chl *a*] vertical profiles (Morel and Berthon, 1989; Uitz et al., 2006; Mignot et al., 2011). They are referred to as "stratified" and "mixed", respectively, and are distinguished according to the relative position of Ze and the MLD. The *DCM* shape is characterized by a subsurface DCM, and the *homogeneous* shape by a positive homogeneous [Chl *a*] in the mixed layer. After an examination of the database, three other standard shapes have been introduced (i.e., *HSC*, *modified DCM* and *complex* shapes) to better describe the observed variability. The *HSC* standard shape was defined for profiles displaying a steady decrease in [Chl *a*] from surface to depth ($\sim$ 100 m), as generally observed during phytoplankton blooms (Chiswell, 2011). The *modified DCM* shape describes profiles with relatively high values in the mixed layer and with a peak of [Chl *a*] just below the MLD. It represents an intermediate condition between the *DCM* and *homogeneous* situations. Finally, profiles with a complex shape, often displaying several peaks and a relatively high surface [Chl *a*], were classed as standard *complex* shapes.

To automatically categorize each profile of the 1994–2014 database into one of the five shape classes, a simple algorithm has been used, computing the following metrics for each profile: the depth of fluorescence maxima (D_{max}, see Fig. 2a and d), the MLD, the fluorescence-integrated content in a 20m layer centered on D_{max} (F_{max}, see Fig. 2a), the fluorescence-integrated content in the 0–20 m surface layer (F_{surf}, see Fig. 2a), the fluorescence-integrated content in the mixed layer (F_{MLD}, see Fig. 2d) and the total fluorescence content (F_T, see Fig. 2b).

The algorithm was applied to each profile. It first tests the *HSC* shape. The *HSC* shape is assigned to a profile if its fluorescence averaged over layers of a width of 10 m decreases from surface to 100 m. Secondly, the *DCM* shape is tested. If the MLD is above D_{max} and if F_{max} is twice as large as F_{surf}, the profile is classed in the *DCM* category. If not, the *homogeneous* shape is tested. The profile is classed in the *homogeneous* category if F_{MLD} / F_T is superior to 0.85 (more than 85 % of biomass contained in the mixed layer). Finally, if the fluorescence profile does not meet any of the previous criteria, it is either classed in the *modified DCM* category, if the corresponding MLD is above D_{max}, or in the *complex* category.

Overall, 2780 profiles were classed in the *DCM* category, 751 in the *homogeneous* category, 413 in the *HSC* category, 637 in the *modified DCM* category and 990 in the *complex* category.

3 Results

3.1 Some climatological behaviors

Although the availability of the calibrated profiles (1998–2014 database) should allow us to generate interpolated products on a regular mesh grid (as, for example, the World Ocean Atlas; Conkright et al., 2002), we preferred to avoid any large

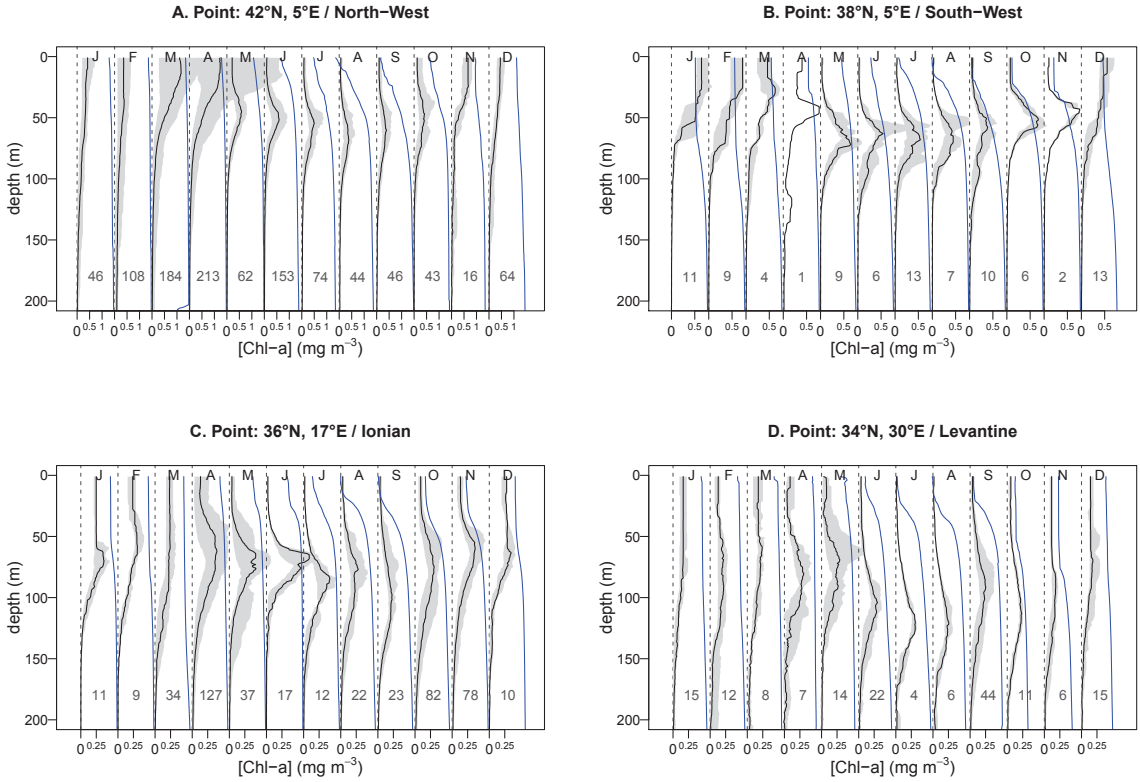


Figure 3. Climatology of [Chl *a*] vertical profiles (black lines) for four points of the Mediterranean Sea (see yellow diamonds in Fig. 1). All profiles located within a 4° × 4° box centered on indicated positions were retained. The median value for each month is the black line. The grey zone indicates the 0.1-quantile–0.9-quantile range. Numbers below climatological profiles indicate the number of available data profiles used to compute them. Normalized average water density profiles are superimposed (blue lines).

interpolation and only present Mediterranean patterns for locations well represented in our database. Hence, monthly climatologies of [Chl *a*] vertical profiles were computed for four geographical areas (i.e., 4° × 4° boxes) where the data density was high. These locations were also placed in four main Mediterranean subbasins (i.e., centered on 42° N, 5° E in the northwestern basin, 38° N, 5° E in the southwestern basin, 36° N, 17° E in the Ionian Sea and 34° N, 30° E in the Levantine Sea; see yellow diamonds in Fig. 1). The monthly time series are presented in the next section (Sect. 3.1.1). Although, in the following, we refer to these time series as "climatological", certain average profiles result from a low number of fluorescence profiles (sometimes less than 10; see numbers in Fig. 3) and therefore do not strictly represent a climatological pattern. To better identify spatial changes in [Chl *a*] fields, we also present climatological transects (Sect. 3.1.2). Due to the weak density of data in the eastern basin, the [Chl *a*] distribution could only be analyzed along a 5° E north–south transect in the western basin (see dotted line in Fig. 1). Nevertheless, this transect encompasses regions with different biological dynamics (D'Ortenzio and Ribera d'Alcalà, 2009), and it is representative of the main patterns of the western Mediterranean.

3.1.1 Seasonality in four geographic locations

For each of the four selected geographic locations (see above), all available profiles in a 4° × 4° side box centered on the chosen geographical position were averaged on a 1 m vertical scale and on a monthly basis to produce climatological profiles. The resulting monthly climatologies are displayed in Fig. 3.

Overall, the climatological time series representing the southwestern basin, the Ionian Sea and the Levantine Sea (Fig. 3b, c and d) display a similar evolution of the vertical [Chl *a*] distribution. From December to March, [Chl *a*] is greater in the surface layer, i.e., from the surface to the base of pycnocline (Fig. 2b), while the April to November period is characterized by the occurrence of a DCM, concurrent with the development of the seasonal pycnocline close to surface. In the southwestern region, winter profiles present relatively high [Chl *a*] in the upper meters ([Chl *a*] $> 0.5\,\mathrm{mg\,m^{-3}}$), whereas in the Ionian, and even more in the Levantine, upper layer [Chl *a*] is lower and the base of the pycnocline is deeper (about 150 m in the Ionian Sea and more than 200 m in the Levantine Sea). DCM, when occurring, is deeper in the Levantine and Ionian seas than in the southwestern region. The climatological time series in the north-

western basin (Fig. 3a) displays a different succession. DCM occurs from May to October, when surface stratification of the water column can be observed. In November and December, [Chl *a*] vertical profiles display homogeneous concentrations from the surface to the upper limit of the pycnocline, which deepens through mixing processes. In January and February, the water density profiles are nearly constant and [Chl *a*] profiles display low and homogeneous concentrations up to 100 m. In March and April, although surface water density slightly decreases, pointing to water column stabilization and/or stratification, surface [Chl *a*] considerably increases. Finally, all time series are characterized by a deepening of the DCM from May to July and a shallowing from August to September. It appears that in the northwest region, the deepening of the DCM coincides with the deepening of the pycnocline. In the other areas, the pycnocline is much shallower than the DCM and their dynamics seem to be uncoupled until September. In October and November, the base of the surface mixed layer seems to be correlated with DCM.

Regarding [Chl *a*] values, regional differences are visible, confirming previous observations on the eastward increase in oligotrophic conditions. The highest [Chl *a*] value is observed in April, in the northwestern climatology (Fig. 3a), reaching 1.2 mg m^{-3}. However, this mean value is derived from extremely variable observations ranging between 0.3 and 4.2 mg m^{-3}. The southwestern time series shows [Chl *a*] values up to 0.5 mg m^{-3}, observed at the surface during winter and at the DCM during summer. In the Ionian climatology, the highest [Chl *a*] values can be observed at the DCM, reaching 0.3 mg m^{-3}. Finally, the Levantine climatology displays the lowest [Chl *a*], with values rarely exceeding 0.25 mg m^{-3}.

Table 2 presents median [Chl *a*] values at the DCM depth, for the four geographic locations analyzed here. Contrary to the DCM [Chl *a*] values seen in Fig. 3, the values reported in Table 2 are derived from the median DCM [Chl *a*] values extracted individually from each fluorescence profile presenting a DCM. In the northwestern region, [Chl *a*] at DCM is often around 1 mg m^{-3}, though it ranges between 0.63 in September and 1.07 mg m^{-3} in April. At the southwestern point, the averaged [Chl *a*] at DCM is 0.88 mg m^{-3}. In the eastern basin, values are twice lower (about 0.55 mg m^{-3} at the Ionian point and 0.40 mg m^{-3} at the Levantine point). A seasonal pattern does not clearly emerge from the analysis of the DCM statistics, except that [Chl *a*] at DCM is generally higher during spring and summer and lower during autumn. Note that median DCM depth [Chl *a*] values (Table 2) are higher than the DCM depth [Chl *a*] values observed on climatological profiles (Fig. 3) because the averaging process on the latter tends to flatten DCMs (see discussion on Sect. 4.1.2, Lavigne et al., 2012).

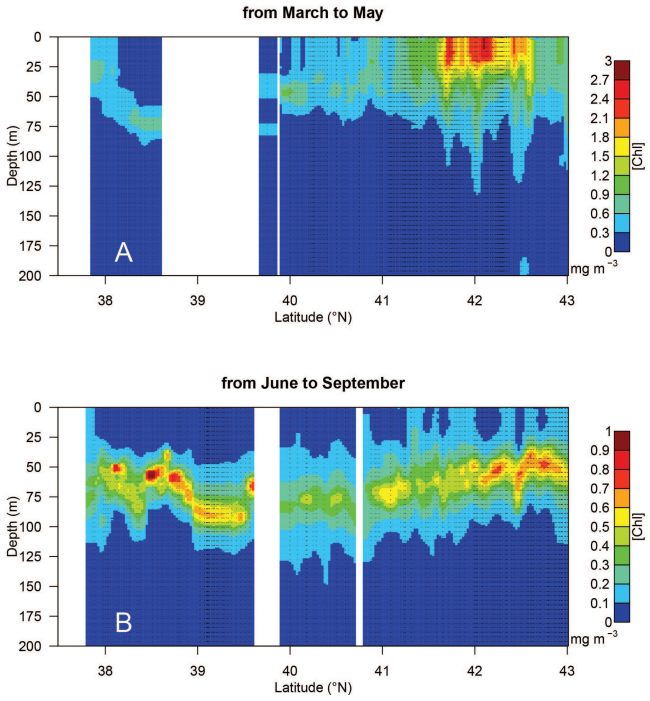


Figure 4. North–south climatological transect of [Chl *a*] along the 5° W meridian (see the black dotted line in Fig. 1). Panel **(a)** represents the averaged situation for the March to May period, and **(b)** does so for the June to September period. Note that color scales are different in **(a)** and **(b)**. For each available data profile, a vertical dotted line was superimposed on the graph.

3.1.2 North–south transect

All the data located within ±2° from the 5° E meridian were selected to produce a climatological picture of [Chl *a*] fields in spring (March–May; Fig. 4a) and in summer (June–September; Fig. 4b).

The spring situation (Fig. 4a) displays various types of profiles and a large range of [Chl *a*] values. North of 41° N, [Chl *a*] values are high (> 1 mg m^{-3}) at the surface and decrease with depth. The highest [Chl *a*] values (∼ 3 mg m^{-3}) are observed around 42° N at the surface (up to 30 m depth). Between 40 and 41° N, surface [Chl *a*] is around 0.5 mg m^{-3} and a DCM is visible at 50m depth. Further south, the climatological transect displays a deeper DCM (around 75m depth) and very low surface [Chl *a*] values (< 0.3 mg m^{-3}).

In the summer transect (Fig. 4b), the presence of a DCM is ubiquitous, although its position in the water column and its [Chl *a*] values vary throughout the transect. A steady deepening of the DCM is observed from 43° N (DCM depth around 50 m) to 39° N (DCM depth around 85 m). A southward decrease in [Chl *a*] at DCM is also observed. It ranges from 0.8 to 0.4 mg m^{-3}. South of 39° N, a shallowing of the DCM depth and an increase in the [Chl *a*] at DCM are observed.

Table 2. Median and interquartile range (IQR) of [Chl *a*] at DCM for each geographical location analyzed in Fig. 3 (i.e., yellow diamonds in Fig. 1). Median and IQR were computed by considering all the DCM depth [Chl *a*] estimations extracted from available profiles of the *DCM* standard shape. IQR is the difference between the third and the first quartile.

	Point: 42° N, 5° E			Point: 38° N, 5° E			Point: 36° N, 17° E			Point: 33.5° N, 33° E		
	(Northwest)			(Southwest)			(Ionian)			(Levantine)		
	MEDIAN	IQR	N	MEDIAN	IQR	N	MEDIAN	IQR	N	MEDIAN	IQR	N
Apr	1.07	0.48	26				0.70	0.26	107	0.47	0.09	6
May	0.83	0.33	38	0.97	0.23	9	0.71	0.25	37	0.49	0.08	6
Jun	0.97	0.36	129	1.08	1.26	6	0.81	0.28	17	0.37	0.25	154
Jul	0.97	0.67	67	0.84	0.20	160	0.42	0.23	9	0.42	0.10	10
Aug	0.57	0.39	45	0.73	0.36	7	0.41	0.15	22	0.32	0.08	11
Sep	0.63	0.21	41	0.62	0.16	9	0.32	0.17	23	0.32	0.06	23
Oct	079	0.32	33	1.06	0.11	6	0.43	0.13	81	0.32	0.03	10

Table 3. Median value (bold) and first and ninth decile for the following parameters: MLD, euphotic depth (Z_e), surface [Chl *a*] observed by satellite (Chl_{SAT}) and percentage of chlorophyll content in the upper 20 m layer compared to the total integrated content (F_{surf}/F_T).

	MLD (m)	Z_e (m)	Chl_{SAT} (mg m^{-3})	F_{surf} / F_T (%)
DCM	**14**	**72**	**0.13**	**5 %**
	11–27	57–90	0.05–0.27	2–11
Modified DCM	**26**	**52**	**0.32**	**21 %**
	13–52	37–66	0.16–0.63	13–32
Homogeneous	**86**	**52**	**0.31**	**21 %**
	27–596	29–71	0.13–1.19	11–43
Complex	**33**	**48**	**0.36**	**28 %**
	17–63	33–62	0.18–0.80	18–47
HSC	**35**	**34**	**0.77**	**31 %**
	13–95	17–57	0.25–2.76	20–53

3.2 Analysis of the profile shapes

3.2.1 Characteristics of standard shapes

As a procedure was established to classify the shapes of the [Chl *a*] profiles included in the 1994–2014 database (Sect. 2.3), certain characteristics related to [Chl *a*] profiles could be computed. They are summarized in Table 3.

The MLD is shallowest when the standard vertical fluorescence shape is *DCM*. Additionally, the MLD is deepest when the standard florescence shape is *homogeneous*. In these two cases, the relative position of the MLD and Z_e confirm, therefore, that the *homogeneous* and *DCM* shapes can be compared with the well-known stratified and mixed shapes introduced by Morel and Berthon (1989). Profiles shapes categorized as *modified DCM*, *complex* and *HSC* display intermediate values for the MLD. For profiles of the *modified DCM* shape, the average distance between the MLD and chlorophyll maxima is 22 m. This relatively short distance may indicate that the *modified DCM* shape derives from erosion by deeper mixing of the DCM structure. For the *HSC* standard shape, the MLD can be relatively deep (ranging between 13 and 95 m). A [Chl *a*] gradient could therefore develop in both stratified and mixed conditions. According to Huisman et al. (1999), the development of a [Chl *a*] gradient in the mixed layer would be possible if mixed layer turbulence were low, thus allowing for the accumulation of phytoplankton cells near the surface.

According to the results presented in Table 3, ocean color surface [Chl *a*] values are related to the shape of the vertical profile. The lowest surface [Chl *a*] values are observed for *DCM* shape profiles, while the highest (0.77 mg m^{-3}) values are observed for *HSC* shape profiles. In spite of its variability, this high value suggests that the *HSC* shape could result from the exponential growth of phytoplankton at the surface in unlimited-nutrient conditions associated with a stable water column. Hence, *HSC* profiles would typically correspond to bloom conditions. A very high variability, with surface [Chl *a*] values ranging from 0.13 to 1.19 mg m^{-3}, is observed for profiles of the standard *homogeneous* shape. This variability likely results from the interactions between the high variability in the MLD and the recent development of phytoplankton biomass.

The F_{surf} / F_T ratio changes with the shape of the [Chl *a*] profile. The lowest ratio (5 %) is observed for the *DCM* shape, even though this value is likely to be underestimated by a factor of 2.5 because of NPQ. The standard *complex* and *HSC* shapes display similar median ratios: 28 and 31 %, respectively. Once again, there is a large variability among *homogeneous* shape profiles that which can be explained by the variability in the MLD. Finally, in the *HSC* situation, the upper 20 m can accumulate up to 50 % of the chlorophyll content.

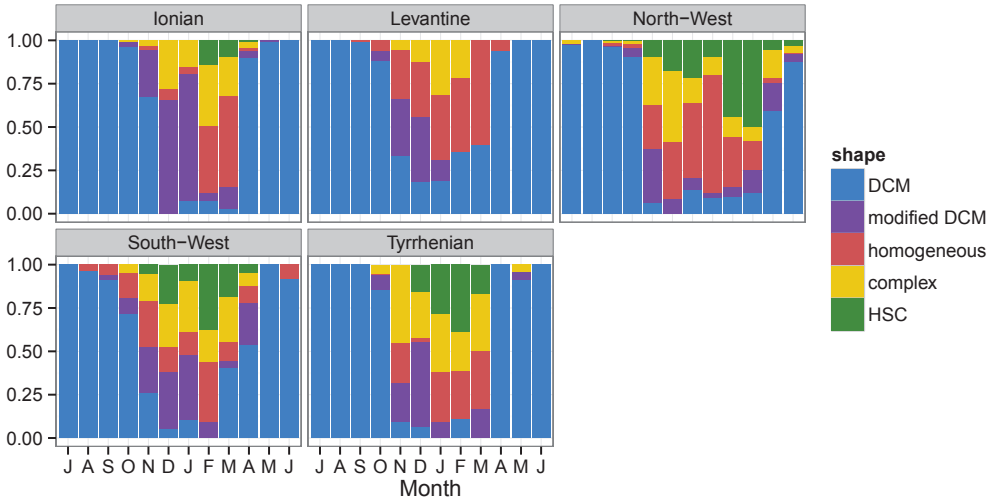


Figure 5. Histograms indicating for each month and each Mediterranean region the proportion of each type of standard shape observed in the 1994–2014 database (i.e., *DCM*, *homogeneous*, *HSC*, *modified DCM* and *complex*; see Fig. 2 and Sect. 2.3). The height of color bars indicates the proportion of profiles which were classed in each category of standard shapes. Note that months range from July to June.

3.2.2 Seasonal distribution of the profile shapes

A study of the seasonal distributions of standard shapes was performed for the main Mediterranean regions (Fig. 5, boundaries of the Mediterranean regions are drawn in Fig. 1). During summer, all the regions are dominated by the *DCM* shape, with occurrences exceeding 90 %. The *DCM* shape disappears everywhere in November. The time of its onset depends on the region: it occurs in April in the Ionian, Levantine and Tyrrhenian regions, in May in the southwest region, and in June in the northwest region. During the autumn and winter period, all the categories of shapes can be observed in the same region and during the same month. Nevertheless, profiles shapes classed as *modified DCM* are more frequent in early winter (i.e., in the Ionian region where the *modified DCM* shape represents more than 60 % of profiles in December and January), which reinforces the impression that this shape may be generated by deeper mixing eroding the DCM structure. Profiles with the *homogeneous* shape are observed everywhere from November to March, except in the Ionian region. Similarly, the *complex* shape is present everywhere from November to March. Profiles displaying an *HSC* shape are absent, or nearly absent, in the Ionian and Levantine regions. In the Tyrrhenian and southwest regions, *HSC* profiles can be observed between November and March and are most abundant in February. In the northwest region, although *HSC* profiles are observed in winter from November to February, they peak in spring (March–April), with occurrences exceeding 60 %. Assuming that the *HSC* profiles denote bloom events, this result suggests that bloom events may occur during winter in the whole western Mediterranean although they only peak in the northwest region during spring.

Table 4. Regional average values and standard deviations (numbers in brackets) for a set of parameters. The winter MLD was computed with January and February MLDs. The DCM depth and the PAR (photosynthetically active radiation) at DCM has been computed only for profiles belonging to the *DCM* standard shape category. PAR at DCM has been determined for each fully calibrated (i.e., 1998–2014 database) [Chl *a*] vertical profiles. The vertical profile of the PAR attenuation coefficient was computed from [Chl *a*] vertical profiles and applied to surface PAR estimates derived from the monthly SeaWiFS PAR climatology. For the nitracline depth, the isoline 1 µM was computed on a large set of nitrates profiles derived from MEDAR and SESAME programs (see Lavigne et al., 2013, for details about their database).

	Winter MLD (m)	Nitracline depth (m)	DCM depth (m)	PAR at DCM (mol photons m^{-2} day^{-1})
Northwest	342 (623)	62 (38)	51.7 (12.5)	1.03 (0.86)
Southwest	47 (63)	78 (24)	73 (17)	0.77 (0.77)
Tyrrhenian	45 (38)	97 (23)	73 (13)	0.57 (0.19)
Adriatic	126 (181)	56 (24)	56 (10)	–
Ionian	67 (46)	119 (46)	83 (29)	0.51 (0.64)
Levantine	122 (122)	185 (47)	102 (17)	0.16 (0.16)

3.2.3 Longitudinal and seasonal distribution of the DCM depth

The DCM is confirmed to be a dominant feature of the [Chl *a*] distribution in the Mediterranean, although its characteristics change from one region to another and with time. A deepening of the DCM depth with longitude is generally observed (Fig. 6), confirming previous findings (Crise et al., 1999). A linear model applied to DCM depth data indicates that, on average, DCM depth deepens by 1.6 m for 1° of longitude. However, large variability exists, especially in the Io-

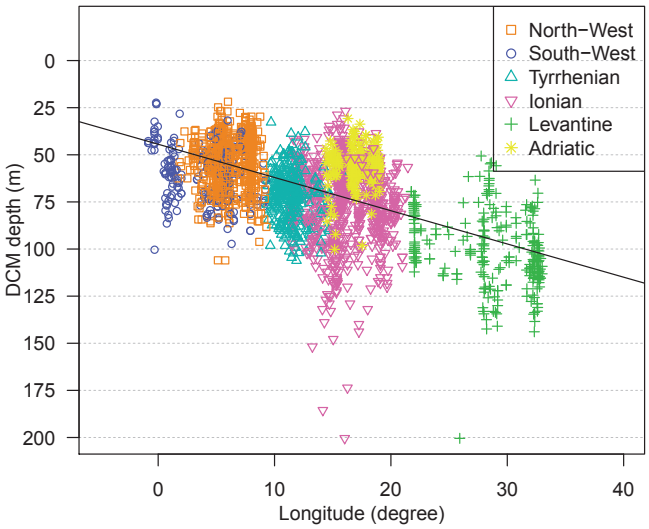


Figure 6. DCM depth as a function of longitude. DCM depths were computed only on profiles of the *DCM* standard shape (see Sect. 2.3 for an objective definition of *DCM* shape). Black line represents the linear model between the DCM depth and the longitude. Its slope is 1.6 m per degree of longitude.

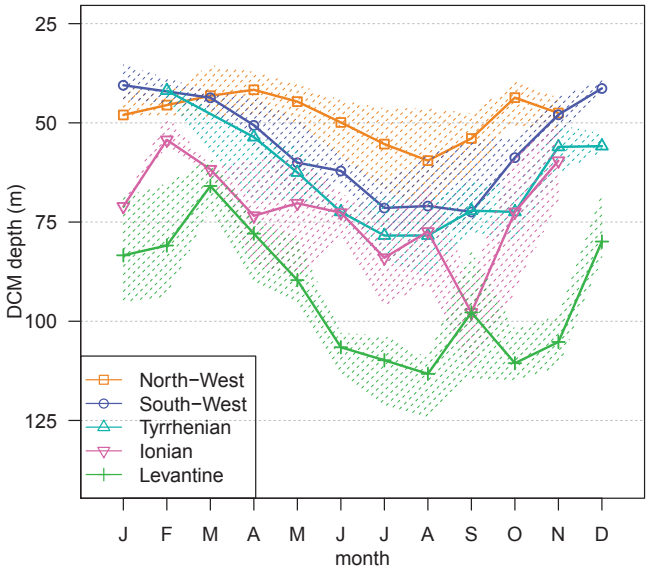


Figure 7. Seasonal evolution of the DCM depth. DCM depths were computed only on profiles of the *DCM* standard shape (see Sect. 2.3 for an objective definition of *DCM* shape). Symbols refer to monthly median, whereas dotted areas indicate the interquartile range.

nian and Levantine seas. Superimposed on this general deepening of DCM with longitude, regional differences can be observed between the main Mediterranean subbasins. Considering profiles at the same range of longitude, the averaged DCM depth is deeper and more variable in the southwest region than in the northwest region (see Table 4). In the eastern basin, the Adriatic region displays shallow and stable DCM depths, whereas the Ionian and Levantine regions display deeper and more variable DCM depths (Table 4).

Part of the variability observed in the different Mediterranean regions can be explained by seasonality. All the Mediterranean regions have a seasonal variability in the DCM depth (Fig. 7), which is characterized by a widespread deepening from March to midsummer and a shallowing from midsummer to November. In all the Mediterranean regions, except the northwest region, there is a 40 % deepening of the DCM between spring and summer (33 % in the northwest).

4 Discussion

4.1 Methodological discussion

4.1.1 Comparison with MEDATLAS

The climatological profiles for each of the four geographical points analyzed in Sect. 3.1 have been computed from the MEDATLAS climatology and compared to their fluorescence-based counterparts evaluated here (Fig. 8). For each geographical point, the two versions of [Chl *a*] vertical profiles (fluorescence-based and MEDATLAS) displayed similar ranges of values, although differences are observed in the form of [Chl *a*] vertical profiles. The fluorescence-based profiles often display thinner DCMs with higher [Chl *a*] values than in the MEDATLAS climatology (see, for instance, Fig. 8b for summer, c for autumn and d for summer). Moreover, in the MEDATLAS climatology, very weak seasonal changes in the DCM depth are visible. These divergences can be explained by the use of discrete data and of interpolation in the MEDATLAS climatology, which prevents the proper characterization of vertical structures. In winter, the MEDATLAS climatology, and sometimes the fluorescence-based climatology, shows profiles with subsurface maxima (Fig. 8a, b, c, winter), which have not been observed in the monthly fluorescence-based time series (Fig. 3). We hypothesize that these winter subsurface maxima could be an artifact caused by the large averaging timescale (from December to March), leading to the combination of [Chl *a*] profiles with highly different vertical distributions (see Fig. 5). Another particular feature of the MEDATLAS climatology that does not show in the fluorescence-based climatology is represented by the rises in summer and autumn surface [Chl *a*] above DCM (Fig. 8a, b and d). We suggest that this feature could result from the propagation by interpolation of the high surface [Chl *a*] observed on coastal regions (see also Bosc et al., 2004). In addition, considering the geographical positions of the available MEDAR observations, in almost all the studied subbasins (except Ionian), coastal observations are included in the database. They might therefore be responsible for the observed difference with the fluorescence-based climatology.

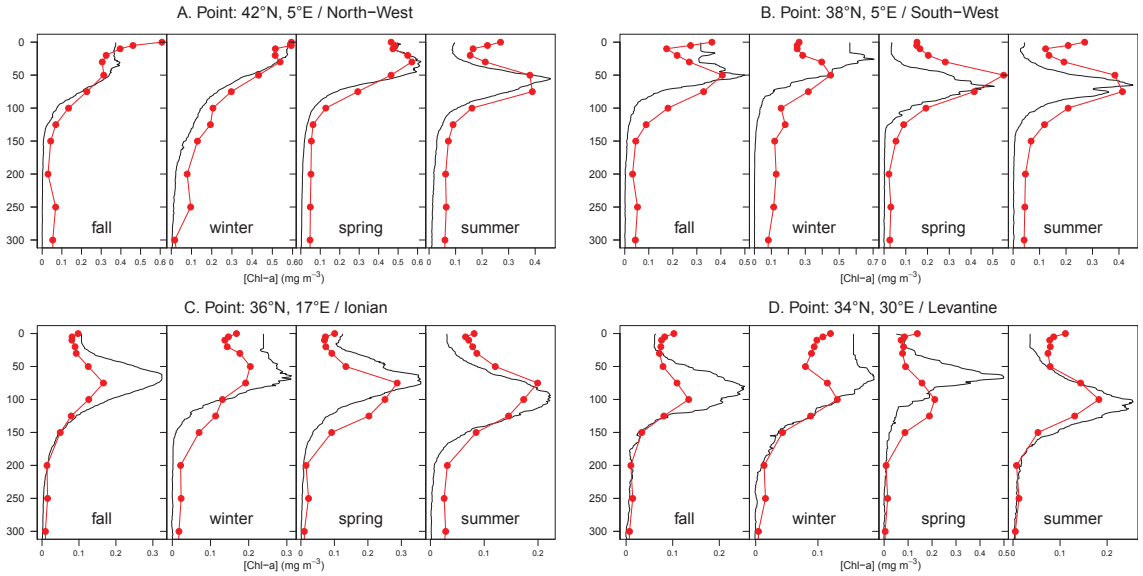


Figure 8. [Chl *a*] profiles obtained from the MEDATLAS climatology for the four locations analyzed in Fig. 3 (red lines and red points). MEDATLAS climatology was downloaded from http://modb.oce.ulg.ac.be/backup/medar/medar_med.html. For comparison, corresponding seasonally averaged profiles were computed from the 1998–2014 [Chl *a*] fluorescence database (black lines). Seasons are calendar-based seasons.

In summary, the results of this comparison demonstrate that, although the MEDATLAS database is extremely valuable, the derived MEDATLAS fields for [Chl *a*] present serious limitations, and they need to be updated.

4.1.2 Methodological approaches

In the present study, two different approaches have been used to describe the monthly variability in [Chl *a*] profiles. On one hand, the standard method consists of averaging [Chl *a*] values for a number of defined standard depths (i.e., Conkright et al., 2002; Sect. 3.1). On the other hand, a "probabilistic" method (Sect. 3.2), for which each [Chl *a*] profile is considered as a whole, focuses the analysis on each profile's general shape and on specific features (e.g., DCM depth). The second approach requires an a priori knowledge of the different profile shapes found in the database as well as the definition of an efficient and automatic procedure to categorize the profiles. In this analysis, the main standard shapes and the classification procedure were defined after individual visualization of all the fluorescence profiles in the database and the determination of their characteristics (i.e., D_{max}, F_{MLD} / F_T, F_{max} / F_{surf}; see Sect. 2.3 for details).

These two approaches are complementary. The standard method highlights the average pattern of the [Chl *a*] profile and provides the ranges of [Chl *a*] values. However, [Chl *a*] values must be considered independently for each depth, and the shape of the resulting climatological profile has to be interpreted carefully because it is a composite. A typical artifact of this method is the tendency of the DCM to be flattened (compare the DCM of Fig. 3 and the values in Table 2). In these cases (i.e., [Chl *a*] profile extremely stable, as during summer, or very dynamic, as during winter), the probabilistic analysis of the shape of the [Chl *a*] profile appears more pertinent. In addition, the probabilistic analysis provides information on the environmental processes that lead to the observed [Chl *a*] shape. As mentioned in Sect. 3.2.1, the *modified DCM* shape likely results from the erosion by upper vertical mixing of the DCM structure, while the *homogeneous* standard shape is likely driven by vertical mixing, which encompasses the whole [Chl *a*] profile. Similarly, the *HSC* profiles, associated with high surface [Chl *a*] values (see Table 3), could be collected (and then associated with) during surface phytoplankton bloom conditions. Under these conditions, if there is no nutrient limitation, growth rate is essentially affected by light availability and then decreases with depth. This can account for the derived decrease in the [Chl *a*] gradient from surface to depth. Nevertheless, these conjectures have to be considered on a statistical basis. Indeed, each individual profile is affected by complex and variable factors (i.e., vertical mixing, 3D dynamic structures, light distribution, grazing pressure; see Longhurst and Harrison, 1989, and see also discussion below), which sometimes lead to erratic [Chl *a*] vertical distributions that become difficult to explain (17 % of profiles have been classed as *complex* standard shapes). Finally, the probabilistic analysis also revealed that seasonal changes in [Chl *a*] profiles are not smooth and steady, as the climatological analysis may suggest, but rather are extremely variable.

4.2 A new vision of the [Chl *a*] in the Mediterranean Sea

4.2.1 Comparison with satellite ocean color observations

The main feature that emerges from the analysis of annual cycles of surface [Chl *a*] from ocean color data over the Mediterranean sea is the coexistence of two main types of cycle (Bosc et al., 2004; D'Ortenzio and Ribera 2009; Lavigne et al., 2013). The two cycles ("NO BLOOM" and BLOOM", following the definition of D'Ortenzio and Ribera d'Alcalà, 2009) can be characterized, firstly, by a 2-fold increase from summer to winter in the normalized [Chl *a*] (so-called NO BLOOM annual cycle) and, secondly, by a moderate (2-fold) increase in normalized [Chl *a*] from summer to winter, followed by an exponential increase (3-fold) in early spring (so-called BLOOM annual cycle). These findings are based on satellite surface [Chl *a*] and result from a complex statistical analysis (i.e., normalization of the seasonal cycles and clustering analysis), but they have also been confirmed by the climatological time series presented here (see Sect. 3.1). Climatologies of [Chl *a*] profiles (Fig. 3) for the southwestern region (Fig. 3b), the Ionian region (Fig. 3c) and the Levantine region (Fig. 3d), which correspond to the NO BLOOM regions identified by D'Ortenzio and Ribera d'Alcalà (2009), display similarities in the seasonal variations in surface [Chl *a*], and they also showed a similar succession of winter homogeneous profiles and summer profiles with DCM. In contrast, the time series corresponding to the northwestern region (Fig. 3a) presents, in March and April, [Chl *a*] vertical profiles characterized by high surface concentrations (i.e., HSC profiles), confirming the specific feature of the northwestern region in the Mediterranean Sea. Unlike NO BLOOM Mediterranean regions, in the northwest region, the average winter MLD is deeper than the DCM and the nitracline depth (see Table 4). This particularity explains the March–April bloom, which could be supported by large winter nutrient supplies and/or the dilution of grazers. It also indicates that winter vertical mixing fully destroys the nitracline, pycnocline and DCM, which have to be restored each year. The annual renewal of these structures contributes to their tight coupling (see Fig. 3a and Table 4), which is not observed in NO BLOOM Mediterranean regions (based on the results in Fig. 3, DCM and pycnocline are uncoupled). In NO BLOOM regions, except for extreme MLD events (Lavigne et al., 2013), the winter MLD does not generally reach the depths reached by the DCM and nitracline during summer (see Table 4).

Beyond the bimodal conception (i.e., BLOOM/NO BLOOM) of annual [Chl *a*] cycles in the Mediterranean Sea, there is an important and unresolved complexity marked by the presence of regional differences within the two main biomass annual cycles. A good illustration of this complexity is the identification by D'Ortenzio and Ribera d'Alcalà (2009) of three different annual cycles (i.e., three bioregions) for the NO BLOOM dynamics. The probabilistic analysis of the general shape of the [Chl *a*] profiles performed in this paper also contributes to refining the basic BLOOM/NO BLOOM scheme and should help to explain the complex patterns observed at the surface. In Fig. 5, regional differences in the distribution of the standard shapes for [Chl *a*] vertical profiles are observed among the NO BLOOM regions (i.e., southwest, Levantine and Ionian regions). The main difference is the significant proportion of *HSC* profiles during the winter months (i.e., January, February and March) in the southwest region, whereas this proportion is very small (less than 10 %) in the Ionian sea and even 0 in the Levantine Sea. The observation of *HSC* profiles in the southwest region suggests that, during winter, mixing is able to supply enough nutrients at the surface to support episodic development of phytoplankton close to the surface when the water column begins to stabilize. This could also explain the higher [Chl *a*] observed in the southwest region and the difference between the southwestern and eastern normalized [Chl *a*] annual cycles (D'Ortenzio and Ribera d'Alcalà, 2009). Compared to the eastern Mediterranean Sea, DCM and nitracline depths are shallow in the southwest region (Table 4). However, winter mixing is constrained, in the Algerian Basin, by the strong halocline associated with the spreading of Atlantic Water, and barely reaches the nitracline depth (D'Ortenzio and Prieur, 2012; Lavigne et al., 2013). As sub-mesoscale activity, associated with jets, fronts and eddies, is also similarly intense in both the southwestern and eastern basin (Rio et al., 2007), our best explanation for the spatial divergences in the occurrence of *HSC* profiles is the regional difference in nutrient stocks below the nitracline. Indeed, for the intermediate layer, the nitrate concentration is much higher in the western than in the eastern basin (Ribera d'Alcalà et al., 2003). In addition, the nitrate-to-phosphate ratio increases eastward, suggesting that phytoplankton growth is mainly limited by phosphate in the eastern Mediterranean Sea (Ribera d'Alcalà et al., 2003; Béthoux et al., 2002; Krom et al., 1991). Hence, the absence of *HSC* profiles in the eastern Mediterranean Sea could be due to a weak mixing efficiency that is not sufficient to support a phytoplankton bloom.

4.2.2 High diversity of the Mediterranean [Chl *a*]

Although the Mediterranean Sea covers a relatively small latitudinal range (from 30 to 45° N), previous findings, essentially based on satellite observations, have shown that in this basin, the annual phytoplankton cycles representative of subtropical and midlatitude regions of the global ocean coexist (D'Ortenzio and Ribera d'Alcalà, 2009; Lavigne et al., 2013). Present results, which focus on the seasonal variability in the whole [Chl *a*] vertical distribution, confirm these previous statements. The climatological time series of [Chl *a*] profiles (Fig. 3) for the southwestern region (Fig. 3b),

the Ionian region (Fig. 3c) and the Levantine region (Fig. 3d) are very close to typical subtropical behavior, marked by the quasi-permanent existence of the DCM (Letelier et al., 2004; Mignot et al., 2014). In particular, the [Chl *a*] climatology of the BATS (Bermuda Atlantic Time-series Study) station in the subtropical North Atlantic gyre (Steinberg et al., 2001; Lavigne et al., 2012) displays many similarities, in terms of ranges of values for [Chl *a*], DCM depths and winter mixing depths, with the climatological time series built in the Levantine Sea (Fig. 3d). The only main difference is that the *homogeneous* climatological profiles begin in December in the Mediterranean regions and only in January at the BATS station (Lavigne et al., 2012). Regarding seasonal cycles obtained for the northwestern Mediterranean Sea, they can be easily compared to midlatitude (40°–60°) regions marked by an intense spring bloom as in the North Atlantic (Siegel et al., 2002) or in certain regions of the Southern Ocean (Thomalla et al., 2011). Similarly to our northwestern Mediterranean observations, the seasonal cycles for [Chl *a*] vertical profiles presented by Boss et al. (2008) in the western North Atlantic (about 50° N) and by Chiswell (2011) in the waters east of New Zealand (about 40° S) display a majority of profiles with a *homogeneous* shape during winter and, in spring, a predominance of profiles displaying an *HSC* shape or a *homogeneous* shape with high [Chl *a*] values. The coexistence of profiles with *homogeneous* and *HSC* shapes during spring could be explained by the intermittent feature of mixing, which continuously modifies the vertical distribution of [Chl *a*] during the spring bloom (Chiswell, 2011). Finally, it is important to mention that the summer situation is very different between the North Atlantic region studied by Boss et al. (2008) and the northwestern Mediterranean Sea. Although DCM profiles are nearly permanent in the northwestern Mediterranean from May or June, Boss et al. (2008) only observed them to start in late summer.

The present study also shows that in the Mediterranean Sea, the specific features of the [Chl *a*] profiles with a *DCM* shape have large variability, comparable to those observed in the global ocean, although occurring on shorter spatial scales. The most relevant indicator is certainly the DCM depth, which was observed to range between 30 m and more than 150m. As expected (e.g., Cullen, 2015), the depth of the Mediterranean DCM is inversely related to the surface [Ch *a*] (Fig. 9). In addition, the relationship between the DCM depth and surface [Chl *a*] (blue curve in Fig. 9) is similar to the relationship reported for the global ocean (red curve in Fig. 9; Mignot et al., 2011). This observation suggests that certain DCM properties in the Mediterranean Sea conform to the same generic properties established for the global ocean.

At the first order, the DCM depth variability in the Mediterranean Sea is related to the spatial component and, in particular, longitude. The deepening of the DCM along a longitudinal gradient (in the present study, DCM deepens by 1.6 m per 1 degree of longitude east) agrees with the previous review, also based on observations, by Crise et al. (1999).

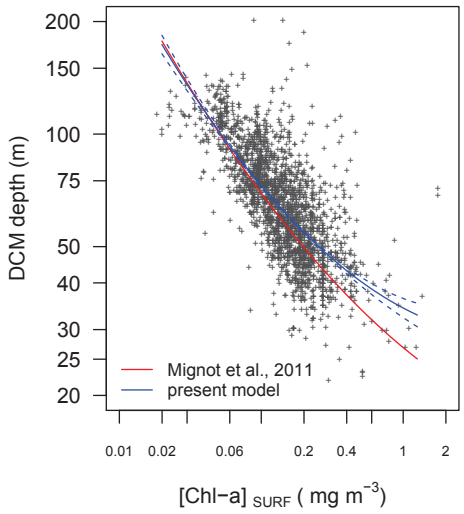


Figure 9. Scatter of the DCM depth as a function of surface [Chl *a*]. Only profiles of the *DCM* standard shape were used for this analysis. Surface [Chl *a*] were obtained from satellite ocean color data. The blue solid line represents a second-order polynomial model determined from present data ($R^2 = 0.52$) with its confidence intervals (blue dotted lines), and the red line represents the model computed by Mignot et al. (2011) from a global ocean data set.

This general deepening of the DCM with longitude covaries with the eastward increase in oligotrophy in the Mediterranean Sea (Béthoux et al., 1998). This pattern is generally attributed to anti-estuarine circulations in the Straits of Gibraltar and Sicily, which generate an eastward inflow of surface nutrient-depleted waters and a westward outflow of deep nutrient-rich waters. In the eastern Mediterranean Sea, oligotrophy is also maintained by poor nutrient inputs from the boundaries (atmosphere and coasts) and by the formation of Levantine Intermediate Water, which is not the product of deep convection but of the subduction of surface water into intermediate water layers (Robinson and Golnaraghi, 1994). As revealed by Table 4, regional changes in DCM depth, nitracline depth and averaged daily PAR at DCM are correlated in the Mediterranean Sea. The eastward deepening of the DCM depth and of the nitracline depth is accompanied by a decrease in the mean daily averaged PAR at DCM (values ranging from 1 mol quanta m^{-2} day^{-1} in the northwest Mediterranean to 0.16 mol quanta m^{-2} day^{-1} in the Levantine Sea). This trend agrees with the "general rule" that states that the DCM builds up where there is an optimal balance between the upward nutrient flux and the downward photon flux and lies on top of the nutricline (Cullen, 2015). The large distance between DCM depth and nitracline depth in the Ionian (36 m) and the Levantine (83m) basins may be considered as contradicting the previous theory. However, as can be seen in Table 4, the estimations of nitracline depth are not likely to be good estimators of the top of the nitracline if the nitrate gradient is not sharp enough, as is the case, for example, in

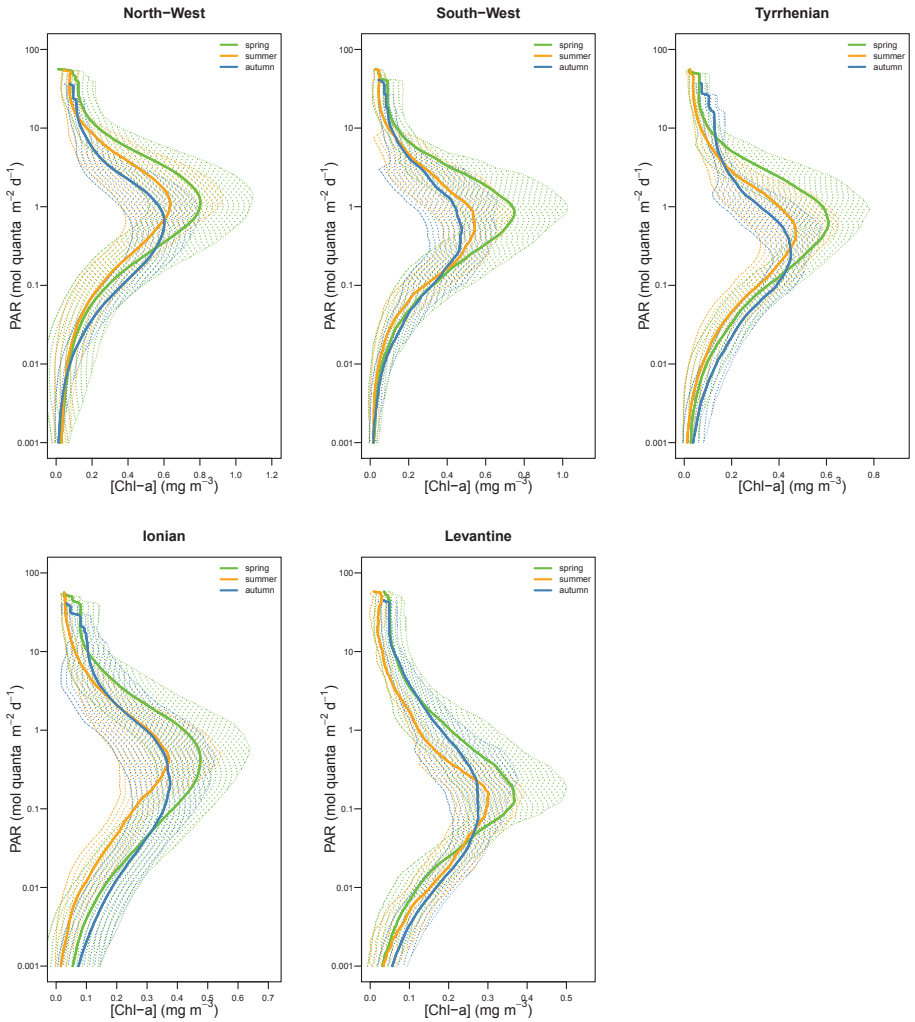


Figure 10. Averaged vertical distribution of [Chl *a*] as a function of PAR with standard deviation (dotted area). Spring refers to the April–June period, summer to July and August, and autumn to the September–November period.

the eastern Mediterranean Sea. Indeed, nitracline depths have been computed from discrete vertical profiles, using the 1 µM isoline (Lavigne et al., 2013).

The results in Fig. 10 also show that a seasonal component contributes to explaining DCM variability in the Mediterranean regions. The observed seasonal pattern of the DCM depth (i.e., deepening from spring to summer and shallowing from summer to autumn) is consistent with previous model results (Macias et al., 2014) and with individual Bio-Argo float observations (Mignot et al., 2014). Letelier et al. (2004) and Mignot et al. (2014) explain this seasonal pattern by considering that the DCM depth might be driven by the light availability and that it would follow the depth of an isolume. This observation is confirmed here by the analysis of the vertical [Chl *a*] profile as a function of irradiance for the spring, summer and autumn periods (Fig. 10). For all regions, from spring to summer, PAR at DCM depth remains unchanged although [Chl *a*] decreases. Accordingly to Letelier et al. (2004), higher spring [Chl *a*] may be explained by the temporal erosion of the upper nitracline from spring to summer, supporting the hypothesis of deep biomass maxima. From summer to autumn, the magnitude of DCMs remains roughly unchanged, similarly to the PAR at DCM.

5 Conclusions

Since the initial work of the MEDAR/MEDATLAS group (Maillard et al., 2005; Manca et al., 2004), the proposed study represents the first attempt to analyze the seasonal variations in the [Chl *a*] vertical distribution over the Mediterranean Sea. The picture of the [Chl *a*] field in the basin has been updated here, as it had been mainly derived from surface satellite data or from limited and scarce in situ observations. Chlorophyll *a* fluorescence data (specifically calibrated and consistently processed with a dedicated method) provided a significantly larger database than the commonly

used in situ bottle estimations. Additionally, a better description of the vertical distribution was made possible. In order to carry out a comprehensive analysis of the seasonal variability in the vertical [Chl *a*] profiles within the main Mediterranean subbasins, 6790 profiles of fluorescence were gathered and processed. The present analysis, in agreement with previous satellite results (D'Ortenzio and Ribera d'Alcalà, 2009), demonstrates the coexistence of two main types of dynamics (i.e., subtropical and midlatitude dynamics) in the Mediterranean Sea. Midlatitude dynamics are observed in the northwestern basin. Their main specificity is the high occurrence of *HSC* profiles in March and April, whereas this type of shape, associated with bloom conditions, is nearly absent elsewhere during this season. The subtropical dynamics encompass most of the remaining basin. They are characterized by an omnipresent DCM from spring to autumn and by a large variety of [Chl *a*] vertical shapes during winter. The present analysis also demonstrated that the [Chl *a*] pattern in the Mediterranean Sea is not uniform. Even among regions with subtropical dynamics, a strong variability was observed in [Chl *a*] values or DCM characteristics. On the basin scale, this variability follows an eastward oligotrophic pattern.

The present study was often limited by the quantity of data, which did not allow for the analysis of each region of the Mediterranean Sea (e.g., the Adriatic Sea). We regret the singular absence of fluorescence profiles in oceanographic databases compared to other parameters. For instance, in the MEDAR database, there are 118 009 salinity profiles, 44 928 oxygen profiles and only 1984 chlorophyll *a* fluorescence profiles. Finally, in this study we were only able to present climatological behavior. Although it is a first and necessary step for a better understanding of processes which impact seasonal variability in [Chl *a*] vertical profiles, it would be interesting to further study certain particular cases showing, with a high-frequency, annual series of vertical [Chl *a*] profiles. These data have now become available with the development of Bio-Argo floats (Johnson et al., 2009) and some studies have already demonstrated their potential for such applications (Boss and Behrenfeld, 2010; Mignot et al., 2014).

Acknowledgements. The authors would like to thank all people involved in the collection and distribution of the oceanographic data used in this paper. We thank PANGAEA, SISMER, the National Oceanographic Data Center and the OGS data center for making data freely available online. The programs LEFE-CYBER, Bio-Argo, MEDAR and IP SESAME are thanked for their valuable contribution to Mediterranean databases. The US NASA Agency is thanked for the easy access to SeaWiFS and MODIS data. We acknowledge the support of the European Commission (Cofunded by the European Union under FP7 – People – Co-funding of Regional, National and International Programmes, GA no. 600407) and of the RITMARE Flagship Project. This work is also the contribution to the PABO (Plateformes Autonomes and Biogeochimie Oceanique), funded by the ANR, to the French "Equipement d'Avenir" NAOS project (ANR J11R107-F) and to the remotely Sensed Biogeochemical Cycles in the Ocean (remOcean) project, funded by the European Research Council (grant agreement 246777). We are also grateful to Louis Prieur, Josephine Ras and Julia Uitz for constructive comments and suggestions.

Edited by: E. Boss

References

Andersen, V. and Prieur, L.: One-month study in the open nw mediterranean sea (dynaproc experiment, may 1995): overview of the hydrobiogeochemical structures and effects of wind events, Deep Sea Res. Pt. I, 47, 397–422, 2000.

Antoine, D., Morel, A., and André, J.-M.: Algal pigment distribution and primary production in the eastern Mediterranean as derived from coastal zone color scanner observations, J. Geophys. Res., 100, 16193–16209, 1995.

Barale, V., Jaquet, J.-M., and Ndiaye, M.: Algal blooming patterns and anomalies in the Mediterranean Sea as derived from the SeaWiFS data set (1998–2003), Remote Sens. Environ., 112, 3300–3313, 2008.

Béthoux, J., Morin, P., Chaumery, C., Connan, O., Gentili, B., and Ruiz-Pino, D.: Nutrients in the Mediterranean Sea, mass balance and statistical analysis of concentrations with respect to environmental change, Mar. Chem., 63, 155–169, 1998.

Béthoux, J. P., Morin, P., and Ruiz-Pino, D. P.: Temporal trends in nutrient ratios: chemical evidence of Mediterranean ecosystem changes driven by human activity, Deep Sea Res. Pt. II, 49, 2007–2016, 2002.

Biermann, L., Guinet, C., Bester, M., Brierley, A., and Boehme, L.: An optimised method for correcting quenched fluorescence yield, Ocean Sci. Discuss., 11, 1243–1264, doi:10.5194/osd-11-1243-2014, 2014.

Bosc, E., Bricaud, A., and Antoine, D.: Seasonal and interannual variability in algal biomass and primary production in the Mediterranean Sea, as derived from 4 years of SeaWiFS observations, Global Biogeochem. Cy., 18, 1–17, 2004.

Boss, E. and Behrenfeld, M.: In situ evaluation of the initiation of the North Atlantic phytoplankton bloom, Geophys. Res. Lett., 37, L18603, doi:10.1029/2010GL044174, 2010.

Boss, E., Swift, D., Taylor, L., Brickley, P., Zaneveld, R., Riser, S., Perry, M., and Strutton, P.: Observations of pigment and particle distributions in the western North Atlantic from an autonomous float and ocean color satellite, Limnol. Oceanogr., 53, 2112–2122, 2008.

Boyer, T. P., Antonov, J. I., Baranova, O. K., Carla, C., Garcia, H. E., Grodsky, A., Johnson, D. R., Locarnini, R. A., Mishonov, A. V., O'Brien, T. D., Paver, C. R., Reagan, J. R., Seidov, D., Smolyar, I. V., and Zwengan, M. M.: World Ocean Database 2013, Sydney Levitus, Ed.; Alexey Mishonov, Technical Ed.; NOAA Atlas NESDIS 72, 209 pp., 2013.

Casotti, R., Landolfi, A., Brunet, C., D'Ortenzio, F., Mangoni, O., Ribera d'Alcalà, M., and Denis, M.: Composition and dynamics of the phytoplankton of the Ionian Sea (eastern Mediterranean), J. Geophys. Res.-Oceans (1978–2012), 108, C98116, doi:10.1029/2002JC001541, 2003.

Cetinic, I., Toro-Farmer, G., Ragan, M., Oberg, C., and Jones, B.: Calibration procedure for Slocum glider deployed optical instruments, Optics Express, 17, 15420–15430, 2009.

Chiswell, S.: Annual cycles and spring blooms in phytoplankton: don't abandon Sverdrup completely, MEPS, 443, 39–50, 2011.

Christaki, U., Giannakourou, A., Van Wambeke, F., and Grégori, G.: Nanoflagellate predation on auto-and heterotrophic picoplankton in the oligotrophic Mediterranean Sea, J. Plank. Res., 23, 1297–1310, 2001.

Claustre, H., Fell, F., Oubelkheir, K., Prieur, L., Sciandra, A., Gentili, B., and Babin, M.: Continuous monitoring of surface optical properties across a geostrophic front: Biogeochemical inferences, Limnol. Oceanogr., 45, 309–321, 2000.

Claustre, H., Hooker, S.B., Van Heukelem, L., Berthon, J-F., Barlow, R., Ras, J., Sessions, S., Targa, C., Thomas, C. S., van der Linde, D., and Marty, J. C.: An intercomparison of hplc phytoplankton pigment methods using in situ samples: application to remote sensing and database activities, Mar. Chem., 85, 41–61, 2004.

Conkright, M., O'Brien, T., Stephens, K., Locarnini, R., Garcia, H., Boyer, T., and Antonov, J.: World Ocean Atlas 2001, Volume 6: Chlorophyll, Ed. S. Levitus, NOAA Atlas NESDIS 54, US Governement Printing Office, Washington, DC, 46 pp., 2002.

Crise, A., Allen, J., Baretta, J., Crispi, G., Mosetti, R., and Solidoro, C.: The Mediterranean pelagic ecosystem response to physical forcing, Prog. Oceanogr., 44, 219–243,1999.

Crombet, Y., Leblanc, K., Quéguiner, B., Moutin, T., Rimmelin, P., Ras, J., Claustre, H., Leblond, N., Oriol, L., and Pujo-Pay, M.: Deep silicon maxima in the stratified oligotrophic Mediterranean Sea, Biogeosciences, 8, 459–475, doi:10.5194/bg-8-459-2011, 2011.

Cullen, J.: The Deep Chlorophyll Maximum – Comparing Vertical Profiles of Chlorophyll-A, Can. J. Fish. Aqua. Sci., 39, 791–803, 1982.

Cullen, J. J.: Subsurface chlorophyll maximum layers: enduring enigma or mystery solved?, Ann. Rev. Mar. Sci., 7, 207–239, 2015.

Cullen, J. and Lewis, M.: Biological Processes and Optical Measurements Near the Sea Surface: Some Issues Relevant to Remote Sensing, J. Geophys. Res.-Oceans, 100, 13255–13266, 1995.

de Boyer Montegut, C., Madec, G., Fischer, A., Lazar, A., and Iudicone, D.: Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology, J. Geophys. Res.-Oceans, 109, C12003, doi:10.1029/2004JC002378, 2004.

Devred, E., Sathyendranath, S., and Platt, T.: Delineation of ecological provinces using ocean colour radiometry, Mar. Ecol. Prog. Ser., 346, 1–13, 2007.

Dolan, J., Vidussi, F., and Claustre, H.: Planktonic ciliates in the Mediterranean Sea: longitudinal trends, Deep Sea Res. Pt. I, 46, 2025–2039, 1999.

Dolan, J., Claustre, H., Carlotti, F., Plounevez, S., and Moutin, T.: Microzooplankton diversity: relationships of tintinnid ciliates with resources, competitors and predators from the Atlantic Coast of Morocco to the Eastern Mediterranean, Deep Sea Res. Pt. I, 49, 1217–1232, 2002.

D'Ortenzio, F. and Ribera d'Alcalà , M.: On the trophic regimes of the Mediterranean Sea: a satellite analysis, Biogeosciences, 6, 139–148, doi:10.5194/bg-6-139-2009, 2009.

D'Ortenzio, F. and Prieur, L.: The upper mixed layer, Life in the Mediterranean Sea: A look at habitat changes, edited by: Noga Stambler, Nova Science Publisher, 127–156, 2012.

D'Ortenzio, F., Iudicone, D., Montegut, C., Testor, P., Antoine, D., Marullo, S., Santoleri, R., and Madec, G.: Seasonal variability of the mixed layer depth in the Mediterranean Sea as derived from in situ profiles, Geophys. Res. Lett., 32, L12605, doi:10.1029/2005GL022463, 2005.

D'Ortenzio, F., Thierry, V., Eldin, G., Claustre, H., Testor, P., Coatanoan, C., Tedetti, M., Guinet, C., Poteau, A., Prieur, L., and Lefevre D., Bourrin F., Carval T., Goutx M., Garçon V., Thouron D., Lacombe M., Lherminier P., Loisiel H., Mortier L., and Antoine D.: White Book on Oceanic Autonomous Platforms for Biogeochemical Studies: Instrumentation and Measure (PABIM) version 1.3, available from: http://www.obs-vlfr.fr/OAO/file/PABIM_white_book_version1.3.pdf, 2010.

D'Ortenzio, F., Lavigne, H., Besson, F., ClaLavigne, H., Coppola, L., Garcia, N., Laes-Huon, A.s-Huon, A., Le Reste, S., Malardé, D., Migon, C., Morin, P., Mortier, L., Poteau, A., Prieur, L., Raimbault, P., and Testor, P.: Observing mixed layer depth, nitrate and chlorophyll concentrations in the northwestern Mediterranean: A combined satellite and NO_3 profiling floats experiment, Geophys. Res. Lett., 41, 6443–6451, doi:10.1002/2014GL061020, 2014.

Dugdale, R. and Wilkerson, F.: Nutrient sources and primary production in the Eastern Mediterranean, Oceanol. Acta, 9, 179–184, 1988.

Durrieu de Madron, X., Guieu, C., Sempéré, R., Conan, P., Cossa, D., D'Ortenzio, F., Estournel, C., Gazeau, F., Rabouille, C., Stemmann, L., Bonnet, F. Diaz, P. Koubbi, O. Radakovitch, M. Babin, M. Baklouti, Bancon-Montignya, C., Belviso, S., Bensoussan, N., Bonsang, B., Bouloubassi, I., Brunet, C., Cadiou, J.-F., Carlotti, F., Chami, M., Charmasson, S., Charrière, B., Dachsa, J., Doxaran, D., Dutay, J.-C., and Elbaz-Poulicheta, F.: Marine ecosystems responses to climatic and anthropogenic forcings in the mediterranean, Prog. Oceanogr., 91, 97–166, 2011.

Estrada, M., Marrase, C., Latasa, M., Berdalet, E., Delgado, M., and Riera, T.: Variability of deep chlorophyll maximum characteristics in the Northwestern Mediterranean, Mar. Ecol.-Prog. Ser., 92, 289–289, 1993.

Feldman, G., Kuring, N., Ng, C., Esaias, W., McClain, C., Elrod, J., Maynard, N., and Endres, D.: Ocean-color: availability of the global data set, EOS, p. 634, 1989.

Franz, B., Werdell, P., Meister, G., Bailey, S., Eplee, R., Feldman, G., Kwiatkowskaa, E., McClain, C., Patt., F., and Thomas, D.: The continuity of ocean color measurements from SeaWiFS to MODIS, in: Earth observing systems: X. Proceedings SPIE, edited by: Butler, J. J., vol. 5882, The International Society for Optical Engineering, 1–13, 2005.

Gieskes, W. W. and Kraay, G. W.: Unknown Chlorophyll a Derivatives in the North Sea and the Tropical Atlantic Ocean Revealed by HPLC Analysis, Limnol. Oceanogr., 28, 757–766, 1983.

Huisman, J., Oostveen, P. v., and Weissing, F. J.: Critical Depth and Critical Turbulence: Two Different Mechanisms for the Development of Phytoplankton Blooms, Limnol. Oceanogr., 44, 1781–1787, 1999.

Ignatiades, L., Gotsis-Skretas, O., Pagou, K., and Krasakopoulou, E.: Diversification of phytoplankton community structure and re-

lated parameters along a large-scale longitudinal east-west transect of the Mediterranean Sea, J. Plank. Res., 31, 411–428, 2009.

Johnson, K. S., Berelson, W. M., Boss, E. S., Chase, Z., Claustre, H., Emerson, S. R., Gruber, N., Kortzinger, A., Perry, M. J., and Riser, S. C.: Observing biogeochemical cycles at global scales with profiling floats and gliders: Prospects for a global array, Oceanography, 22, 216–225, 2009.

Kiefer, D. A.: Fluorescence properties of natural phytoplankton populations, Mar.Biol., 22, 263–269, 1973.

Kolber, Z. and Falkowski, P. G.: Use of Active Fluorescence to Estimate Phytoplankton Photosynthesis in Situ, Limnol. Oceanogr., 38, 1646–1665, 1993.

Krom, M., Brenner, S., Kress, N., Neori, A., and Gorgon, L.: Nutrient Dynamics and New Production in a Warm-Core Eddy from the Eastern Mediterranean-Sea, Deep-Sea Res. Pt. A, 39, 467–480, 1992.

Krom, M. D., Kress, N., Brenner, S., and Gordon, L. I.: Phosphorus Limitation of Primary Productivity in the Eastern Mediterranean Sea, Limnol. Oceanogr., 36, 424–432, 1991.

Lavigne, H., D'Ortenzio, F., Claustre, H., and Poteau, A.: Towards a merged satellite and in situ fluorescence ocean chlorophyll product, Biogeosciences, 9, 2111–2125, doi:10.5194/bg-9-2111-2012, 2012.

Lavigne, H., D'Ortenzio, F., Migon, C., Claustre, H., Testor, P., d'AlcalÃ , M. R., Lavezza, R., Houpert, L., and Prieur, L.: Enhancing the comprehension of mixed layer depth control on the Mediterranean phytoplankton phenology, J. Geophys. Res.-Oceans, 118, 3416–3430, 2013.

Letelier, R. M., Karl, D. M., Abbott, M. R., and Bidigare, R. R.: Light Driven Seasonal Patterns of Chlorophyll and Nitrate in the Lower Euphotic Zone of the North Pacific Subtropical Gyre, Limnol. Oceanogr., 49, 508–519, 2004.

Longhurst, A. R.: Ecological Geography of the Sea, Second Edition, Academic Press, 2 Edn., 2006.

Longhurst, A. R. and Harrison, G. W.: The biological pump: profiles of plankton production and consumption in the upper ocean, Prog. Oceanogr., 22, 47–123, 1989.

Lorenzen, C. J.: A method for the continuous measurement of in vivo chlorophyll concentration, Deep-Sea Res., 13, 223–227, 1966.

Macias, D., Stips, A., and Garcia-Gorriz, E.: The relevance of deep chlorophyll maximum in the open Mediterranean Sea evaluated through 3D hydrodynamic-biogeochemical coupled simulations, Ecol. Modell., 281, 26–37, 2014.

Maillard, C. et al.: MEDAR/MEDATLAS 1998–2001: A Mediterranean and Black Sea oceanographic data base and network, 46, 329–344, 2005.

Manca, B., Burca, M., Giorgetti, A., Coatanoan, C., Garcia, M.-J., and Iona, A.: Physical and biochemical averaged vertical profiles in the Mediterranean regions: an important tool to trace the climatology of water masses and to validate incoming data from operational oceanography, J. Mar. Syst., 48, 83–116, 2004.

Marty, J.-C., Chiaverini, J., Pizay, M.-D., and Avril, B.: Seasonal and interannual dynamics of nutrients and phytoplankton pigments in the western Mediterranean Sea at the DYFAMED time-series station (1991–1999), Deep Sea Res. Pt. II, 49, 1965–1985, 2002.

McClain, C., Cleave, M., Feldman, G., Gregg, W., Hooker, S., and Kuring, N.: Science quality SeaWiFS data for global biosphere research, Sea Technol., 39, 10–16, 1998.

Mignot, A., Claustre, H., D'Ortenzio, F., Xing, X., Poteau, A., and Ras, J.: From the shape of the vertical profile of in vivo fluorescence to Chlorophyll *a* concentration, Biogeosciences, 8, 2391–2406, doi:10.5194/bg-8-2391-2011, 2011.

Mignot, A., Claustre, H., Uitz, J., Poteau, A., D'Ortenzio, F., and Xing, X.: Understanding the seasonal dynamics of phytoplankton biomass and the deep chlorophyll maximum in oligotrophic environments: A Bio-Argo float investigation, Global Biogeochem. Cy., 28, 856–876, 2014.

Müller, P., Li, X.-P., and Niyogi, K. K.: Non-Photochemical Quenching. A Response to Excess Light Energy, Plant Physiol., 125, 1558–1566, 2001.

Morel, A. and André, J.-M.: Pigment Distribution and Primary Production in the Western Mediterranean as Derived and Modeled From Coastal Zone Color Scanner Observations, J. Geophys. Res., 96, 12685–12698, 1991.

Morel, A. and Berthon, J.: Surface Pigments, Algal Biomass Profiles, and Potential Production of the Euphotic Layer – Relationships Reinvestigated in View of Remote-Sensing Applications, Limnol. Oceanogr., 34, 1545–1562, 1989.

Morel, A. and Maritorena, S.: Bio-optical properties of oceanic waters: A reappraisal, J. Geophys. Res.-Oceans, 106, 7163–7180, 2001.

Moutin, T. and Raimbault, P.: Primary production, carbon export and nutrients availability in western and eastern Mediterranean Sea in early summer 1996 (MINOS cruise), J. Mar. Syst., 33-34, 273–288, 2002.

Moutin, T. and Prieur, L.: Influence of anticyclonic eddies on the Biogeochemistry from the Oligotrophic to the Ultraoligotrophic Mediterranean (BOUM cruise), Biogeosciences, 9, 3827–3855, doi:10.5194/bg-9-3827-2012, 2012.

Moutin, T., Van Wambeke, F., and Prieur, L.: Introduction to the Biogeochemistry from the Oligotrophic to the Ultraoligotrophic Mediterranean (BOUM) experiment, Biogeosciences, 9, 3817–3825, doi:10.5194/bg-9-3817-2012, 2012.

Navarro, G., Vázquez, Á., Macás, D., Bruno, M., and Ruiz, J.: Understanding the patterns of biological response to physical forcing in the Alborán Sea (western Mediterranean), Geophys. Res. Lett., 38, L23606, doi:0.1029/2011GL049708, 2011.

Platt, T., Sathyendranath, S., White III, G. N., Fuentes-Yaco, C., Zhai, L., Devred, E., and Tang, C.: Diagnostic properties of phytoplankton time series from remote sensing, Estuar. Coast., 33, 428–439, 2010.

Psarra, S., Tselepides, A., and Ignatiades, L.: Primary productivity in the oligotrophic Cretan Sea (NE Mediterranean): seasonal and interannual variability, Prog. Oceanogr., 46, 187–204, 2000.

Ribera d'Alcala', M., Civitarese, G., Conversano, F., and Lavezza, R.: Nutrient ratios and fluxes hint at overlooked processes in the Mediterranean Sea, J. Geophys. Res., 108, 8106, doi:10.1029/2002JC001650, 2003.

Rio M.-H., Poulain P.-M., Pascual A., Mauri, E., Larnicol, G., and Santoleri, R.: A mean dynamic topography of the Mediterranean Sea computed from altimetric data, in-situ measurements and a general circulation model, J. Mar. Syst., 65, 484–508, 2007.

Robinson, A. and Golnaraghi, M.: The physical and dynamical oceanography of the Mediterranean Sea, in: Ocean processes in

climate dynamics: Global and Mediterranean examples, vol. 419 of NATO Science Series, 255–306, 1994.

Röttgers, R. and Koch, B. P.: Spectroscopic detection of a ubiquitous dissolved pigment degradation product in subsurface waters of the global ocean, Biogeosciences, 9, 2585–2596, doi:10.5194/bg-9-2585-2012, 2012.

Sackmann, B. S., Perry, M. J., and Eriksen, C. C.: Seaglider observations of variability in daytime fluorescence quenching of chlorophyll *a* in Northeastern Pacific coastal waters, Biogeosciences Discuss., 5, 2839–2865, doi:10.5194/bgd-5-2839-2008, 2008.

Sharples, J., Moore, C. M., Rippeth, T. P., Holligan, P. M., Hydes, D. J., Fisher, N. R., and Simpson, J. H.: Phytoplankton Distribution and Survival in the Thermocline, Limnol. Oceanogr., 46, 486–496, 2001.

Siegel, D. A., Doney, S. C., and Yoder, J. A.: The North Atlantic Spring Phytoplankton Bloom and Sverdrup's Critical Depth Hypothesis, Science, 296, 730–733, 2002.

Siokou-Frangou, I., Christaki, U., Mazzocchi, M. G., Montresor, M., Ribera d'Alcalà, M., Vaque, D., and Zingone, A.: Plankton in the open Mediterranean Sea: a review, Biogeosciences, 7, 1543–1586, doi:10.5194/bg-7-1543-2010, 2010.

Sosik, H.M., Chisholm, S. C., and Olson, R. J.: Chlorophyll Fluorescence from Single Cells: Interpretation of flow cytometric signals, Limnol. Oceanogr., 34, 1749–1761, 1989.

Steinberg, D., Carlson, C., Bates, N., Johnson, R., Michaels, A., and Knap, A.: Overview of the US JGOFS Bermuda Atlantic Time-series Study (BATS): a decade-scale look at ocean biology and biogeochemistry, Deep-Sea Res. Pt. II, 48, 1405–1447, 2001.

Strass, V.: On the calibration of large-scale fluorometric chlorophyll measurements from towed undulating vehicles, Deep Sea Res. Pt. A. Oceanogr. Res. Papers, 37, 525–540, 1990.

Strickland, J.: Production of organic matter in primary stages of the marine food chain, in: Chemical Oceanography, Academic Press, London, edited by: J. P. Riley and G. Skirrow, 477–610, 1965.

Taupier-Letage, I., Puillat, I., Millot, C., and Raimbault, P.: Biological response to mesoscale eddies in the Algerian Basin, J. Geophys. Res.-Oceans (1978–2012), 108, 3245, doi:10.1029/1999JC000117, 2003.

Thomalla, S. J., Fauchereau, N., Swart, S., and Monteiro, P. M. S.: Regional scale characteristics of the seasonal cycle of chlorophyll in the Southern Ocean, Biogeosciences, 8, 2849–2866, doi:10.5194/bg-8-2849-2011, 2011.

Uitz, J., Claustre, H., Morel, A., and Hooker, S.: Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll, J. Geophys. Res.-Oceans, 111, C08005, doi:10.1029/2005JC003207, 2006.

Volpe, G., Nardelli, B. B., Cipollini, P., Santoleri, R., and Robinson, I. S.: Seasonal to interannual phytoplankton response to physical processes in the Mediterranean Sea from satellite observations, Remote Sens. Environ., 117, 223–235, 2012.

Xing, X., Morel, A., Claustre, H., Antoine, D., D'Ortenzio, F., Poteau, A., and Mignot, A.: Combined processing and mutual interpretation of radiometry and fluorimetry from autonomous profiling Bio-Argo floats: Chlorophyll a retrieval, J. Geophys. Res.-Oceans, 116, C0620, doi:10.1029/2010JC006899, 2011.

Xing, X., Claustre, H., Blain, S., D'Ortenzio, F., Antoine, D., Ras, J., and Guinet, C.: Quenching correction for in vivo chlorophyll fluorescence measured by instrumented elephant seals in the Kerguelen region (Southern Ocean), Limnol. Oceanogr.-Methods, 10, 483–495, 2012.

Yoder, J. A., McClain, C. R., Feldman, G. C., and Esaias, W. E.: Annual cycles of phytoplankton chlorophyll concentrations in the global ocean: A satellite view, Global Biogeochem. Cy., 7, 181–193, 1993.

Résumé :
Les travaux présentés dans cette thèse concernent la paramétrisation de la distribution verticale de la biomasse et de la structure des communautés phytoplanctoniques dans l'océan global. Nous avons tout d'abord développé une méthode neuronale de calibration de la fluorescence en concentration en chlorophylle-*a* ([Chl]) associée à la biomasse phytoplanctonique totale et à trois classes de taille de phytoplancton (i.e. micro-, nano- et pico-phytoplancton). Cette méthode, FLAVOR (Fluorescence to Algal communities Vertical distribution in the Oceanic Realm), a été entrainée et validée à l'aide une base de données de ~900 profils concomitants de fluorescence et de pigments mesurés pat HPLC. Une base de données globale de ~49000 profils de fluorescence a ensuite été assemblée et calibrée en termes de biomasse chlorophyllienne et composition du phytoplancton. Ce travail représente une première étape vers une vision globale tridimensionnelle de la biomasse phytoplanctonique. Nous avons ensuite développé deux réseaux de neurones (SOCA : Satellite Ocean Color and Argo data to infer vertical distribution of bio-optical properties) pour estimer la distribution verticale de deux indicateurs bio-optiques de la biomasse phytoplanctonique, [Chl] et le coefficient de rétrodiffusion particulaire. Ces réseaux de neurones requièrent comme données d'entrée des données satellites de couleur de l'eau co-localisées avec un profil hydrologique collecté par un flotteur Argo. Ils ont été entrainés et validés avec une base de données globale composée de ~5 000 profils de propriétés bio-optiques et hydrologiques acquises par des flotteurs Bio-Argo de façon concomitante aux données satellitaires. Les bases de données utilisées pour développer les méthodes FLAVOR et SOCA proviennent de régions océaniques très variées, représentatives de l'océan global, permettant ainsi l'application de ces deux méthodes à la majorité des eaux océaniques. Finalement, nous avons mené une étude focalisée sur l'Atlantique Nord qui exploite les outils développés. Les champs tridimensionnels de biomasse obtenus, couplés à un modèle bio-optique de production primaire, permettent d'étudier les cycles saisonniers de la distribution verticale de la biomasse phytoplanctonique et de la production primaire dans différentes bio-régions de l'Atlantique Nord.

Mots clés : [phytoplancton, chlorophylle-*a*, bio-optique, fluorescence, coefficient de rétrodiffusion particulaire, couleur de l'eau, flotteurs profileurs Bio-Argo, réseau de neurones, échelle globale]

Study and parameterization of the vertical distribution of phytoplankton biomass in the global ocean

Abstract :
This PhD work focuses on the parameterization of the vertical distribution of phytoplankton biomass and community structure in the global open ocean. First we have developed a neural network-based method for the calibration of the fluorescence in chlorophyll *a* concentration [Chl] associated with the total phytoplankton biomass and with three phytoplankton size classes (i.e. micro, nano- and pico-phytoplankton). This method, (FLAVOR for Fluorescence to Algal communities Vertical distribution in the Oceanic Realm), was trained and validated using a database of ~900 concomitant fluorescence and HPLC-determined pigment profiles. A global database comprising ~49 000 fluorescence profiles was assembled and calibrated with FLAVOR. The resulting database represents a first step towards a global three-dimensional view of phytoplankton biomass and community composition. Second, two neural networks (SOCA for Satellite Ocean Color and Argo data to infer vertical distribution of bio-optical properties) were developed to infer the vertical distribution of two bio-optical proxies of the phytoplankton biomass, [Chl] and the particulate backscattering coefficient, using as input satellite-derived products matched up with a hydrological Argo profile. The SOCA methods were trained and validated using a global database of ~5 000 profiles of bio-optical and hydrological properties collected from Bio-Argo floats with concomitant satellite products. The database used to develop FLAVOR and SOCA originates from various oceanic regions largely representative of the global ocean, making the methods applicable to most oceanic waters. Finally, we proposed a study dedicated to the North Atlantic where the tools developed in this thesis are used in conjunction with a bio-optical primary production model. This allows us to characterize the seasonal cycle of the vertical distribution of the phytoplankton biomass and primary production in various bio-regions of the North Atlantic.

Keywords : [phytoplankton, chlorophyll-*a*, bio-optics, fluorescence, particulate backscattering, ocean color, Bio-Argo profiling floats, neural networks, global scale]