



HAL
open science

Téledétection des groupes phytoplanctoniques via l'utilisation conjointe de mesures satellites, in situ et d'une méthode de classification automatique

Zied Ben Mustapha

► **To cite this version:**

Zied Ben Mustapha. Téledétection des groupes phytoplanctoniques via l'utilisation conjointe de mesures satellites, in situ et d'une méthode de classification automatique. Océanographie. Université du Littoral Côte d'Opale, 2013. Français. NNT : 2013DUNK0405 . tel-01367831

HAL Id: tel-01367831

<https://theses.hal.science/tel-01367831v1>

Submitted on 28 Jun 2017

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.

Ecole doctorale « Sciences de la Matière, du Rayonnement et de l'Environnement »

Université du Littoral Côte d'Opale

CNRS UMR LOG 8187

Equipe d'Océanographie Physique, Transport et Télédétection

ZIED BEN MUSTAPHA

THESE DE DOCTORAT

En vue de l'obtention du titre de Docteur en Sciences de la Matière, du Rayonnement et de l'Environnement

Spécialité : Océanographie

Sujet de thèse

Télédétection des groupes phytoplanctoniques via l'utilisation conjointe de mesures satellites, in situ et d'une méthode de classification automatique

Directeur de thèse :	Mr. Hubert LOISEL (Pr, ULCO/LOG)
Co-Directeurs de thèse :	Mme. Séverine ALVAIN (CR, CNRS/LOG) Mr. Cédric JAMET (MC, ULCO/LOG)
Examineur et président du jury:	Mr. Yves DANDONNEAU (DR émérite, IRD)
Rapporteurs :	Mme. Sylvie THIRIA (Pr, LOCEAN/UPMC) Mr. Cyril MOULIN (HDR, LSCE/CEA/IPSL)

Remerciements

D'abord, je remercie les membres du jury d'avoir accepté d'évaluer ce travail, notamment, Mme Sylvie Thiria, Mr Cyril Moulin et Mr Yves Dandonneau.

J'adresse également ma sincère reconnaissance à Mr François Schmitt pour m'avoir accueilli au sein du Laboratoire d'Océanologie et de Géosciences.

Cette étude n'aurait pu être menée sans le soutien financier de la Région Nord-Pas de Calais et du Syndicat Mixte de la Côte d'Opale. Le travail accompli s'inscrit également dans le cadre de deux programmes de recherche : TOSCA « PHYTOCOT » du Centre National d'Etudes Spatiales et le projet INTERREG « DYMAPHY ».

Mes sincères remerciements vont également à Mr Rachid Amara pour son accueil au sein de la Maison de la Recherche en Environnement Naturel, de même que pour sa bienveillance et précieux conseils.

Un grand merci à Mr Hubert Loisel d'avoir dirigé cette thèse et m'avoir accueilli au sein de l'équipe d'Océanographie physique, transport et télédétection du LOG. Je le remercie bien vivement pour les conseils scientifiques et humains qu'il m'a donnés. «... Il faut faire preuve d'humilité, ce qui permet d'avancer souvent deux fois plus vite, et bien écouter les conseils...». «...Les vietnamiens plantent le riz, les cambodgiens le regardent pousser, les Laotiens l'écoutent pousser...». Je tacherai de viser les vietnamiens Hubert. Je n'oublierai jamais ces conseils !

Après ces trois années passées au LOG, j'exprime toute ma gratitude à Mme Séverine Alvain et Mr Cédric Jamet pour m'avoir offert la possibilité de travailler sur ce sujet de thèse, grâce auquel, j'ai découvert la couleur de l'océan et ses potentielles applications en océanographie, de même que pour avoir méthodiquement dirigé et encadré ce travail. Leurs conseils scientifiques pointus et avisés, disponibilité et patience ont joué un rôle pertinent dans le déroulement de la thèse. Par ailleurs, je leur suis très reconnaissant pour m'avoir donné l'occasion de partager ce que j'ai appris et d'apprendre, à travers les diverses conférences et ateliers. Ce fut également une expérience humainement, culturellement et scientifiquement très enrichissante que de participer à la mission MELVILLE 2011 qui restera, pour moi, une aventure inoubliable. Merci Séverine !

J'adresse mes sincères remerciements à tout le personnel du LOG pour le cadre scientifique et humain d'exception ainsi qu'à l'équipe « Océanographie physique,

transport et télédétection », notamment à Mr David Dessailly pour son précieux soutien en programmation, Mr Régis Sion pour son assistance en informatique (pour sa bonne bière artisanale également), Mr Vincent Vantrepotte et sa bonne humeur inégalable, Clémence Goyens avec qui j'ai eu le plaisir de « bavarder » et partager le bureau durant ces trois années dans la bonne humeur (bon courage pour la suite Clem !), sans oublier le reste de l'équipe. Merci également à Mr Charles Verpoorter, dont les conseils avisés et les discussions scientifiques intéressantes m'ont apporté beaucoup de soutien dans la phase de rédaction de ma thèse.

Je pense également aux personnes qui m'ont moralement soutenu durant ce projet difficile. Je ne saurais comment remercier mes parents et ma sœur qui n'ont jamais cessé de croire en moi et m'ont toujours donné les bons conseils, sans lesquels, cette étape aurait été encore plus difficile.

Merci à ma chère amie Isabelle, son magnifique « petit Macaque » ainsi qu'à toute sa famille pour leur présence et leur accueil chaleureux, grâce auquel je me suis si facilement intégré chez les « Ch'tis » ! Une grande pensée à mon « GI JOE » (il se reconnaîtra) pour les précieux moments passés en sa compagnie. J'ai eu le plaisir de connaître quelqu'un que tout (ou presque) passionne ! On se retrouvera sûrement quelque part mon ami et comme tu le dis.....bip-bip !!!

Je remercie Mr Slim Gana qui m'a accompagné, soutenu et bien conseillé durant tout mon cursus universitaire.

Je salue toutes les personnes que j'ai eu le plaisir de connaître et côtoyer tout au long de ces trois années passées au LOG, notamment, Achwak, Amélie, Céline, Erwan, Natacha, Renosh, Virginie, Elodie, Morgane, Corinne, Vincent...Merci également à Amine, Aurélie, Mouadh, Riadh et Tarek, votre présence m'a continuellement ressourcé malgré la distance !

Merci à toutes les personnes que j'aurais oubliées...

Résumé

La télédétection de la couleur de l'océan représente un outil adapté à l'observation du phytoplancton avec des résolutions spatio-temporelles élevées et pouvant être adaptées à chaque cas d'étude. Plusieurs méthodes ont été développées ces dernières années afin de permettre la distinction de différents groupes de phytoplancton en utilisant les données des capteurs de la couleur de l'océan. Dans le cadre de cette thèse, on présente une nouvelle approche, appelée PHYSAT-SOM (Ben Mustapha et al., 2013), qui se base sur l'application d'un algorithme de classification automatique non supervisée (SOM ou Self-Organizing Maps, Kohonen, 1984) à l'extraction de différentes formes et amplitudes de spectres d'anomalies de luminances (Ra ou Radiance Anomaly). Cette anomalie spectrale a été définie par Alvain et al. (2005), lors du développement de la méthode PHYSAT et il est actuellement admis que sa variabilité est reliée à celle de la composition des communautés phytoplanctoniques (Alvain et al., 2012). L'utilisation des SOM vise à améliorer la caractérisation de la variabilité des Ra en termes de forme et amplitude ainsi que l'expansion du potentiel de leur utilisation à de grandes bases de données in situ de pigments.

En considérant un même jeu de données de spectres de Ra, une comparaison entre la précédente version de PHYSAT (Alvain et al., 2005, 2008) et la nouvelle approche, basée sur SOM (Ben Mustapha et al., 2013), a montré qu'il est maintenant possible de couvrir toute la variabilité spectrale des Ra. Ceci n'était pas le cas avec l'ancienne approche du fait de l'utilisation de seuils, définis dans le but d'éviter les chevauchements entre les signatures spectrales des différents groupes de phytoplancton. La méthode basée sur SOM est pertinente pour caractériser une grande variété de spectres de Ra, de par sa capacité à gérer de grandes quantités de données et de sa fiabilité statistique. La première approche aurait pu, de ce fait, introduire des biais potentiels et donc, les possibilités de son extension à de plus grandes bases de données in situ étaient relativement restreintes. Par la suite, SOM a été utilisé pour classer les spectres de Ra fréquemment observés à l'échelle globale. Ces spectres ont ensuite été empiriquement reliés à différents groupes de phytoplancton, identifiés à partir de données in situ de pigments. Cette classification a été appliquée aux archives satellite du capteur SeaWiFS, permettant l'étude de la distribution globale de chaque groupe. Grâce à sa capacité à caractériser un large éventail de spectres de Ra et de gérer une plus grande base de données in situ, l'outil SOM permet de classer un nombre plus élevé de pixels (2 fois plus) que la précédente approche de PHYSAT. En outre, différentes signatures spectrales de Ra ont été associées aux diatomées. Ces signatures sont situées dans divers environnements où les propriétés optiques inhérentes affectant les spectres Ra sont susceptibles d'être significativement différentes. Par ailleurs, les floraisons de diatomées dans certaines conditions sont plus clairement visibles avec la nouvelle méthode.

La méthode PHYSAT-SOM offre ainsi plusieurs perspectives afin d'aller plus loin dans l'utilisation des données de la couleur de l'océan pour la détection des groupes de phytoplancton. On peut citer l'exemple d'une application future dans les eaux du Cas 2, moyennant une approche de normalisation adéquate du signal de luminances. Une étude préliminaire en Manche et Mer du Nord est présentée dans le dernier chapitre, montrant qu'il sera possible d'utiliser PHYSAT-SOM dans cet environnement optiquement complexe.

Mots Clés : Phytoplancton, Couleur de l'Océan, Pigments, Réseaux de neurones, Cartes auto-organisatrices, PHYSAT.

Abstract

Remote sensing of ocean color is a powerful tool for monitoring phytoplankton in the ocean with a high spatial and temporal resolution. Several methods were developed in the past years for detecting phytoplankton functional types from satellite observations. In this thesis, we present an automatic classification method, based on a neural network clustering algorithm, in order to classify the anomalies of water leaving radiances spectra (Ra), introduced in the PHYSAT method by Alvain et al. (2005), and analyze their variability at the global scale (Ben Mustapha et al., 2013). The use of an unsupervised classification aims at improving the characterization of the spectral variability of Ra in terms of shape and amplitude as well as the expansion of its potential use to larger *in situ* datasets for global phytoplankton remote sensing. The Self-Organizing Map algorithm (SOM, Kohonen, 1984) aggregates similar spectra into a reduced set of pertinent groups, allowing the characterization of the Ra variability, which is known to be linked with phytoplankton community composition (Alvain et al. 2012).

Based on the same sample of Ra spectra, a comparison between the previous version of PHYSAT (Alvain et al., 2005, 2008) and the new one using SOM (Ben Mustapha et al., 2013) shows that it is now possible to take into consideration all the types of spectra. This was not possible with the previous approach, based on thresholds, defined in order to avoid overlaps between the spectral signatures of each phytoplankton group. The SOM-based method is relevant for characterizing a wide variety of Ra spectra through its ability to handle large amounts of data, in addition to its statistical reliability compared to the previous PHYSAT. The former approach might have introduced potential biases and thus, its extension to larger databases was very restricted. In a second step, some new Ra spectra have been related to phytoplankton groups using collocated field pigments inventories from a large *in situ* database. Phytoplankton groups were identified based on biomarker pigments ratios thresholds taken from the literature. SOM was then applied to the global daily SeaWiFS imagery archive between 1997 and 2010. Global distributions of major phytoplankton groups were analyzed and validated against *in situ* data. Thanks to its ability to capture a wide range of spectra and to manage a larger *in situ* pigment dataset, the neural network tool allows to classify a much higher number of pixels (2 times more) than the previous PHYSAT method for the five phytoplankton groups taken into account in this study (*Synechococcus*-Like-Cyanobacteria, diatoms, *Prochlorococcus*, Nanoeucaryots and *Phaeocystis-like*). In addition, different Ra spectral signatures have been associated to diatoms. These signatures are located in various environments where the inherent optical properties affecting the Ra spectra are likely to be significantly different. Local phenomena such as diatoms blooms in the upwelling regions or during climatic events (i.e. *La Niña*) are more clearly visible with the new method.

The PHYSAT-SOM method provides several perspectives concerning the use of ocean color remote sensing data for phytoplankton groups identification, such as, the potential application of the method in Case 2 waters, using an appropriate nLw signal normalization approach. A preliminary case study in the English Channel and North Sea waters is presented in the last chapter of the thesis, showing the possibility of a future use of PHYSAT-SOM in these optically complex waters.

Keywords: Phytoplankton, Ocean Color, Pigments, Neural Networks, Self-Organizing Maps, PHYSAT

Valorisation des travaux

Publication scientifique :

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, sous presse.

Communications

- Atelier national sur la couleur de l'océan du GIS COOC à Nice en Janvier 2013.
- Ocean Optics conference à Glasgow (Royaume-Uni) en Octobre 2012.
- Ocean Sciences Meeting à Salt Lake City (Etats-Unis) en Fevrier 2012.
- 44th International Liège Colloquium on Ocean Dynamics à Liège (Belgique) en Mai 2012.
- Séminaire organisé par l'association « ADOC SEMINAIRES» à Lille en Janvier 2012.
- La journée des doctorants Irepe à Lille en Mars 2012.
- La journée des doctorants du LOG à Wimereux en Mars 2012 et Septembre 2013.

Table des matières

Chapitre 1 : Introduction générale	1
1.1. Diversité du phytoplancton et empreinte sur les cycles biogéochimiques	1
1.1.1. Importance du compartiment phytoplanctonique dans l'écosystème marin.....	1
1.1.2. Classification du phytoplancton.....	2
1.2. La couleur de l'océan : un outil pertinent pour l'observation du phytoplancton depuis l'espace	6
1.2.1. Généralités	6
1.2.2. La couleur de l'océan : spécificités et principes de mesure	10
1.2.2.1. Paramètres mesurés.....	10
1.2.2.2. Différences entre eaux dites du « Cas1 » et du « Cas2 ».....	12
1.2.2.3. Corrections atmosphériques et algorithmes bio-optiques.....	12
1.3. Apport de la couleur de l'océan à la détection des communautés phytoplanctoniques	14
1.3.1. Présentation générale des approches disponibles.....	14
1.3.2. La méthode PHYSAT : principe général et limitations	17
1.4. Plan et objectifs de la thèse	18
 Chapitre 2 : La méthode PHYSAT : Approche détaillée	 20
2.1. Introduction	20
2.2. Variabilité de second ordre des luminances marines normalisées : les anomalies spectrales	20

2.3. Identification des groupes de phytoplancton à partir des inventaires pigmentaires	22
2.3.1. Diversité des approches disponibles pour l'exploitation des inventaires pigmentaires	22
2.3.2. Identification des groupes de phytoplancton à partir des mesures in situ	23
2.4. Relation entre les anomalies spectrales et la dominance des groupes de phytoplancton, définition des signatures spectrales des groupes phytoplanctoniques	26
2.4.1. Anomalies spectrales associés aux groupes dominants identifiés dans les inventaires de pigments	26
2.4.2. Liens entre les anomalies spectrales et l'environnement bio-optique	30
2.5. Limitations de PHYSAT	31

Chapitre 3: Les cartes auto-organisatrices : description détaillée de l'algorithme et application à la classification automatique des anomalies spectrales

3.1. Réseaux de neurones artificiels	34
3.2. Cartes auto-organisatrices	35
3.2.1. Généralités	35
3.2.2. Spécificités et adaptabilité au traitement des données satellite	36
3.2.3. L'algorithme de SOM	37
3.2.3.1. Architecture neuronale et notion de voisinage	38
3.2.3.2. Apprentissage	40
3.2.4. Evaluation de la qualité de l'apprentissage.....	45
3.2.4.1. L'erreur de quantification vectorielle	45
3.2.4.2. L'erreur topologique.....	46

3.2.4.2. La distorsion	47
3.3. Exploitation de la carte topologique en fin d'apprentissage	48
3.4. Application de SOM à la classification automatique des anomalies spectrales	51
3.4.1. Echantillonnage des données d'entrée et apprentissage de SOM	51
3.4.2. Variabilité des anomalies spectrales à l'échelle globale	53
3.4.3. Amélioration de la caractérisation de la variabilité des anomalies spectrales avec SOM par rapport à l'approche actuelle de PHYSAT	56
Chapitre 4: Application de SOM à la télédétection des groupes phytoplanctoniques à l'échelle globale	60
4.1. Description de la base d'inventaires pigmentaires utilisée	60
4.2. Identification des groupes de phytoplancton et procédure de labellisation des anomalies spectrales.....	62
4.2.1. Identification des groupes de phytoplancton à partir des données <i>in situ</i>	62
4.2.2. Procédure de labellisation des spectres de Ra classés avec SOM	63
4.3. Analyse des formes et amplitudes des spectres associés aux groupes phytoplanctoniques.....	67
4.4. Validation des cartes journalières de distribution des groupes de phytoplancton obtenues avec PHYSAT-SOM	71
4.5. Observations globales et cycles saisonniers.....	74
4.6. Augmentation de la couverture spatiale avec SOM	82
4.7. Conclusion	85

Chapitre 5: Distribution spatio-temporelle des signatures spectrales associées aux diatomées et amélioration de la détection de leurs blooms avec PHYSAT-SOM 87

5.1. Introduction	87
5.2. Distribution spatio-temporelle des différentes signatures spectrales des diatomées	89
5.2.1. Distribution spatiale des Ra et relations avec la composition phytoplanctonique et les propriétés optiques inhérentes du milieu	89
5.2.2. Variabilité temporelle des anomalies spectrales dans l’Océan Austral	98
5.3. Amélioration de la détection des blooms de diatomées dans des conditions spécifiques avec PHYSAT-SOM	102
5.3.1. Amélioration de la détection des blooms de diatomées dans les upwellings côtiers	102
5.3.2. Blooms de diatomées succédant à un évènement climatique	104
5.4. Conclusion	109

Chapitre 6: Application de PHYSAT-SOM dans les eaux du Cas 2 de la Manche et Mer du Nord : première approche et perspectives 111

6.1. Introduction	111
6.2. La Manche et la Mer du Nord	112
6.2.1. Contexte hydrographique	112
6.2.2. L’étude du phytoplancton en Manche orientale et sud de la Mer du Nord	113
6.3. Méthodologie suivie pour la classification des anomalies spectrales en Manche et Mer du Nord	114
6.3.1. Données satellites	114
6.3.2. Classification préliminaire des eaux	115
6.3.3. Calcul des anomalies spectrales relatives à chaque classe d’eau	117

6.3.4. Application des cartes auto-organisatrices séparément dans chaque classe d'eau .	120
6.4. Variabilité spatio-temporelle des anomalies spectrales en Manche et Mer du Nord : résultats préliminaires	123
6.5. Perspectives	128
Conclusions et perspectives	132
Bibliographie	139

CHAPITRE 1

INTRODUCTION GENERALE

1.1. Diversité du phytoplancton et empreinte sur les cycles biogéochimiques

1.1.1. Importance du compartiment phytoplanctonique dans l'écosystème marin

Le phytoplancton représente un compartiment clé dans l'écosystème marin, de par son positionnement à la base du réseau trophique. Il joue un rôle essentiel dans l'équilibre du climat, notamment en pompant le dioxyde de carbone (CO₂) atmosphérique pour réaliser la photosynthèse ou dans l'équilibre acido-basique des océans (Longhurst et al., 1995 ; Claustre et al., 1997 ; Behrenfeld et al., 2002 ; Hashimoto et Shiomoto, 2002). Le phytoplancton, qui ne représente que 1 % de la biomasse (masse totale des organismes vivants mesurée dans une population, une aire ou une autre unité) des organismes photosynthétiques sur la planète, constitue pourtant un puits biologique de gaz carbonique atmosphérique dans les océans. Il assure environ 45 % de la production primaire totale (Claustre et al., 1997 ; Uitz et al., 2006). Son développement est fortement saisonnier et souvent constitué de successions de populations dominantes (majoritairement présentes par rapport à d'autres populations) (Alvain et al., 2008). L'identification de ces différentes populations, et la connaissance de leurs mécanismes de successions, représentent des enjeux majeurs pour les océanographes et les biogéochimistes. Par ailleurs, l'efficacité avec laquelle le carbone est fixé (Claustre *et al.*, 1997 ; Hashimoto et Shiomoto, 2002) ainsi que les voies vers lesquelles ce carbone sera dirigé (Michaels et Silver, 1988) sont étroitement liées aux groupes phytoplanctoniques mis en jeu et à leur taille. De nombreuses recherches, actuellement en cours, soupçonnent l'existence d'un rétrocontrôle spécifique des populations de poissons pélagiques par les niveaux trophiques inférieurs (contrôle appelé « bottom-up »), en lien avec les différents groupes ou espèces phytoplanctoniques (Frederiksen et al., 2006 ; Sinistro, 2010). C'est-à-dire que des groupes de poissons spécifiques pourraient être attirés par une population de phytoplancton spécifique. Les enjeux scientifiques et économiques sont donc importants. Aussi, l'identification des

différents groupes de phytoplancton représente un domaine de recherche essentiel pour appréhender ces différents phénomènes, et constitue la thématique principale de cette thèse.

1.1.2. Classification du phytoplancton

Le plancton représente l'ensemble des organismes vivant dans les eaux douces, saumâtres et salée, le plus souvent en suspension, ils sont inaptes à lutter contre le courant (Hensen, 1887). Ils sont communément divisés en deux classes : le plancton animal ou zooplancton et le plancton végétal ou phytoplancton. Plusieurs approches existent pour la classification du plancton. Sieburth et al. (1978) ont introduit une classification basée sur trois classes de taille : le microplancton (20-200 μm), le nanoplancton (2-20 μm) et le picoplancton (0.2-2 μm). Cette classification constitue un outil pertinent pour décrire de façon simple les communautés phytoplanctoniques. L'influence de la taille sur la physiologie du phytoplancton a bien été confirmée dans le cadre de plusieurs études (Chisholm, 1992 ; Jassby et Platt, 1976 ; Raven, 1998). La variabilité de quelques fonctions biogéochimiques peut être abordée avec cette approche. Par exemple, compte tenu de leur ratio surface/volume élevé, les cellules picophytoplanctoniques (e.g. *Prochlorococcus* et *Synechococcus*) absorbent plus efficacement les nutriments dans les milieux oligotrophes (limités en nutriments), ce qui les rend généralement dominants au niveau de ces zones (Dandonneau et al., 2004 ; Longhurst, 2007). Inversement, le microphytoplancton (e.g. diatomées, dinoflagellés) dominant généralement les eaux riches en nutriments (hautes latitudes, upwellings) et représentent le principal facteur d'exportation du carbone vers les eaux profondes vu leur taille qui leur procure une plus grande vitesse de sédimentation par rapport aux autres classes de taille (Smetacek et al., 1990 ; Tremblay et al., 2002).

Cette organisation des groupes présente l'avantage d'être basée sur un critère relativement simple, mais a pour inconvénient de ne pas tenir compte des caractéristiques intrinsèques des groupes. Une approche basée sur les classes de taille peut en effet conduire à un échec de distinction de certaines fonctions biogéochimiques, si des groupes phytoplanctoniques caractérisés par différentes fonctions appartiennent à une même classe de taille (Le Quéré et al., 2005 ; Nair et al., 2008). Par exemple, les producteurs de diméthylsulfure (DMS) (e.g. *Phaeocystis*) et les calcifiants (e.g. Coccolithophores) appartiennent tous les deux au nanophytoplancton et à l'embranchement des haptophytes alors qu'ils agissent différemment d'un point de vue biogéochimique. En effet, les producteurs de

DMS ont la capacité de former de nouveaux aérosols (aérosols de sulfite) qui, dans certaines conditions, peuvent se comporter comme des noyaux de condensation de nuages (Belviso et al., 2004). Ces noyaux de condensation rétrodiffusent alors les rayons solaires et contribuent ainsi au refroidissement de l'atmosphère (Malin et al., 1992). En revanche, les organismes calcifiants engendrent une diminution de l'alcalinité des eaux de surface (via la production du carbonate de calcium composant leur squelette externe) et par conséquent, une production de CO₂ qui sera libéré dans l'atmosphère (Belviso et al., 2004 ; Le Quéré et al., 2005 ; Nair et al., 2008). Ce gaz à effet de serre va donc participer au réchauffement de l'atmosphère. Cet effet biogéochimique antagoniste engendré par une même classe de taille de phytoplancton représente un exemple, parmi d'autres, qui montre l'importance de considérer les fonctions biogéochimiques du phytoplancton pour sa classification.

Cette approche de classification consiste donc à regrouper le phytoplancton par groupes fonctionnels (ou PFT pour Phytoplancton Functional Type), en se basant sur leurs processus biologiques de croissance et de mortalité (Claustre, 1994 ; Falkowski et al., 2004, Bouman et al., 2003, Le Quéré et al., 2005) ainsi que leurs fonctions biogéochimiques au sein de l'écosystème. Même si cette tâche n'est pas aisée, l'objectif est de tendre vers cela en utilisant différentes techniques, basées, par exemple, sur l'étude des différentes compositions en pigments. Ces derniers, appelés également pigments photosynthétiques ou assimilateurs, représentent des composés chimiques permettant la transformation de l'énergie lumineuse en énergie chimique chez les organismes effectuant la photosynthèse. Lorsqu'un photon (particule élémentaire de lumière) heurte une molécule de pigment photosynthétique, son énergie excite un atome de cette molécule, et la fait passer à un état excité, de niveau énergétique élevé. L'énergie accumulée dans la molécule de pigment est libérée lors du retour à l'état fondamental de la molécule. Selon leur mode de transmission de l'énergie lumineuse accumulée (fluorescence, transmission de l'état d'excitation et conversion de l'énergie qui consiste en la modification de la structure chimique du pigment ou d'une molécule voisine) et leur réaction vis-à-vis de cette lumière (absorption plus ou moins importante à certaines longueurs d'ondes), on distingue deux grands types de pigments photosynthétiques :

- Les pigments actifs : capables d'effectuer les 3 modes de libération de l'énergie lumineuse accumulée. On trouve dans cette catégories la chlorophylle a et b et les bactériochlorophylles (pigments photosynthétiques présents chez différentes bactéries autotrophes).

- Les pigments accessoires : incapables d'effectuer la conversion de l'énergie. On trouve principalement dans cette catégorie les caroténoïdes, les xanthophylles et les phycobilines.

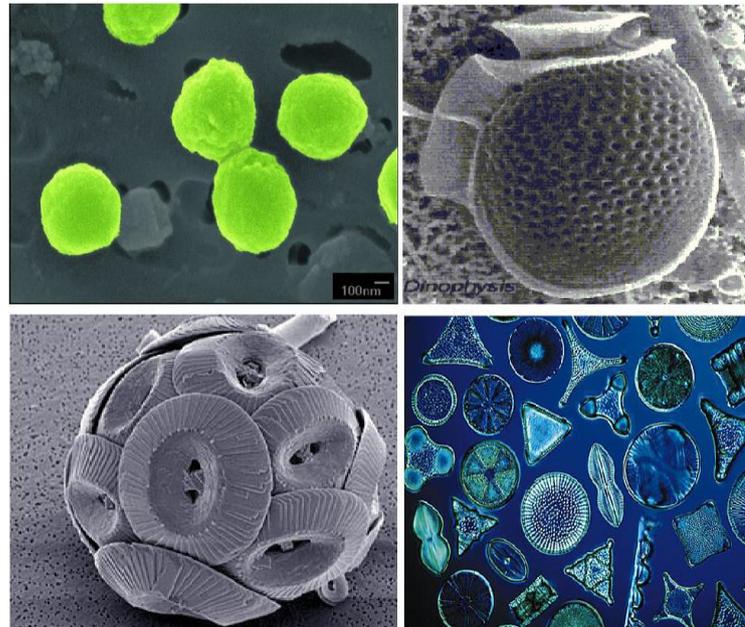
Le tableau 1 liste quelques exemples de groupes fonctionnels et les processus biogéochimiques dans lesquels ils sont impliqués.

Tableau 1.1 : Les grands groupes fonctionnels phytoplanctoniques et processus associés (d'après Le Quéré et al., 2005)

Groupes fonctionnels	Processus particuliers
Fixateurs d'azote (e.g. <i>Trichodesmium</i>)	<ul style="list-style-type: none"> - Utilisation de l'azote atmosphérique en tant que nutriment pour la croissance - Impact direct sur le cycle de l'azote et sur d'autres facteurs qui agissent sur le changement climatique.
Phytoplancton siliciants (e.g. Diatomées)	<ul style="list-style-type: none"> - Utilisation de la silice pour la formation du squelette externe (frustules). Le squelette siliceux procure une plus grande densité et permet donc un export de carbone plus efficace vers les couches profondes. - Participation importante à la production primaire.
Phytoplancton calcifiants (e.g. coccolithophoridés)	<ul style="list-style-type: none"> - Utilisation du calcaire pour la formation du squelette externe (coccolithes). - La formation de carbonate de calcium engendre la production de CO₂ et son expulsion vers l'atmosphère. - Le carbonate de calcium constitue une source potentielle de carbone inorganique particulière qui est exporté vers les couches profondes.
Producteurs de DMS (e.g. <i>Phaeocystis</i>)	<ul style="list-style-type: none"> - Le DMS influe sur le climat terrestre à travers la formation d'aérosols sulfurés. Ces derniers vont contribuer à l'équilibre radiatif global en agissant en tant que noyaux de condensation de nuages. Le DMS contribue aussi à la formation de pluie acide.

Au-delà des exemples présentés ci-dessus, il existe une grande diversité de groupes fonctionnels, de tailles, de formes et de préférences trophiques différentes, dont quelques exemples sont affichés sur la figure 1.1. Dans la suite de ce manuscrit, nous tendrons vers cette catégorisation des différents ensembles de phytoplancton. Cependant, nous utiliserons le

terme « groupe » plutôt que PFT pour caractériser plus aisément les sous-ensembles de phytoplancton dont il sera question. Il est toutefois important de signaler que le terme « groupe » a un sens bien précis dans l'arborescence taxinomique qui ne correspond pas littéralement à celui que nous lui donnerons par la suite. Certains des groupes, ainsi défini plus en détail ci-dessus, sont en fait, d'un point de vue taxinomique, des classes ou des genres.



*Figure 1.1. Exemples de différentes formes de groupes de phytoplancton. De haut en bas et de gauche à droite, une cyanobactérie (*Prochlorococcus*, <http://proportal.mit.edu/>), un genre de dinoflagellés (*Dinophysis*, <http://www.u-picardie.fr>), un coccolithophore (*Coccolithus Pelagicus*, <http://fr.wikipedia.org/wiki/Coccosphaerales>) et différentes espèces de diatomées (<http://www.larousse.fr/encyclopedie/>). A noter que les figures ne sont pas à la même échelle.*

D'un point de vue pigmentaire, les organismes photoautotrophes contiennent tous de la chlorophylle *a* (à l'exception des prochlorophytes, chez lesquels elle est remplacée par la divinyl-chlorophylle *a*) ainsi que des pigments accessoires. Contrairement à la chlorophylle *a*, les pigments accessoires varient d'un groupe phytoplanctonique à l'autre et peuvent servir de biomarqueurs de groupes taxinomiques (Jeffrey et Vesk, 1997). Les pigments biomarqueurs sont aussi appelés pigments diagnostiques. L'intérêt taxinomique réside dans le fait que la majorité des groupes phytoplanctoniques contient un ou plusieurs pigments diagnostiques "spécifiques". La détection de ces pigments permet donc l'identification du groupe auquel ils appartiennent. La méthode HPLC (High Performance Liquid Chromatography) permet d'identifier et de quantifier l'ensemble des pigments donnant accès à la composition taxinomique des communautés phytoplanctoniques. Cette méthode de spéciation a été

largement utilisée dans de précédentes études portant sur l'identification des groupes phytoplanctoniques (e.g. Mackey et al., 1996 ; Vidussi et al., 2001). Bien que cette technique soit peu sensible quant à la détection de certains composés, tels que les substances coéluées (substances adsorbées simultanément), elle permet d'avoir des résultats facilement exploitables et très précis (de l'ordre du picogramme). La méthode HPLC permet aussi d'éviter la surestimation des phaeopigments et la sous-estimation de la chlorophylle *a*, tel que c'est le cas avec d'autres méthodes quand la chlorophylle *b* est présente (Gibbs, 1979 ; Mantoura et al., 1997 ; Uitz et al., 2006).

Il existe différentes méthodes d'observation du phytoplancton et d'estimation de sa biomasse, à travers la mesure de la concentration en chlorophylle *a*, avec, en premier lieu, les mesures *in situ*. Parmi ces dernières, on peut citer celles réalisées dans le cadre de différents programmes, tels que JGOFS (Joint Global Ocean Flux Study), GeP&CO (Geochemistry, Phytoplankton and Color of the Ocean) ou AMT (Atlantic Meridional Transect). Ces dernières sont essentielles mais relativement limitées dans l'espace et dans le temps au regard des échelles spatio-temporelles caractérisant les efflorescences phytoplanctoniques (courtes périodes de blooms et efflorescences locales pour certains groupes). L'élaboration de bases de données représentatives de l'océan mondial est donc nécessaire afin de passer à un niveau d'intégration globale des informations disponibles. C'est ce qui, par exemple, réalisé dans le programme qui permet de regrouper les mesures *in situ* disponibles dans la base de données NOMAD (NASA bio-Optical Marine Algorithm Data set, <http://seabass.gsfc.nasa.gov/>). Dans ce contexte, l'utilisation des capteurs spatiaux, embarqués à bord de satellites, constitue une des méthodes complémentaires pouvant être efficaces pour caractériser la distribution spatiale et temporelle des principaux groupes de phytoplancton.

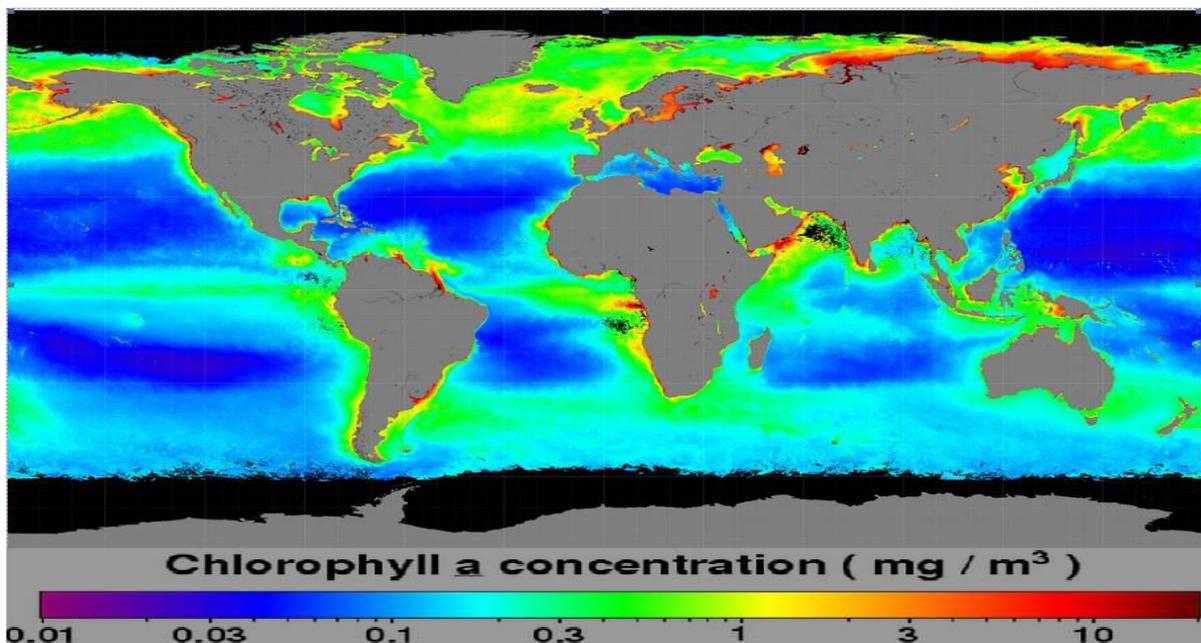
1.2. La couleur de l'océan : un outil pertinent pour l'observation du phytoplancton depuis l'espace

1.2.1. Généralités

Depuis l'espace, il est possible d'observer une large gamme de paramètres utiles à l'étude et à la compréhension du fonctionnement du système océanique. Cela va de la détermination de la température de surface, du niveau de la mer, de la vitesse des courants,

des vents ou de ce qui va nous intéresser particulièrement par la suite : la « couleur de l’océan ». Cette dernière prend en compte les variations spectrales de la lumière diffusée par la couche supérieure de la surface de l’eau. Celle-ci varie d’un bleu profond en eaux pures, à des verts sombres dans les eaux riches en phytoplancton ou à des verts "laiteux" dans les eaux riches en sédiments, par exemple. De même, des variations de la couleur de la mer, allant du bleu par beau temps au gris sous un épais couvert nuageux sont observables à l’œil nu. Cet aspect rend surtout compte de la réflexion de l’éclairement solaire et de la voûte céleste sur la surface de la mer. En règle générale, l’eau de mer nous paraît bleue à l’œil nu, en raison de sa forte absorption dans les longueurs d’onde vertes et au-delà.

La télédétection de la “couleur de l’océan” permet d’estimer, depuis l’espace, de nombreux paramètres, dont le contenu en chlorophylle des eaux océaniques de surface. La variation des nuances de couleurs observées depuis l’espace sont dues, au premier ordre, aux interactions de la lumière incidente avec l’eau, les cellules phytoplanctoniques ainsi que d’autres composés organiques. Tous ces éléments engendrent des variations spectrales de la lumière réfléchi par les eaux de surface. Ces variations nous renseignent donc indirectement sur le contenu de l’eau, permettant ainsi d’estimer différents paramètres, dont la concentration en chlorophylle *a* (Chl*a*) qui donne une information sur la biomasse phytoplanctonique présente dans l’eau (Figure 1.2).



*Figure 1.2. Carte de concentration moyenne en chlorophylle *a* ($\text{mg}\cdot\text{m}^{-3}$) en Aout, réalisée à partir des données du capteur SeaWiFS sur une climatologie de 1998 à 2010 (synthèse pluriannuelle)*
(<http://oceancolor.gsfc.nasa.gov/>)

La fin des années 1970 a vu l'apparition d'un nouvel outil permettant d'observer la couleur de l'océan dans de nombreuses gammes spectrales (e.g. visible, infrarouge) avec une résolution spatiale élevée (de quelques kilomètres), une couverture pratiquement globale (environ 80% de la superficie de la Terre) et une résolution temporelle haute fréquence (pratiquement journalière) : les techniques de mesure par satellite de la « couleur de l'océan ». Ces techniques ont vu le jour en 1978 avec le lancement du capteur CZCS (Coastal Zone Color Scanner). Les mesures se sont poursuivies depuis, avec toutefois une interruption assez longue entre 1986 et 1996, jusqu'à l'arrivée de capteurs plus récents comme POLDER (POLarization and Directionality of the Earth's Reflectances), SeaWiFS (Seaviewing Wide Field of view Sensor), MODIS (Moderate Resolution Imaging Spectroradiometer), MERIS (MEDium Resolution Imaging Spectrometer) et plus récemment VIIRS (Visible Infrared Imager Radiometer Suite). Le tableau 1.2 liste quelques capteurs de couleur de l'océan effectuant ou ayant effectué des observations à l'échelle globale.

Tableau 1.2. Liste de quelques capteurs de couleur de l'océan effectuant ou ayant effectué des observations à l'échelle globale avec leurs périodes d'observations respectives, résolution de base et nombre de bandes respectifs (nombre de longueurs d'onde)

Capteur	Période d'observation	Résolution de base (m)	Nombre de bandes
CZCS (Coastal Zone Color Scanner)	10/1978 - 06/1986	825	6
POLDER (POLarization and Directionality of the Earth's Reflectances)	08/1996 - 07/1997	6000	9
POLDER 2	12/2002 - 10/2003	6000	9
POLDER 3	12/2004 - Présent	6000	9
SeaWiFS (Seaviewing Wide Field of view Sensor)	09/1997 - 12/2010	1100	8
MERIS (MEDium Resolution Imaging Spectrometer)	03/2002 - 04/2011	300 / 1200	15
MODIS-Aqua (MODerate Resolution Imaging Spectroradiometer)	05/2002 - Présent	1000	36
GOCI (Geostationary Ocean Color Imager)	06/2010 - Présent	500	8
VIIRS (Visible Infrared Imager Radiometer Suite)	10/2011 - Présent	370 / 740	22

D'une façon générale, pour tous les capteurs spatiaux utilisés en couleur de l'océan, les produits opérationnels (images) sont disponibles à différents niveaux de traitement :

- Niveau 1 : comptes numériques correspondant aux réflectances brutes mesurées par le capteur dans chaque canal, avec les données de télémétrie, calibration et navigation associées. Ces données sont disponibles en mode LAC (« Local Area Coverage », pleine résolution, 1 km) ou en mode GAC (« Global Area Coverage »). Dans ce mode les données sont échantillonnées à partir des données LAC, à raison d'un pixel sur 4 (résolution réduite, 4 km/pixel).
- Niveau 2 : Paramètres géophysiques issus du niveau 1 par application des facteurs de calibration, des corrections atmosphériques et des différents algorithmes bio-optiques. La résolution de ces produits est de 4 km (GAC) et de 1 km en LAC.
- Niveau 3 « binned » : données GAC de niveau 2, moyennées sur 1 jour, 8 jours, 1 mois ou une année et projetées sur une grille avec des pixels d'égale surface, de 81 km² environ (pixel de 9 km x 9 km). C'est ce type de produit qui a été utilisé pour développer la méthode présentée dans ce manuscrit. D'autres résolutions existent. Par exemple, pour le capteur MODIS, ces produits sont aussi disponibles à 4 km de résolution.
- Niveau 3 « mapped » ou SMI (Standard Mapped Image) : ces produits sont des cartes, au format image, de synthèse de paramètres sur une projection équi-cylindrique du globe. La résolution de ces produits est, en général, de 9,2 km (à l'équateur) même si des cartes à d'autres résolutions existent.

Les images acquises par les capteurs de la couleur de l'océan ont permis d'envisager plusieurs études à diverses échelles d'observation, allant de l'échelle régionale jusqu'à celle des bassins océaniques et de l'océan mondial. Par exemple, les données acquises ont permis d'analyser le taux de fixation du carbone dans l'océan via des modèles de production primaire (photosynthèse), la quantification de l'échange du gaz carbonique à l'interface océan-atmosphère ou encore du flux de carbone organique quittant la couche océanique superficielle (Bidigare et al., 1992 ; Behrenfeld et Falkowski, 2004; Campbell et al., 2002). Ces modèles calculent un taux de fixation de carbone qui dépend de l'éclairement disponible pour la photosynthèse (PAR), de Chl_a mais aussi de paramètres photophysiques (i.e. l'absorption spécifique du phytoplancton).

Compte tenu de la très forte variabilité spatio-temporelle des processus physiques, chimiques et biologiques régulant la composition, la distribution et l'activité du phytoplancton (i.e. mélange vertical, flux de nutriments, sédimentation, broutage), la couleur de l'océan se présente comme un outil particulièrement adapté pour l'étude de l'évolution de la biomasse phytoplanctonique. Néanmoins, l'observation par les satellites présente également ses limites. La présence de nuages représente l'obstacle majeur. Par exemple, pour le capteur SeaWiFS, il fallait environ une semaine pour obtenir une couverture globale en raison de la présence importante de nuages qui empêche l'observation de la surface sous-jacente. D'autres obstacles liés, par exemple, à la présence d'aérosols atmosphériques ou à certains défauts inhérents aux capteurs (défauts radiométriques ou rapport signal / bruit faible) peuvent significativement limiter les observations. Pour ces diverses raisons, des algorithmes prenant en compte ces limites ont vu le jour, permettant à l'observation de la couleur de l'océan de devenir un outil précieux en océanographie (Ruddick et al., 2000 ; Stumpf et al., 2003 ; Brajard et al., 2006 ; Kuchinke et al., 2009 ; Bailey et al., 2010).

1.2.2. La couleur de l'océan : spécificités et principes de mesure

1.2.2.1. Paramètres mesurés

Le paramètre océanographique « couleur de l'eau » sera plus précisément défini ici comme étant les variations spectrales (variations avec la longueur d'onde considérée, notée λ) de la réflectance diffuse des couches océaniques supérieures, notée $R(\lambda)$. La réflectance diffuse $R(\lambda)$ représente le rapport entre l'éclairement ascendant qui émerge juste sous la surface de l'eau, $E_u(\lambda)$, à l'éclairement descendant y pénétrant, $E_d(\lambda)$.

$$R(\lambda) = E_u(\lambda) / E_d(\lambda) \quad \text{(Eq. 1.1)}$$

Notons que $E_u(\lambda)$ correspond à l'intégrale des « luminances marines » ou « L_w » (tenant compte de leurs angles zénithaux). L_w correspond à l'énergie lumineuse émise juste au-dessus de l'interface air-mer dans une direction donnée. Cette grandeur dépend donc de la géométrie de l'observation (c'est une grandeur anisotrope, c'est-à-dire, présentant des variations en fonction de la direction). Le paramètre standard utilisé en couleur de l'océan se nomme « réflectance de télédétection » ou R_{rs} (pour « remote sensing reflectance ») et s'exprime suivant cette équation :

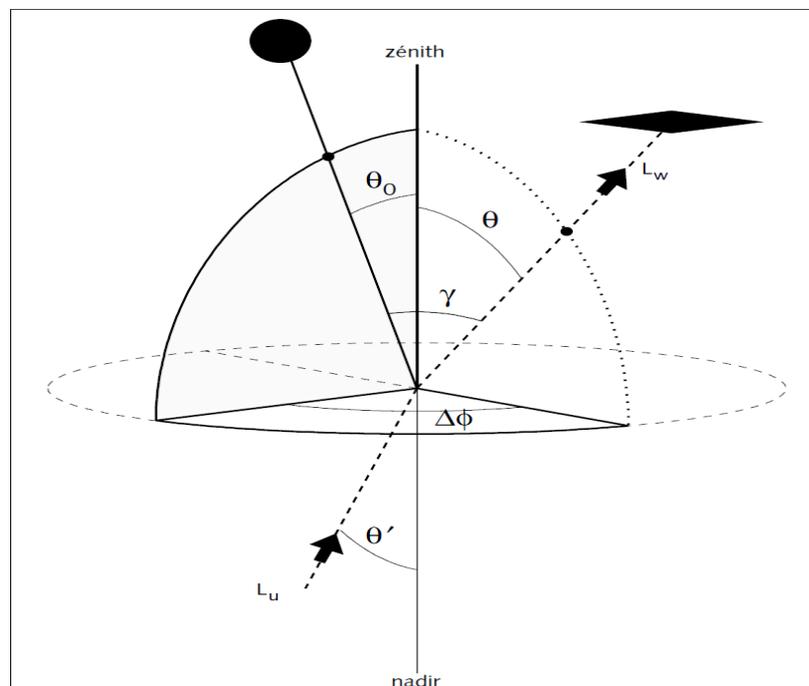
$$Rrs(\lambda;0+) = Lw(\lambda;0+;\theta, \Phi) / E_d(\lambda;0+) \quad (\text{Eq. 1.2})$$

Avec $Lw(\lambda;0+;\theta, \Phi)$ la luminance sortante juste au-dessus de l'interface pour une longueur d'onde donnée et dans la direction (θ, Φ) et $E_d(\lambda;0+)$ l'éclairement au-dessus de la surface pour une longueur d'onde donnée (les différentes directions et angles d'observations sont définis dans la Figure 1.3).

La Rrs présente l'avantage d'être affranchie de la géométrie de l'observation. En effet, ce paramètre est supposé avoir les mêmes valeurs dans toutes les directions (c'est une grandeur isotrope, c'est-à-dire qu'elle est homogène dans toutes les directions). C'est aussi le cas des luminances marines dite normalisées, ou nLw , qui représentent les luminances telles qu'elles seraient observées si le satellite faisait toutes ses mesures au nadir, sans atmosphère et avec le soleil au zénith. Cette grandeur est définie par :

$$nLw(\lambda) = Rrs(0+;\theta, \Phi) / F_0 \quad (\text{Eq. 1.3})$$

Avec $F_0(\lambda)$ l'éclairement solaire incident au sommet de l'atmosphère, dont les valeurs sont bien connues.



*Figure 1.3. Représentation schématique de la géométrie d'une observation satellitaire.
D'après Morel et Gentili (1996).*

1.2.2.2. Différences entre eaux dites du « Cas 1 » et du « Cas 2 »

Il existe dans l'océan ouvert toute une variété de constituants (phytoplancton, détritiques associés, matières dissoutes) qui vont modifier la manière dont la lumière va se propager dans la couche de surface des océans. Il suffit de savoir que les variations de Rrs sont dues aux diverses substances dissoutes ou particules en suspension dans l'eau, pour comprendre alors que, par des méthodes d'inversion, on pourra obtenir des informations sur les propriétés ou les concentrations de ces divers « additifs » à l'eau de mer pure à partir des mesures de $Rrs(\lambda)$ (mesures à des longueurs d'onde discrètes ou bien sur tout le spectre visible, de 400 nm à 700 nm environ). Précisons que les causes de variations de la couleur de l'océan dépendent du type d'eau considéré. Il est, en effet, actuellement admis que les eaux peuvent être séparées en deux catégories :

i) Les eaux dites du Cas 1 (Morel et Prieur, 1977) : en général, ce sont les eaux « du large », non affectées par les apports fluviaux et plus généralement par toute influence de la côte ou du fond. Dans ces eaux, qui représentent environ 95% de l'océan mondial, les propriétés optiques sont déterminées, d'une part, par celles de l'eau de mer pure, et d'autre part par celles du phytoplancton et de l'ensemble de ses produits associés (tels que les détritiques organiques particuliers ou dissous).

ii) Les eaux dites du Cas 2, où les propriétés optiques dépendent en partie des mêmes composantes que dans les eaux du Cas 1, mais aussi d'autres substances actives d'un point de vue optique, comme les sédiments en suspension ou les substances organiques dissoutes (appelées aussi « matières jaunes ») provenant du lessivage des terres. Les multiples combinaisons possibles entre phytoplancton, particules et substances dissoutes exogènes (c'est-à-dire non produites par l'activité phytoplanctonique) génèrent une multitude de possibilités pour le spectre de la réflectance. L'inversion de ce signal pour estimer les concentrations des diverses composantes de ces eaux devient alors difficile. L'exploration des propriétés optiques des eaux du Cas 2 et la recherche d'algorithmes permettant d'inverser le spectre de réflectance de ces eaux sont en fort développement actuellement, dans la mesure où l'intérêt pour ces eaux côtières, dont l'importance économique est de plus en plus ressentie, ne cesse de croître.

1.2.2.3. Corrections atmosphériques et algorithmes bio-optiques

Les capteurs dédiés à la mesure de la réflectance marine sont embarqués à bord de satellites en orbite autour de la terre, se trouvant au-dessus de l'atmosphère terrestre. Le capteur observe donc la surface des océans à travers l'atmosphère. Celle-ci représente un écran hautement diffusant qui interagit avec le rayonnement de façon importante. En réalité, la part du signal marin sur le signal total au sommet de l'atmosphère n'excède pas 10%. Dans le domaine du visible (λ entre 400 et 700 nm), le reste du signal provient des interactions de la lumière avec les molécules de l'atmosphère (eau, gaz) et les aérosols (Gordon et al., 1997). Pour obtenir les nLW à partir des réflectances mesurées par le capteur, il est essentiel d'estimer la contribution atmosphérique pour être en mesure de déduire du signal total mesuré la part du signal marin. C'est ce que l'on appelle les corrections atmosphériques (Gordon et Wang, 1994 ; Gordon, 1997 ; Gordon et al., 1997, Stumpf et al., 2003). Si la contribution moléculaire est assez bien connue, celle des aérosols est plus délicate à estimer. Des algorithmes ont été et sont toujours développés dans ce sens en utilisant généralement des mesures effectuées dans le proche infrarouge, là où le signal marin est nul ou faible (Gordon et Wang, 1994, Gordon, 1997). Dans ces longueurs d'onde le signal mesuré est donc essentiellement atmosphérique, il est alors possible de quantifier la contribution des aérosols à partir de la dépendance spectrale de ce signal et d'en corriger l'effet en extrapolant vers les plus courtes longueurs d'onde. Précisons que cette hypothèse de « signal nul » dans le proche infrarouge n'est pas valide dans certaines conditions, essentiellement lorsque la concentration en chlorophylle a est relativement élevée (Siegel et al., 2000). Une fois ces corrections effectuées, on utilise des modèles bio-optiques (reliant les propriétés optiques et les paramètres biologiques) permettant de retrouver la $Chla$. Ces modèles consistent à établir des relations empiriques entre un jeu de données de mesures *in situ* de $Chla$ et les données radiométriques mesurées. Sachant que la chlorophylle a présente un maximum d'absorption vers 440 nm et son minimum vers 550 nm, le rapport des réflectances à ces deux longueurs d'onde apporte une information sur le contenu de l'eau en chlorophylle a (Morel, 1988). L'algorithme standard actuellement utilisé par la NASA pour le traitement des données satellite SeaWiFS et le calcul de la $Chla$ s'appelle OC4v6 (O'Reilly et al., 2000). Il est basé sur le rapport entre le maximum des Rrs à 443, 490 et 510 nm sur la valeur de Rrs à 555 nm (O'Reilly et al., 1998, 2000). Ce rapport « bleu/vert » est relié à la concentration en chlorophylle a par un polynôme logarithmique d'ordre 4 :

$$Chla = 0,3272 - 2,9940\alpha + 2,7218\alpha^2 - 1,2259\alpha^3 - 0,5683\alpha^4 \quad (\text{Eq. 1.4})$$

où $\alpha = \log_{10}(Rrs)$ et $Rrs = \max(Rrs(443, 490, 510)) / Rrs(555)$

Notons que les coefficients de l'équation 1.4 sont utilisés pour le capteur SeaWiFS. Ils changent généralement d'un capteur à un autre, en fonction des longueurs d'onde prises en compte. Par exemple, pour le capteur MODIS-Aqua, la longueur d'onde 490 nm de SeaWiFS est remplacée par 488 nm, 510 nm par 531 nm et 555 nm par 551 nm et des coefficients différents sont donc utilisés. Par ailleurs, il existe également des algorithmes dits semi-analytiques qui mêlent des approches physiques à des relations empiriques (Garver et Siegel, 1997 ; Maritorena et al., 2002). Ils permettent de mieux comprendre les causes de variations de la couleur de l'eau, mais sont moins précis que les algorithmes empiriques pour des applications comme la quantification du contenu en chlorophylle *a* (O'Reilly et al., 1998).

1.3. Apports de la couleur de l'océan à la détection des communautés phytoplanctoniques

1.3.1. Présentation générale des approches disponibles

L'existence d'une archive de plus de 30 ans de cartes globales de concentration en chlorophylle *a* (avec toutefois une interruption assez longue entre 1986 et 1996) a contribué à la réalisation de nombreuses études visant à mieux comprendre la répartition du phytoplancton, son évolution dans le temps et dans l'espace ainsi que ses interactions avec d'autres paramètres biophysiques (Huot et al., 2007 ; Timmermans et al., 2008 ; Martinez et al., 2009, 2011; Bricaud et al., 2012 ; D'Ortenzio et al., 2012; Uitz et al., 2012 ; Siegel et al., 2013). Durant plusieurs années, la communauté « couleur de l'océan » a focalisé son attention sur l'étude de la biomasse phytoplanctonique dans son ensemble, avec l'estimation de la *Chla*, sans pouvoir tenir compte de sa composition taxinomique. Cependant, la chlorophylle étant le pigment commun à tous les groupes de phytoplancton, il ne permet pas, à lui seul, d'entreprendre une distinction taxinomique de ces différents groupes. Des études complémentaires ont donc été menées ces dernières années et ont abouti à différentes approches visant à extraire plus d'informations, au-delà de la seule concentration en chlorophylle-*a* (Brown et Yoder, 1994 ; Sathyendranath et al., 2001, 2004 ; Subramaniam et al., 2002 ; Alvain et al., 2005, 2008 ; Ciotti et Bricaud, 2006 ; Devred et al., 2006 ; Uitz et al.,

2006, Aiken et al., 2007 ; Hirata et al., 2008; Raitzos et al., 2008 ; Bracher et al., 2009 ; Brewin et al., 2010 ; Kostadinov et al., 2009; Pan et al., 2010, 2011). Ces algorithmes, permettant d'identifier ou de quantifier la biomasse spécifique de différents groupes de phytoplancton à partir de données de couleur de l'océan, peuvent être classés en trois approches :

i) Les approches basées sur l'abondance (Sathyendranath et al., 2001; Uitz et al., 2006; Devred et al., 2006; Aiken et al., 2007; Hirata et al., 2008; Pan et al., 2010 et 2011)

ii) Les approches écologiques (Raitzos *et al.*, 2008)

iii) Les approches basées sur les réponses spectrales du phytoplancton (Brown et Yoder, 1994 ; Subramaniam et al., 2002 ; Sathyendranath et al., 2004; Alvain et al., 2005 et 2008; Ciotti et Bricaud, 2006; Bracher et al., 2009; Brewin et al., 2010 ; Uitz et al., 2006).

Les principes des trois approches, leurs avantages et inconvénients sont résumés dans le tableau 1.3.

Tableau 1.3. Spécificités, avantages et inconvénients des 3 approches actuellement disponibles pour la télédétection des groupes de phytoplancton

Type d'approche	Spécificités	Avantages	Inconvénients
Abondance	<ul style="list-style-type: none"> - Relations entre la biomasse phytoplanctonique (Chla ou absorption par le phytoplancton) et la dominance, présence d'un groupe ou d'une classe de taille 	<ul style="list-style-type: none"> - Approches robustes dans les zones caractérisées par la dominance d'un groupe ou classe de taille particulière - Pertinence dans le paramétrage des modèles de production primaire qui produisent des estimations basées sur la biomasse phytoplanctonique 	<ul style="list-style-type: none"> - Efficacité limitée dans les régions où les variations des proportions des différents groupes pourraient confondre le signal recherché. - Possibilité d'échec dans la distinction des blooms de groupes ou de classes de tailles dominants ayant la même biomasse - La variabilité intra-groupe ou intra-spécifique des propriétés optiques (e.g. stades de croissance) peut introduire des erreurs significatives
Ecologique	<ul style="list-style-type: none"> - Prise en compte de paramètres environnementaux (e.g. température, vent, lumière) en plus des paramètres optiques. - Détermination de la dominance de groupes en relation avec ces paramètres et en se basant sur des connaissances sur leurs régimes biologiques dans la région étudiée 	<ul style="list-style-type: none"> - Utilisation d'une information supplémentaire aux paramètres optiques - Peuvent donner des informations sur la variabilité des groupes phytoplanctoniques en relation avec les changements environnementaux et quels facteurs influencent le plus ces changements (Brewin et al., 2011) 	<ul style="list-style-type: none"> - Complexité des approches statistiques utilisées et intrications complexes des différents paramètres - D'un point de vue analytique, l'interprétation des résultats peut être difficile et dépend fortement de la qualité et quantité des données
Réponses spectrales	<ul style="list-style-type: none"> - Relations entre spectres (réflectances, absorption spécifique, luminances marines normalisées, etc.) et dominance des groupes de phytoplancton - Relations entre la dominance de classes de tailles et l'absorption spécifique du phytoplancton 	<ul style="list-style-type: none"> - Peuvent détecter différents groupes de phytoplancton ayant la même biomasse. - Sont efficace pour détecter les groupes ayant des signatures spectrales contrastées avec le milieu environnant (e.g. coccolithophores, <i>Trichodesmium</i> ou diatomées). 	<ul style="list-style-type: none"> - Distinction difficile des groupes ayant des signatures spectrales similaires (Brewin et al., 2011) - Difficile de distinguer les variations des signatures spectrales d'un même groupe ou espèce de phytoplancton se trouvant dans des conditions différentes (différents stades de croissance, disponibilité des nutriments, régimes de lumière, etc.) (Nair et al., 2008).

L'une des méthodes disponibles à ce jour, dénommée PHYSAT (Alvain et al., 2005), faisant partie des approches spectrales, permet de détecter des groupes de phytoplancton à l'échelle globale quand ils sont dominants. La dominance d'un groupe de phytoplancton, telle que nous l'expliquerons par la suite, dépend de la concentration relative (par rapport à celle de la chlorophylle *a*) de son pigment biomarqueur dans un échantillon donné (Alvain et al., 2005). PHYSAT se base sur des signatures optiques spécifiques, dérivées des luminances marines normalisées. Nous nous sommes basées sur le principe de cette approche pour développer la méthode décrite dans le cadre de cette thèse.

1.3.2. La méthode PHYSAT : principe général et limitations

Cette méthode permet de détecter plusieurs groupes de phytoplancton dominants à l'échelle globale dans les eaux de surface du Cas 1. Comme nous l'avons détaillé dans le paragraphe 1.1, il existe une grande diversité de phytoplancton en termes de tailles et de fonctions biogéochimiques. La méthode PHYSAT est basée sur l'exploitation de la variabilité de second ordre des nLw, en lien avec la diversité du phytoplancton. En effet, des anomalies de nLw, pour des valeurs de Chl-*a* données, liée à la présence de groupes spécifiques ont été mises en évidence à partir de deux types de données : des mesures *in situ* de pigments (chlorophylle *a* et autres pigments), collectées dans le cadre de campagnes océanographiques, et des mesures coïncidentes de nLw du capteur SeaWiFS (Alvain et al. 2005, 2008). Les inventaires pigmentaires comportant les principaux biomarqueurs ont permis l'identification de groupes de phytoplancton dominants en se basant sur des critères issus de la littérature (Alvain et al. 2005). Des spectres d'anomalies de luminances, spécifiques en termes d'amplitude et de forme ont été mis en évidence et ont permis la détection de groupes de phytoplancton à l'état de dominance (Nanoeucaryotes, *Prochlorococcus*, *Synechococcus* Like Cyanobacteria ou SLC et diatomées). Ces signatures spectrales ont par la suite été exploitées pour déterminer la distribution des groupes de phytoplancton à l'échelle globale à partir de l'archive d'images satellites. Les résultats ont montré que la distribution globale des groupes était en accord avec les précédentes études ayant analysé la répartition du phytoplancton à l'échelle globale (Alvain et al., 2005 ; Alvain et al., 2008). Par ailleurs, une étude récente a montré de quelle manière les formes et amplitudes de ces spectres étaient reliés aux propriétés optiques du milieu (Alvain et al., 2012). Cette explication physique a ouvert plusieurs perspectives quant à une exploitation possible de ces spectres pour détecter des groupes de phytoplancton dans des situations autres que la dominance (présence du même groupe dans différents types d'environnements et stades de croissance, par exemple).

Bien que PHYSAT ait mis en évidence et confirmé l'existence de signaux spécifiques pour des groupes phytoplanctoniques dominants (Alvain et al., 2005 ; 2008 ; 2012), la méthode restait limitée par l'approche suivie pour déterminer les signatures spectrales de ces groupes, nécessitant la définition d'enveloppes spectrales. Pour cela, il était nécessaire d'introduire des seuils minimaux et maximaux (formant des tubes) ainsi que des critères sur la forme des spectres afin d'éviter les chevauchements entre les différentes signatures spectrales des groupes de phytoplancton. Il était donc parfois difficile de prendre en compte certaines formes de spectres si celles-ci se retrouvaient «à cheval» entre différents tubes. Cette approche représentait une limite de PHYSAT et restreignait ses possibilités de développement. En effet, il était difficile de caractériser avec précision l'entière variabilité spatio-temporelle de ces anomalies de luminances marines et de les utiliser pour détecter les signatures spectrales des groupes phytoplanctoniques à partir de grandes bases de données *in situ* sans l'utilisation d'une approche statistique objective. Or l'utilisation d'une plus grande base de données de pigments, incluant une plus grande variabilité de conditions environnementales, et sa mise en relation avec les données satellite constitue un atout important pour améliorer la méthode et aller de l'avant dans la détection des groupes de phytoplancton. Il était donc nécessaire de développer une autre approche permettant à la fois une caractérisation plus fine et fiable des spectres de Ra en termes de forme et d'amplitude et une intégration efficace de plus grandes bases de données *in situ* afin d'élargir les perspectives en termes de détection des groupes de phytoplancton.

1.4. Plan et objectifs de la thèse

L'objectif principal de ma thèse a consisté à développer une méthode de classification automatique des anomalies spectrales de PHYSAT, dans le but d'améliorer leur caractérisation en termes de formes et d'amplitudes, puis, d'appliquer cette classification à la détection des principaux groupes de phytoplancton dominants à l'échelle globale.

Le développement d'une nouvelle approche, basée sur l'utilisation conjointe de réseaux de neurones artificiels et d'une base de données *in situ* de pigments plus large que lors de la mise au point de la version de PHYSAT originale, constitue la démarche que nous avons choisie pour atteindre cet objectif. La famille de réseaux de neurones utilisée, appelée « cartes auto-organisatrices » ou « SOM » (pour Self-Organizing Maps, Kohonen, 1984) est

un outil de classification automatique de type non-supervisée qui permet de regrouper les spectres de Ra selon leurs formes et amplitudes. Les cartes auto-organisatrices constituent un outil adapté à notre objectif et ont déjà été utilisées dans le cadre de plusieurs applications en télédétection, notamment, en couleur de l'océan (e.g. Richardson et al., 2003 ; Niang et al., 2003, 2006 ; Brajard et al., 2007 ; Jouini et al., 2013 ; Diouf et al., 2013). Dans le cadre de notre étude, l'intérêt de l'utilisation des cartes auto-organisatrices consiste à avoir une description précise de la variété d'anomalies spectrales existante à l'échelle globale, en effectuant une classification automatique à partir d'une grande base de données de spectres. Cet outil nous permet d'apporter une première série de réponses aux questions suivantes : Dans quelles gammes de valeurs varient ces spectres ? Quelles sont les formes majoritairement représentées à l'échelle globale ?

L'utilisation de la base de données *in situ* de pigments nous a permis par la suite d'attribuer un « label », c'est-à-dire, un (ou plusieurs) groupe (s) de phytoplancton à différents spectres. A partir de cela, les questions suivantes ont été étudiées : Quelle est la répartition globale de ces groupes par rapport à l'approche d'Alvain et al. (2005) et quelles sont les améliorations apportées par la nouvelle approche de classification ? Les différentes formes et amplitudes de spectres classées, sont-elles toutes associées à des groupes de phytoplancton dominants ? Par ailleurs, quelles hypothèses peut-on formuler dans le cas où un groupe de phytoplancton se trouve associé à des signatures spectrales très différentes en termes de formes et amplitudes ?

Toutes ces questions seront progressivement abordées dans le manuscrit. La méthode PHYSAT ainsi que ses limitations sont décrites en détail dans le chapitre 2. L'algorithme des cartes auto-organisatrices et son application à la classification automatique des anomalies spectrales sont présentées dans le chapitre 3. Dans le 4^{ème} chapitre, on applique la classification automatique faite avec les cartes auto-organisatrices à la télédétection des groupes de phytoplancton à l'échelle globale : c'est la méthode PHYSAT-SOM. Le 5^{ème} chapitre est réservé à l'analyse des signatures spectrales des diatomées et leurs blooms dans certaines régions. Enfin, le dernier chapitre présente les perspectives ouvertes par l'utilisation des cartes auto-organisatrices, notamment à travers un premier test de leur application à la classification automatique des anomalies spectrales à l'échelle régionale en Manche et Mer du Nord.

CHAPITRE 2

La méthode PHYSAT : Approche détaillée

2.1. Introduction

La méthode PHYSAT (Alvain et al. 2005, 2008, 2012) fait partie des approches spectrales pour la détection des groupes de phytoplancton. Les différentes étapes clefs de cette méthode sont décrites dans ce chapitre. L'idée originale était de travailler à partir de la variabilité de second ordre des mesures de nLw, après avoir éliminé la part liée à la concentration en chlorophylle *a*. Ensuite, grâce à des mesures *in situ* de pigments, cette variabilité est ensuite utilisée afin de détecter différents groupes de phytoplancton, quand ils sont dominants.

Ce chapitre a pour objectif la présentation détaillée de la méthode PHYSAT. Dans un premier temps, la variabilité de second ordre des luminances marines normalisées sera expliquée à travers la définition des anomalies spectrales (Ra ou Radiance Anomaly), paramètre base de la méthode. L'approche utilisée classiquement pour identifier les groupes de phytoplancton à partir des inventaires pigmentaires sera ensuite décrite. Enfin, la méthodologie suivie pour définir les signatures spectrales de chaque groupe phytoplanctonique ainsi que les limitations générales de PHYSAT seront détaillées.

2.2. Variabilité de second ordre des luminances marines normalisées : les anomalies spectrales

Les données satellites utilisées pour développer PHYSAT sont les niveaux 3 journaliers du capteur SeaWiFS, fournis par la NASA/GSFC/DAAC (NASA Goddard Space Flight Center). Ces données consistent en des cartes journalières de concentration en chlorophylle *a*, de l'épaisseur optique des aérosols à 865 nm et des luminances marines normalisées (nLw) à 412, 443, 490, 510 et 555 nm, dont certaines sont colocalisées (prises simultanément) avec les mesures *in situ* de pigments. La méthode PHYSAT est basée sur l'étude des anomalies de nLw, appelées anomalies spectrales ou Ra (Radiance Anomaly), définies (Alvain et al. 2005) telles que :

$$Ra(\lambda) = nLw(\lambda) / nLw_{ref}(\lambda, Chla) \quad (\text{Eq. 2.1})$$

où $nLw_{ref}(\lambda, Chla)$ représente le spectre moyen de nLw pour une concentration en chlorophylle *a* donnée, indépendamment des groupes phytoplanctoniques présents au moment de la mesure.

$Ra(\lambda)$ est donc, par définition, une grandeur indépendante de la concentration en chlorophylle *a* et représente la variation de second ordre de nLw (Alvain et al., 2005). Idéalement, $nLw_{ref}(\lambda, Chla)$ aurait dû être la fonction inverse de l’algorithme bio-optique utilisé. Cependant, ce dernier étant basé sur un choix entre plusieurs rapports de luminance (O’Reilly et al., 1998, 2000), il ne pouvait pas être inversé analytiquement. Cette référence a donc été définie empiriquement, à partir d’un grand jeu de données satellite de chlorophylle *a* et de nLw associées.

Les valeurs moyennes de nLw (λ) ont ainsi été calculées pour 26 plages étroites de valeurs de chlorophylle *a* situées entre 0.04 mg.m^{-3} et 3 mg.m^{-3} , permettant ainsi de créer un tableau de valeur ou LUT (Look Up Table) de $nLw_{ref}(\lambda, Chla)$. Ces valeurs sont présentées dans la figure 2.1. La concentration en chlorophylle *a* doit être inférieure à un maximum fixé arbitrairement (à 3 ou 5 mg.m^{-3} selon les versions) afin de diminuer la probabilité de présence d’eaux aux floraisons très fortes, non classées cas 1, également potentiellement côtières et impactées par de la matière terrigène, et supérieure à 0.04 mg.m^{-3} afin d’exclure les eaux ultra-oligotrophes où il devient difficile de distinguer un groupe de phytoplancton dominant dans les mesures de couleur de l’océan. (Alvain et al. 2005). De plus, afin de garder uniquement les mesures de bonne qualité, l’épaisseur optique des aérosols à 865 nm doit être inférieure ou égale à 0,15. Ce seuil vise à minimiser l’impact des erreurs liées aux corrections atmosphériques sur les nLw. (Alvain et al. 2005).

En utilisant l’équation 2.1, il est ainsi possible de calculer des cartes journalières de $Ra(\lambda)$ (qui sera noté « Ra » dans le reste du manuscrit), à partir des cartes de nLw et de la concentration en chlorophylle *a* (en interpolant si nécessaire les valeurs de la LUT définie ci-dessus). Pour les eaux du Cas 1, l’hypothèse de base de PHYSAT est que les formes et amplitudes des spectres de Ra sont liées à la composition en groupes de phytoplancton. La seconde étape passe donc par une utilisation d’inventaires pigmentaires pour calibrer la méthode.

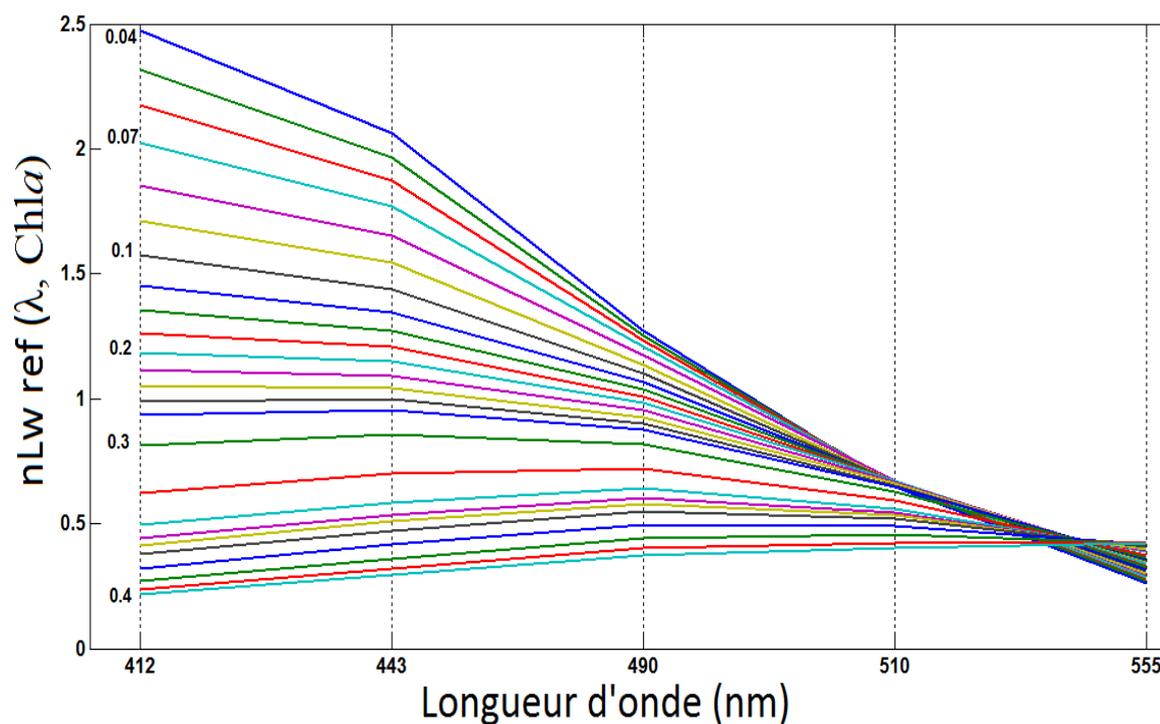


Figure 2.1. Valeurs des luminances marines de référence $nLw_{ref}(\lambda, Chl a)$ calculées à partir de moyennes pour des petites plages de valeurs de concentration en chlorophylle *a* (calculées d'après le modèle bio-optique OC4v6). Ces dernières sont indiquées d'une manière synthétique à gauche des courbes.

2.3. Identification des groupes de phytoplancton à partir des inventaires pigmentaires

2.3.1. Diversité des approches disponibles pour l'exploitation des inventaires pigmentaires

Les inventaires pigmentaires sont communément utilisés pour distinguer et identifier les groupes phytoplanctoniques. En effet, afin d'exploiter l'information contenue dans les inventaires de pigments, des « indices de pigments » ont été développés pour caractériser la composition taxinomique en phytoplancton à partir des pigments biomarqueurs (pigments caractéristiques d'un groupe de phytoplancton) (Claustre, 1994 ; Barlow et al., 2002 ; Gieskes et al., 1988 ; Mackey et al., 1996 ; Vidussi et al., 2001; Uitz et al., 2006). Ces études ont proposé plusieurs approches en se basant sur les concentrations relatives en pigments accessoires. S'ils sont communément admis et utilisés, les indices concernant la concentration relative de chaque pigment pour chaque groupe phytoplanctonique restent très variables dans la littérature.

Les méthodes proposées diffèrent selon le jeu de données utilisé, de même qu'avec le but de l'étude (recherche d'un groupe, d'un mélange de groupes, d'une proportion de groupes, de classes de tailles, etc.). Claustre et al. (2004) ont par exemple proposé un ratio qui permet de quantifier la proportion relative des diatomées et des dinoflagellés en se basant sur les concentrations en fucoxanthine et péridinine qui sont respectivement les biomarqueurs des diatomées et des dinoflagellés. L'approche CHEMTAX (Mackey et al., 1996) consiste en un algorithme qui calcule des ratios pigmentaires spécifique d'un échantillon donné et donne en sortie les abondances de différentes classes de phytoplancton. D'autres approches (Gieskes et al., 1988) sont basées sur des relations (régressions multiples) entre la concentration en chlorophylle *a* et celle de plusieurs pigments biomarqueurs. L'exploitation de ces relations permet, ensuite, d'estimer avec précision le ratio (Chl *a* / pigment biomarqueur) pour le jeu de données considéré (Uitz et al., 2006), donnant ainsi une information sur la composition en groupes de phytoplancton.

Des exercices d'intercomparaison basés sur des mesures HPLC de pigments ont montré que, si la concentration en chlorophylle *a* pouvait être déterminée avec une incertitude d'environ 8% (Claustre et al., 2004), la précision pour les pigments accessoires est généralement moindre. Il a cependant été montré que les indices pigmentaires (ratios) pouvaient être déterminés avec une précision plus élevée que celle des concentrations brutes de pigments biomarqueurs. Cette amélioration est à priori liée à la normalisation (par exemple, la division de la concentration brute du pigment biomarqueur par la concentration totale en chlorophylle *a* dans le cas de PHYSAT) qui annule certaines erreurs analytiques (Uitz et al., 2006).

2.3.2. Identification des groupes de phytoplancton à partir des mesures in situ

En utilisant les méthodes existantes basées sur les pigments biomarqueurs, il est possible d'identifier à ce jour 5 groupes de phytoplancton potentiellement dominants dans les bases de données pigmentaires disponibles: les *Prochlorococcus* (cyanobactéries prochlorophytes), les diatomées, les dinoflagellés, les haptophytes (ou nanoeucaryotes) et les *Synechococcus* (cyanobactéries). Ces groupes sont en effet caractérisés par des pigments spécifiques (Tableau 2.1). Cinq des pigments classiquement mesurés sont caractéristiques d'un groupe de phytoplancton et peuvent être utilisés pour en détecter la présence dans un échantillon d'eau (la chlorophylle *a* étant utilisée pour la normalisation des spectres de

luminance marine et la pheophytine *a* étant exploitée comme indice de présence de matériel dégradé).

Tableau 2.1. Principaux pigments et message taxinomique associé

	Notation	Message taxinomique
Chlorophylle <i>a</i>	Chl <i>a</i>	Tous les groupes sauf <i>Prochlorococcus</i>
Divinyl-Chlorophylle- <i>a</i>	Div <i>a</i>	<i>Prochlorococcus</i>
Pheophytine <i>a</i>	Pheo <i>a</i>	Produit de dégradation
Fucoxanthine	Fucox	Diatomées
Péridinine	Perid	Dinoflagellés
19'-Hexanoyloxyfucoxanthine	19'HF	Prymnésiophytes
Zéaxanthine	Zeax	<i>Synechococcus / prochlorococcus</i>

Les approches utilisées pour identifier les groupes phytoplanctoniques à partir des inventaires de pigments étant très variables d'une étude à une autre, un choix a dû être fait pour la recherche de liens entre les groupes dominants et les valeurs spécifiques de Ra (Alvain et al. 2005). Ainsi, dans le cas de PHYSAT, l'identification des groupes de phytoplancton s'est basée sur un choix de critères pigmentaires établis de façon à conserver les cas de dominances. L'indice de concentration relative en pigments biomarqueurs, communément utilisé pour l'identification des groupes de phytoplancton dans les inventaires pigmentaires, a été utilisé. Il est défini par :

$$P_{rel} = P / (Chl\ a + Div\ a) \quad (\text{Eq. 2.2})$$

où *P* est la concentration mesurée du pigment dans l'eau de mer, et *Chl a* et *Div a* respectivement la concentration en chlorophylle *a* et en divinyl chlorophylle *a*.

Un ensemble de seuils permettant de caractériser un maximum d'inventaires de pigments en termes de groupes dominants ont été définis et appliqués sur les données pour lesquelles des mesures spatiales coïncidentes étaient disponibles. Ces critères sont basés sur

l'analyse bibliographique de Mackey et al. (1996) pour les populations des eaux équatoriales et, sur l'étude de Lampert (2001), pour les eaux de l'Atlantique nord. Ceci a permis d'établir les critères résumés dans le tableau 2.2. A noter qu'à ce jour aucun inventaire n'a été labellisé comme « dinoflagellés » malgré le très faible seuil sur la concentration relative en péridinine (tableau 2.2). Cela n'est toutefois pas surprenant car les dinoflagellés sont généralement connus pour être présents sur une large échelle spatiale sans jamais dominer les populations de phytoplancton dans les eaux de l'océan ouvert (Jeffrey et al., 1997). En raison de leur abondance permanente dans les observations *in situ*, la 19'HF et la fucoxanthine ne présentent pas de seuils maximums pour les groupes de phytoplancton dont ils ne sont pas caractéristiques. En effet, ces deux pigments sont en général connus pour être présents dans d'autres groupes que ceux dont ils sont le biomarqueur principal.

Un critère additionnel concernant la présence d'éléments organiques dégradés a été ajouté en se basant sur un seuil de la pheophytine *a* lorsque celui-ci est mesuré dans les échantillons. Ces critères permettent donc de labelliser les inventaires pigmentaires en termes de groupes phytoplanctoniques dominants pour des eaux de l'océan ouvert.

Tableau 2.2. Seuils en pigments relatifs utilisés pour associer un inventaire de pigments à un seul groupe de phytoplancton spécifique. Les valeurs en gras correspondent aux biomarqueurs de chaque groupe (Alvain et al. 2008).

	Pheo <i>a</i>	Div <i>a</i>	Perid	Fucox	19'HF	Zeax
Diatomées	< 0,30	< 0,40	< 0,10	> 0,18	----	< 0,20
Prochlorococcus	< 0,30	> 0,40	< 0,10	----	----	> 0,35
Haptophytes	< 0,30	< 0,40	< 0,10	----	> 0,14	< 0,20
SLC (<i>Synechococcus</i> <i>Like</i> <i>Cyanobacteria</i>)	< 0,30	< 0,40	< 0,10	----	----	> 0,20
Dinoflagellés	< 0,30	< 0,40	> 0,10	----	----	< 0,20

L'étape suivante consiste à appliquer ces critères aux inventaires pigmentaires pour lesquels des mesures spatiales coïncidentes sont disponibles, ceci de façon à pouvoir identifier par la suite un signal satellitaire pour chaque groupe de phytoplancton. Par exemple, en 2005,

parmi les 176 inventaires disponibles (avec une mesure satellite coïncidentes), 41 ont pu être associés à un groupe dominant en appliquant les seuils sur les pigments biomarqueurs. Les autres inventaires n'ont pas été conservés car ils ne remplissaient pas tous les critères correspondant à un groupe de phytoplancton dominant. Le même travail peut être reproduit sur d'autres bases de données, comme cela a par exemple été le cas en 2012 (Alvain et al. 2012) pour la validation de PHYSAT avec une base de données complémentaire ou lors du travail présenté dans la suite de ce manuscrit. Disposant à présent d'inventaires pigmentaires dont le groupe dominant est identifié, il s'agit d'étudier dans la prochaine étape les mesures satellite de luminance correspondantes afin d'identifier un signal spécifique à chaque groupe de phytoplancton.

2.4. Relation entre les anomalies spectrales et la dominance des groupes de phytoplancton, définition des signatures spectrales des groupes phytoplanctoniques

Dans cette sous-partie, nous décrivons la méthode suivie par Alvain et al. (2005, 2008) pour relier les anomalies spectrales à des groupes de phytoplancton dominants identifiés dans les inventaires pigmentaires (paragraphe 2.2.2).

2.4.1. Anomalies spectrales associés aux groupes dominants identifiés dans les inventaires de pigments

Afin d'associer des signatures spectrales de R_a spécifiques aux groupes de phytoplancton dominants, un jeu de spectres est sélectionné à partir des mesures du capteur étudié en sélectionnant uniquement les pixels lors de conditions claires, du même jour que la mesure *in situ* et dans un rayon de 1 pixel autour de celle-ci (fenêtre de 3x3 pixels). Le nombre de pixels valides, associés aux mesures *in situ* est donc au maximum de 9. Pour chaque observation et pixel valide, les R_a ont été calculés sur les 5 longueurs d'onde citées précédemment en utilisant l'équation 2.1.

On dispose donc ainsi d'un jeu de spectres de R_a et d'inventaires de pigments associés. Les anomalies spectrales correspondant aux mesures pour lesquelles un groupe de phytoplancton dominant peut être identifié sont par la suite analysées. La figure 2.2 montre un exemple de sous-ensemble de spectres associés à différents groupes dominants (Alvain et al. 2005).

Ce premier résultat a permis de montrer pour la première fois en 2005, un lien entre la forme et l'amplitude des spectres et le groupe de phytoplancton présent. Ainsi, selon ces premiers travaux, les nanoeucaryotes sont caractérisés par des valeurs de Ra relativement faibles avec une hausse des valeurs entre les petites longueurs d'onde (412 et 443 nm) et les grandes longueurs d'onde. *Prochlorococcus* est associé à des spectres plus aplatis avec des valeurs de Ra autour de 1. Les SLC présentent des spectres ayant une forme plus ou moins similaire à celle des *Prochlorococcus* avec toutefois des valeurs plus élevées et une légère hausse autour de 412 nm. Les similarités entre les spectres des *Prochlorococcus* et des SLC peuvent être expliquées par les caractéristiques proches de ces cyanobactéries en termes de taille et d'environnement occupé (Alvain et al., 2012). Ce sont également tous deux des organismes faisant partie du picoplancton (0,2 – 2 µm) et qui dominent généralement les eaux oligotrophes.

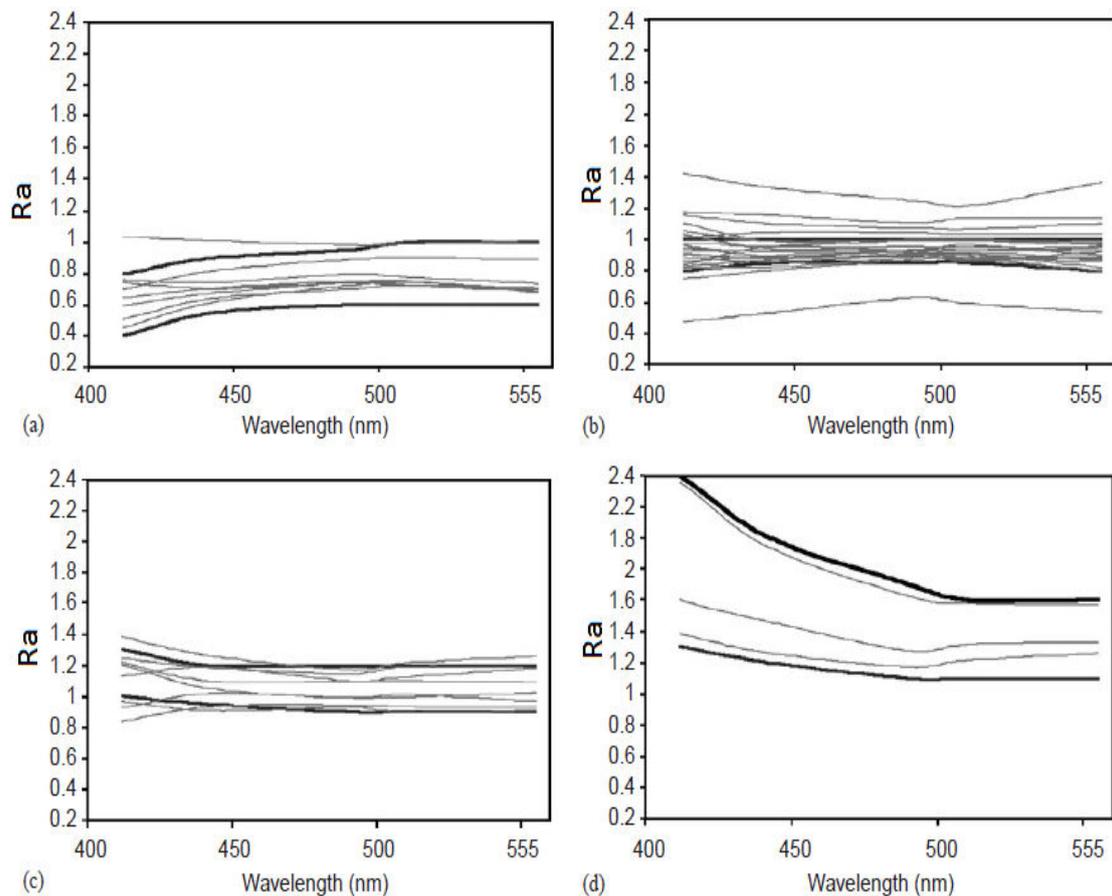


Figure 2.2. Signatures spectrales de Ra pour chacun des 4 assemblages de phytoplancton identifiés à partir de la base disponibles en 2005 dominés par « nanoeucaryotes » (a), « *Prochlorococcus* » (b), « SLC » (c) et « diatomée » (d). Les spectres individuels sont en gris clair et les bornes minimales et maximales définissant la gamme de variabilité spectrale de chaque assemblage sont représentées en traits pleins noirs (figure issue de Alvain et al. 2005).

Les diatomées sont détectables par des valeurs de Ra relativement fortes (entre 1.2 et 2.4) avec une forte chute entre 412 nm et les plus grandes longueurs d'onde. Les valeurs relativement élevées de Ra associées aux diatomées ont été récemment expliquées par les fortes de valeurs de rétrodiffusion (diffusion dans la direction opposée à celle de la lumière incidente) associées aux diatomées (Alvain et al., 2012). En effet, les valeurs de rétrodiffusion sont généralement d'autant plus élevées que la taille des cellules (et des particules en général) est plus grande (Loisel et al., 2007 ; Whitmire et al., 2010 ; Neukermans et al., 2012).

Ces résultats permettent par la suite de fixer un ensemble de seuils et de critères pour caractériser chaque ensemble de spectres associés à un groupe donné (tableau 2.3). Ce travail a été réalisé tout d'abord en procédant à une analyse des formes et amplitudes des spectres associés à chaque groupe de phytoplancton. Des critères ont été définis afin de maximiser le nombre de spectres à l'intérieur de tubes (valeurs maximales et minimales), tout en évitant au maximum les recouvrements pouvant apparaître entre les signatures spectrales de plusieurs groupes.

Tableau 2.3. Seuils de Ra utilisés par PHYSAT pour définir la signature spectrale des 4 groupes de phytoplancton

	412 nm	443 nm	490 nm	510 nm	555 nm	Critères supplémentaires
Nano-min	0,4	0,55	0,6	0,6	0,6	Ra(412)<Ra(443)
Nano-max	0,8	0,9	0,95	1	1	Ra(443)>Ra(490)
Proch-min	0,8	0,85	0,85	0,85	0,8	
Proch max	1	1	1	1	1	
SLC-min	1,0	0,95	0,9	0,9	0,9	Ra(412)>Ra(443)
SLC-max	1,3	1,2	1,2	1,2	1,2	Ra(412)<Ra(490)
Diat-min	1,3	1,2	1,1	1,1	1,1	Ra(412)>Ra(490)
Diat-max	2,4	2	1,7	1,6	1,6	Ra(490)>Ra(555)

Il a ainsi été montré, depuis 2005, que différents groupes, identifiés comme étant dominants à partir d'inventaires de pigments, correspondent à des signaux de Ra d'amplitudes

et de formes différentes. On peut donc associer un groupe de phytoplancton dominant à un signal satellite particulier de Ra en éliminant l'effet de la concentration en chlorophylle *a*. Il devient possible de produire des cartes de groupes de phytoplancton dominants en appliquant les critères du tableau 2.3 pixel par pixel. Chaque pixel d'une image sera alors associé à un groupe de phytoplancton dominant, si son spectre de Ra remplit les critères prédéfinis lors de l'étape de calibration. Les pixels dont les spectres de Ra ne correspondent à aucun groupe dominant ne sont pas identifiés mais restent valides. Les cartes journalières sont utilisées par la suite pour réaliser des moyennes mensuelles ou des climatologies pluriannuelles (moyennes mensuelles pluriannuelles). La figure 2.3 montre un exemple de climatologie de la distribution des groupes phytoplanctoniques dominants entre 1997 et 2010, pour les mois de Janvier, Avril, Juin et Octobre.

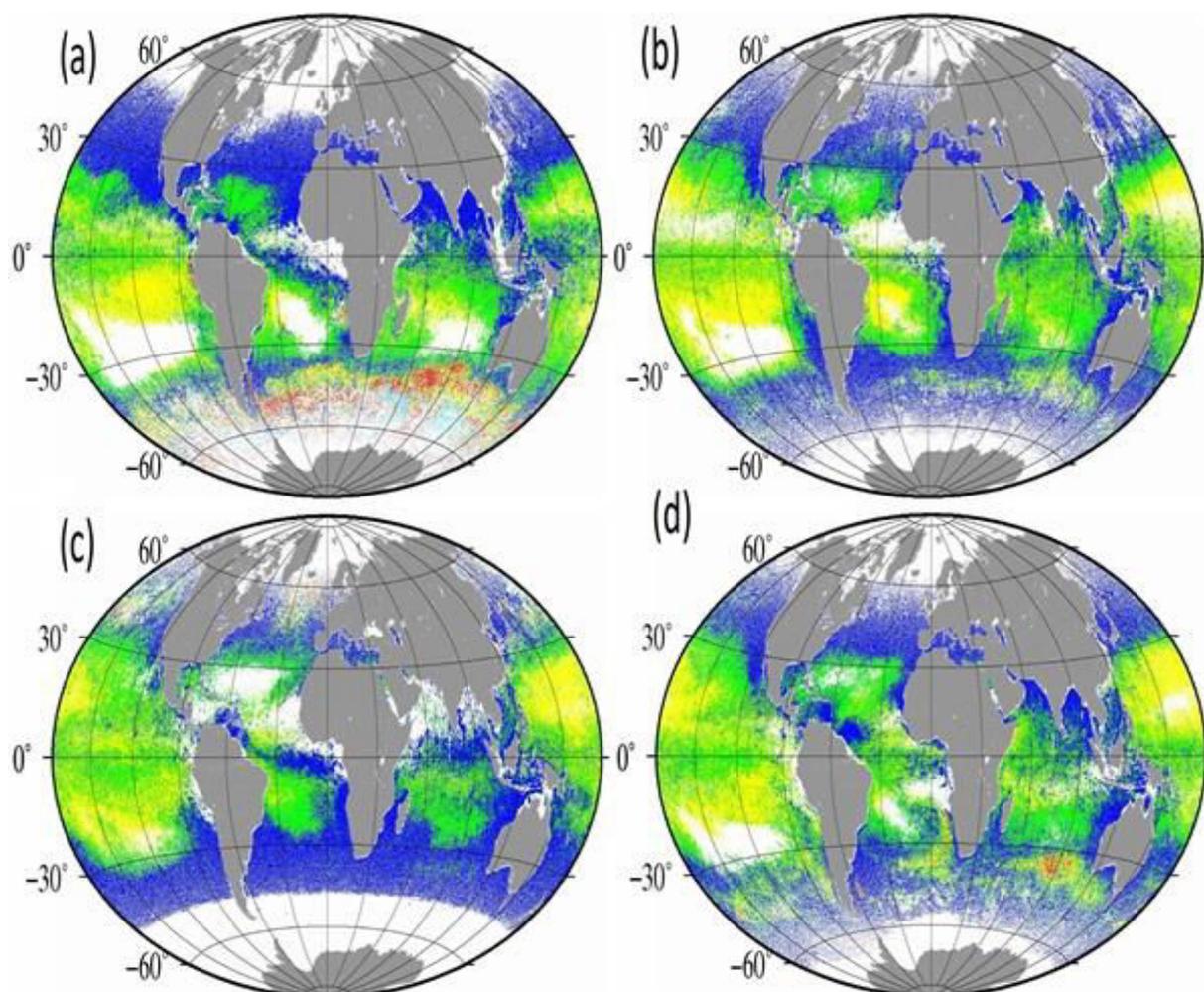


Figure 2.3. Synthèses de la distribution des groupes de phytoplancton dominants pour les mois de Janvier (a), Avril (b), Juin (c) et Octobre (d) entre 1997 et 2010. En bleu les nanoeucaryotes, en vert, les Prochlorococcus, en jaune les SLC et en rouge les diatomées (Cartes réalisées à partir des données SeaWiFS dans leur version de traitement de 2010)

Les structures globales sont caractérisées par une dominance des cyanobactéries (*Prochlorococcus* et SLC) dans les eaux oligotrophes tropicales, leurs petites tailles les rendant moins sensibles aux faibles teneurs en nutriments, alors que les nanoeucaryotes et les diatomées sont dominants dans les eaux plus riches en nutriments des hautes et moyennes latitudes (Alvain et al. 2008). On observe un cycle saisonnier marqué dans les hautes latitudes où les nanoeucaryotes dominent en hiver et où de larges régions de floraisons de diatomées apparaissent en été. L'extension de ces floraisons de diatomées est particulièrement importante dans l'océan Austral au mois de Janvier.

2.4.2. Liens entre les anomalies spectrales et l'environnement bio-optique

La méthode PHYSAT a mis en évidence un lien entre un signal spécifique de R_a et la présence dans l'eau d'un groupe de phytoplancton particulier. Une étude récente a montré que les liens entre les anomalies et les groupes sont liés à l'environnement optique spécifique associé à chaque cas. Ceci a été montré théoriquement par Alvain et al. (2012) à partir d'une étude de la sensibilité de R_a issues de simulation via le code Hydrolight (Mobley, 1994). Si de nombreuses études ont visé la compréhension de l'influence de la taille, de l'absorption et de la rétrodiffusion de la lumière par différents types de particules (Stramski et al., 2001, 2004, Bricaud et al., 1983, 1988, 2004) chacune de ses études ne permettait pas d'expliquer l'organisation des spectres de PHYSAT.

Alvain et al. (2012) ont pris en compte l'ensemble de ces différents paramètres, ce qui a permis de montrer, pour la première fois, les liens existant entre les anomalies spectrales et les propriétés optiques inhérentes du milieu (IOP pour Inherent Optical Properties). Ces propriétés optiques sont représentées par 3 paramètres qui sont : la rétrodiffusion par les particules, notée b_{bp} , l'absorption par le phytoplancton (a_{phy}) et l'absorption par les matières organiques dissoutes (a_{cdom}). L'ensemble de ces paramètres constitue ce que l'on désignera par « environnement bio-optique ». Les valeurs de ces paramètres dépendent de la nature du milieu, de ses constituants et de leur structure (taille, forme).

Les résultats montrent que, pour une concentration en chlorophylle a donnée, la rétrodiffusion par les particules explique la majeure partie de la variabilité spectrale des R_a , en termes d'amplitude. Les variations de l'absorption par les matières organiques dissoutes (a_{cdom}) ainsi que l'absorption par le phytoplancton peuvent aussi avoir un impact important sur les signatures spectrales des R_a , surtout en termes de formes. Pour des environnements bio-optiques spécifiques, les eaux dominées par les cyanobactéries (pico-plancton) sont associées

aux plus fortes valeurs d' a_{phy}^* (absorption par le phytoplancton divisée par la concentration en chlorophylle a). En revanche, ce groupe est associé à de faibles valeurs d' a_{cdom} et de b_{bp} par rapport aux valeurs moyennes. Les fortes valeurs d' a_{phy}^* sont probablement associées à l'effet d'agglomération des cellules phytoplanctoniques (Alvain et al., 2012).

Les eaux dominées par les diatomées sont associées aux plus fortes valeurs de b_{bp} et d' a_{cdom} . Les valeurs de b_{bp} augmentent proportionnellement avec le diamètre de la cellule, ce qui explique les valeurs relativement fortes d'anomalies spectrales caractérisant les diatomées (Alvain et al., 2012). En revanche, les eaux dominées par les diatomées sont caractérisées par les plus faibles valeurs d' a_{phy}^* relativement aux valeurs moyennes. Les nanoeucaryotes se caractérisent par de fortes valeurs d' a_{cdom} et des valeurs modérées de b_{bp} (moins élevées que celles des diatomées et plus élevées que celles des cyanobactéries). Ces résultats montrent que chaque assemblage de groupes dominants se trouve dans un environnement bio-optique particulier, caractérisé par des proportions différentes d'IOP. En considérant leurs valeurs d'IOP moyennes respectives, les amplitudes des spectres de Ra retrouvées théoriquement pour ces 3 groupes de phytoplancton sont en accord avec celles des anomalies spectrales mise en évidence empiriquement (Alvain et al., 2005).

Ces résultats démontrant la sensibilité des Ra et, par conséquent, des groupes de phytoplancton détectés à l'environnement bio-optique considéré, ouvrent de nouvelles perspectives quant à la détection des groupes phytoplanctoniques. On peut en effet imaginer qu'une analyse fine des formes et amplitudes des Ra en relation avec ces différents paramètres optiques puisse permettre la détection d'avantages de situations en termes de compositions phytoplanctoniques ou l'ajout de critères de confiances dans les observations.

2.5. Limitations de PHYSAT

La méthode PHYSAT a permis de démontrer, pour la première fois, qu'un proxy de la mesure satellite pouvait être relié à différents groupes de phytoplancton. Cependant, l'approche par laquelle la signature spectrale de chaque groupe a été définie présente plusieurs limitations quant à une caractérisation précise et souple de la variabilité des spectres de Ra en termes de forme et d'amplitude ainsi que l'exploitation de plus grandes bases de données *in situ*. Avec l'approche initiale (utilisation de tubes), il est nécessaire d'introduire des critères additionnels empiriques et subjectifs (de plus, développés initialement à partir d'une base de données *in situ* restreinte) sur la forme des spectres afin d'éviter les recouvrements entre les signatures spectrales des groupes phytoplanctoniques, tout en essayant de maximiser le

nombre de spectres pris en compte. Ceci engendrait une élimination d'office d'une partie des cas de Ra labellisés « dominants » à partir de l'expertise *in situ* mais non caractérisable pour leur détection ultérieure en raison des critères visant à éviter les recouvrements. Par exemple, dans la version de 2005 de PHYSAT, l'utilisation de tous ces critères sur les 41 spectres identifiés au départ, n'a permis de conserver que 20 anomalies spectrales (figure 2.4). Les 21 spectres restants, associés à une situation de groupes dominants, ne répondaient pas aux seuils définis ou engendraient un chevauchement entre plusieurs signatures spectrales. Ils n'étaient toutefois pas mal classés mais non classés puisqu'ils n'appartenaient à aucun groupe, et non pas à un mauvais groupe.

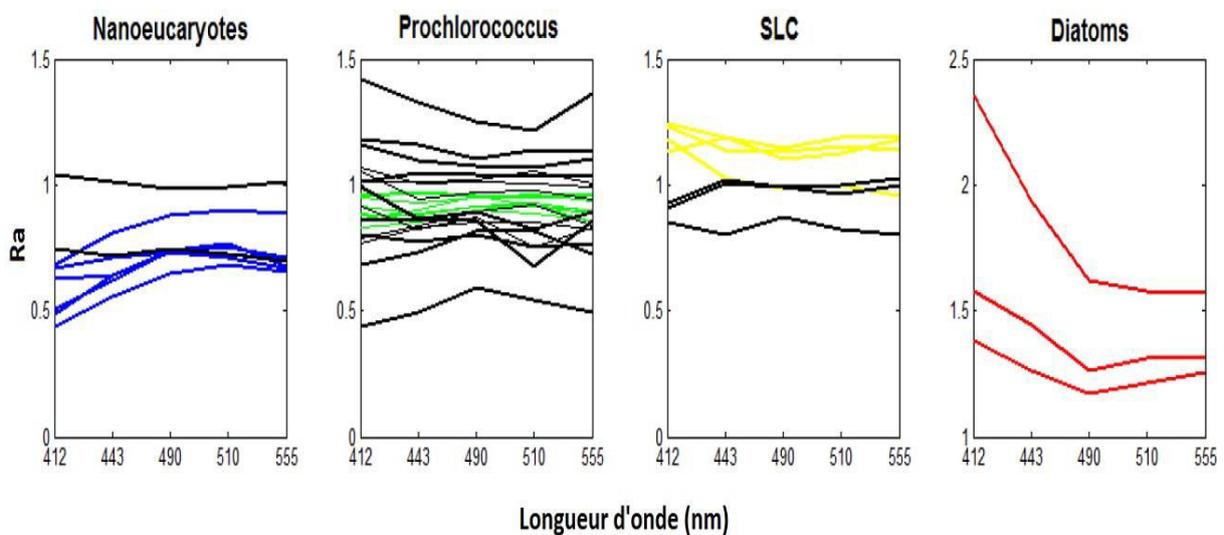


Figure 2.4. Les 41 spectres de Ra correspondants aux inventaires de pigments dominés par les 4 groupes de phytoplancton. Les spectres qui ne répondent pas aux seuils et critères introduits pour définir la signature spectrale de chaque groupe sont représentés en noir. Sur les 41 spectres, 21 n'ont finalement pas été pris en compte par la méthode dans sa version de 2005 et 2008.

Il apparaît donc qu'avec cette approche, plus de 50% des spectres reliés à un groupe dominant grâce aux données pigmentaires n'étaient finalement pas exploités pour définir les signatures spectrales des groupes de phytoplancton, ce qui réduisait la représentativité spatiale de la labellisation en termes de groupes de phytoplancton. En effet, les spectres de Ra qui n'ont pas été pris en compte après l'application des critères peuvent constituer une information potentielle. Une méthode, permettant la caractérisation d'une plus grande variabilité spectrale des Ra en termes de formes et d'amplitudes à partir de l'ensemble des bases de données spatiales, était nécessaire afin d'améliorer la détection des groupes phytoplanctoniques. De plus, l'extension de la méthode à la caractérisation de plus grandes bases de données *in situ*, incluant une plus grande variabilité d'environnements bio-optiques

était difficile à concevoir en l'état. La possibilité d'identification d'autres groupes phytoplanctoniques et/ou des mêmes groupes dans des conditions différentes (mélanges de groupes, différents stades de blooms) était également limitée.

D'un point de vue bio-optique, l'exploitation des spectres de Ra pour détecter les groupes de phytoplancton n'est pas une tâche aisée compte-tenu de la complexité des paramètres qui entrent en jeu et qui impactent la forme et l'amplitude des spectres (Alvain et al., 2012). Par exemple, d'après cette étude théorique, il est tout à fait possible de retrouver dans un tube caractérisant des diatomées, un type de spectre nanoeucaryotes présentant une forme caractéristique qu'il aurait été nécessaire de redéfinir de nouveau avec de seuils. Les conditions et stades de croissances, les mélanges de groupes, la diversité au sein du même groupe (espèces) ou encore la présence des mêmes espèces phytoplanctoniques dans des milieux différents représentent autant de cas, parmi d'autres, impactant les formes et amplitudes des spectres de Ra. En outre, la présence de certains groupes ayant des formes spectrales très similaires, tels que *Prochlorococcus* et SLC qui sont très proches en termes de tailles et d'environnements occupés, rend la tâche de distinction entre ces groupes difficile. Une méthode plus adéquate était donc souhaitable.

Ces différents facteurs et limitations soulignent la nécessité d'une caractérisation objective et précise des spectres de Ra. Une méthode permettant une exploitation optimale des spectres de Ra et leur liaison à de grande bases de données *in situ*, comprenant une large variabilité d'environnements bio-optiques représente une solution pour atteindre cet objectif et faire un pas en plus dans la détection des groupes de phytoplancton. Dans le cadre de ma thèse nous avons donc opté pour l'approche par cartes auto-organisatrices ou Self-Organizing Maps (SOM, Kohonen, 1984) qui a déjà fait ses preuves dans diverses applications par le passé, notamment en couleur de l'océan (Brajard et al., 2007 ; Niang et al., 2003, 2006 ; Jouini et al., 2013 ; Diouf et al., 2013).

CHAPITRE 3

Les cartes auto-organisatrices : description détaillée de l'algorithme et application à la classification automatique des anomalies spectrales

Les limitations de PHYSAT dans sa version de 2005 et 2008, décrites dans le chapitre 2, nous ont motivés à chercher une méthode objective afin d'obtenir une meilleure caractérisation et description de la variabilité globale des anomalies spectrales en termes de formes et d'amplitudes. Le choix a été fait d'utiliser une approche neuronale de classification automatique de ces spectres d'anomalies: les cartes auto-organisatrices ou SOM (pour Self-Organizing Maps, Kohonen, 1984). Ce chapitre est dédié à l'explication du principe des SOM et à leur application à la classification automatique des anomalies spectrales de PHYSAT. La variabilité des anomalies spectrales à l'échelle globale ainsi que l'amélioration de leur caractérisation avec SOM sont également démontrées.

3.1. Réseaux de neurones artificiels

Un réseau de neurones artificiels est un modèle de calcul dont la conception est très schématiquement inspirée du fonctionnement des neurones biologiques. Ces modèles se basent généralement sur des méthodes d'apprentissage, dont le but est d'imiter, à l'aide d'algorithmes, la capacité qu'ont les êtres vivants à apprendre par « l'exemple ». Cette discipline a commencé à être développée dans les années 1940 (McCulloch et Pitts, 1943 ; Rosenblatt, 1958) et connaît, depuis plusieurs années, un essor dans un vaste domaine d'applications (e.g. fouille de données, bio-informatique, télécommunication, etc.). Un réseau de neurones est en général composé d'une succession de couches connectées, dont chacune prend ses entrées sur les sorties de la précédente (McCulloch et Pitts, 1943 ; Rosenblatt, 1958 ; Dreyfus, 2002).

D'une façon générale, l'apprentissage statistique regroupe plusieurs familles de modèles tels que les réseaux de neurones artificiels et les machines à vecteurs de support (Vapnik, 1998 ; Burges, 1998 ; Shawe-Taylor et Cristianini, 2000 ; Schölkopf et al., 2002 ; Bishop, 2006). Les méthodes d'apprentissage des réseaux de neurones peuvent être de type « supervisé » ou « non supervisé ». Un apprentissage est dit « supervisé » lorsque l'on force le

réseau à converger vers un état final précis, en même temps qu'on lui présente une donnée. À l'inverse, lors d'un apprentissage non-supervisé, le réseau est laissé libre de converger vers n'importe quel état final lorsqu'on lui présente une donnée.

La famille des réseaux de neurones artificiels regroupe plusieurs méthodes, entrant dans la catégorie des modèles statistiques non linéaires, telles que les « Perceptrons Multi Couches » ou « M.L.P » (Rosenblatt, 1958) et les cartes auto-organisatrices (ou auto-adaptatives) ou SOM (Kohonen, 1984). Dans la première des méthodes précitées, le neurone (ou perceptron) représente une fonction non linéaire, paramétrée. Il est caractérisé par des valeurs d'entrée, d'une fonction de combinaison et d'une fonction d'activation qui sert à introduire une non-linéarité dans le fonctionnement du neurone (Dreyfus, 2002).

Dans le cadre de cette thèse, les cartes auto-organisatrices (Kohonen, 1984) ont été utilisées. Cette classe de réseaux de neurones à apprentissage non supervisé est très adaptée à la problématique devant être traitée. En effet, elle permet d'étudier de larges bases de données en les regroupant suivant des caractéristiques communes : c'est une méthode de classification automatique. L'appartenance des cartes auto-organisatrices à la famille des méthodes neuronales s'explique par le fait que le formalisme neuronal permet une présentation claire et compacte des différents phénomènes mis en jeu (Dreyfus, 2002).

3.2. Cartes auto-organisatrices

3.2.1. Généralités

L'algorithme des cartes auto-organisatrices a été développé par Teuvo Kohonen (Kohonen, 1984). Les SOM sont utilisées pour cartographier un espace réel afin d'étudier la répartition de données dans un espace à grande dimension. En pratique, cette cartographie sert à réaliser des tâches de classification. Comme beaucoup d'autres méthodes d'intelligence artificielle, les SOM sont fortement inspirés par la biologie. Ils reproduisent le principe neuronal du cerveau : des *stimuli* de même nature excitent une région du cerveau bien particulière. Les neurones sont organisés dans le cortex de façon à interpréter tous les types de *stimuli* imaginables. De la même manière, les cartes auto-organisatrices se comportent de façon à représenter un ensemble de données complexes, pouvant être de dimension multiple. Chaque neurone se spécialise pour représenter un groupe bien particulier de données selon les points communs qui les rassemblent (forme, intensité, etc...).

Les SOM sont utilisées dans diverses applications qui font intervenir le codage de l'information, l'organisation des bases de données, l'exploration des données, la reconnaissance de formes ou en modélisation (Cotrell et al., 1997; Deboeck et Kohonen, 1998, Kohonen, 2013). En effet, elles représentent un outil statistique performant pour étudier, résumer et représenter des données hétérogènes multidimensionnelles qui peuvent contenir à la fois des variables quantitatives (à valeurs continues ou réelles) et qualitatives (discrètes, ordinales ou nominales). Parmi les différentes applications réalisées à l'aide des cartes auto-organisatrices, un assez grand nombre concerne des tâches de classification (Moreira et Costa, 1996; Merkl, 1999 ; Gonçalves et al., 2011 ; Hasan et Shamsuddin, 2011 ; Barnerjee et al., 2009).

3.2.2. Spécificités et adaptabilité au traitement des données satellite

L'extraction d'information spectrale facilement analysable et interprétable à partir des images satellite, n'est pas une tâche facile. Cette difficulté est d'autant plus accrue que les données sont volumineuses, généralement non linéaires et nécessitant une lourde phase de prétraitement des données (Richardson et al., 2003 ; Loyola et al., 2006). Le prétraitement des données (e.g. normalisation) est généralement nécessaire pour l'élimination des données aberrantes ou l'équilibrage du poids des différentes variables traitées pouvant introduire des biais dans leur analyse (Richardson et al., 2003 ; Vesanto et al., 1999; Kohonen et al., 2013).

Les méthodes de classification par réseaux de neurones artificiels, tels que SOM, possèdent de nombreux avantages par rapport aux méthodes de classification classiques (Wassermann, 1989). En effet, ces outils ont la capacité de résoudre des problèmes non linéaires très complexes (Dayhoff, 1990). Ils sont aussi beaucoup plus robustes et efficace quant au traitement des données très bruitées ou des données manquantes (Cotrell *et al.*, 1997 ; Richardson et al., 2003 ; Jouini et al., 2013). La capacité de SOM à gérer les données bruitées ou aberrantes a été comparée par rapport à d'autres méthodes de classification des données multi-variées, telles que les analyses en composantes principales (Richardson et al., 2003). SOM permet d'alléger la tâche de prétraitement des données. Par exemple, si les données aberrantes n'affectent qu'une seule unité de la carte auto-organisatrice (neurone) et son voisinage, elles pourraient avoir un effet drastique et disproportionné sur d'autres méthodes de classification de données. Ces atouts sont, par exemple, particulièrement recherchés pour le traitement des images satellite qui présentent souvent une portion considérable de données manquantes ou bruitées (e.g. présence des nuages ou de composés

atmosphériques tels que les aérosols) sans oublier les erreurs dues aux corrections atmosphériques (Gordon et Wang, 1994 ; Gordon, 1997 ; Ruddick et al., 2000 ; rapport IOCCG, 2010; Kuchinke et al., 2009 ; Jamet et al., 2011 ; Goyens et al., 2013). En outre, les méthodes de classification non supervisées ne nécessitent pas une connaissance préliminaire sur les données ou des hypothèses sur leur distribution (telles que la normalité ou l'égalité des variances) (Chen et Ware, 1999).

Plusieurs études ont exploré les potentialités offertes par SOM quant à la reconnaissance des formes spectrales à partir des images satellites ou de méthodes spectroscopiques de laboratoire (Richardson et al., 2003). Les potentialités offertes par SOM en termes de classification de données complexes et de reconnaissance de formes ont ouvert plusieurs perspectives à diverses applications en océanographie, notamment en télédétection de la couleur de l'océan. Parmi les études réalisées dans ce domaine, Ainsworth (1999) et Ainsworth et Jones (1999) ont utilisé SOM afin d'améliorer les estimations de chlorophylle à partir des données satellites, à travers une classification fine des pixels. Richardson et al. (2002) et Silulwane et al. (2001) ont utilisé SOM afin d'identifier des profils caractéristiques de chlorophylle dans l'océan. Par ailleurs, des recherches ont eu lieu pour la caractérisation des aérosols (Brajard et al., 2007, Niang et al., 2003, 2006, Diouf et al., 2013), la restitution des valeurs de chlorophylle manquantes en présence de nuages (Jouini et al., 2013) ainsi que pour l'analyse des spectres d'absorption du phytoplancton (Chazottes et al., 2006, 2007).

3.2.3. L'algorithme de SOM

C'est un algorithme d'auto-organisation qui projette l'espace des données, de grande dimension, sur un espace discret de faible dimension (en général 1, 2 ou 3) appelé « carte topologique » et noté *C*. Le principe de ce modèle consiste à partitionner de très grands ensembles d'observations en groupements similaires. Les groupements réalisés possèdent la particularité d'avoir une structure de voisinage qui peut être matérialisée par une carte topologique (Kohonen, 1988). Dans une première approche, ce modèle est utilisé dans un but descriptif. Les données à analyser sont constituées d'observations dont on cherche à comprendre la structure et les dépendances qui sont inconnues a priori (Kohonen, 1997).

L'algorithme SOM est en réalité une extension de l'algorithme d'apprentissage non supervisé « K-moyennes » : il cherche à minimiser une fonction de coût (ou erreur). Cette fonction de coût doit chercher, d'autre part, à assurer la conservation de la topologie. Dans ce contexte, la topologie constitue l'ensemble des propriétés qui déterminent les formes des

divers éléments constituant l'espace d'entrée (les données). La conservation de la topologie est réalisée par l'intermédiaire d'une distance définie sur la carte et d'une notion de voisinage qui lui est attachée. C'est cette dernière caractéristique qui fait la particularité de cartes auto-organisatrices. Mathématiquement, la différence entre les SOM et les K-moyennes réside dans le fait que l'algorithme de Kohonen modifie simultanément les paramètres de connexions d'une unité de la carte (neurone), appelés aussi « poids » ou « vecteurs poids », et de son voisinage. C'est de cette particularité que le terme « auto-organisation » (ou « auto-adaptation ») provient. C'est ainsi que s'introduit peu à peu l'ordre topologique : des unités proches sur la carte représentent des observations proches dans l'espace des données. Il s'agit d'une méthode de classification automatique très adaptée aux grandes bases de données.

3.2.3.1. Architecture neuronale et notion de voisinage

La carte auto-organisatrice est constituée de deux couches totalement connectées entre elles (Figure 3.1). La couche d'entrée sert à présenter au réseau les données à traiter. La seconde couche (carte topologique) est un treillis régulier à deux ou trois dimensions dont chacun des nœuds est un neurone (unité de classification). Le choix de la géométrie de la carte est fait a priori. Les liens entre les neurones de C se font par une structure de graphe « non orienté » (Contrairement aux graphes dits « orientés », dans lesquels le lien entre deux points donnés du graphe a un sens unique, les graphes « non orientés » sont caractérisés par une symétrie des liens entre tous les points : il n'y a pas de sens unique), qui induit une distance discrète d sur C .

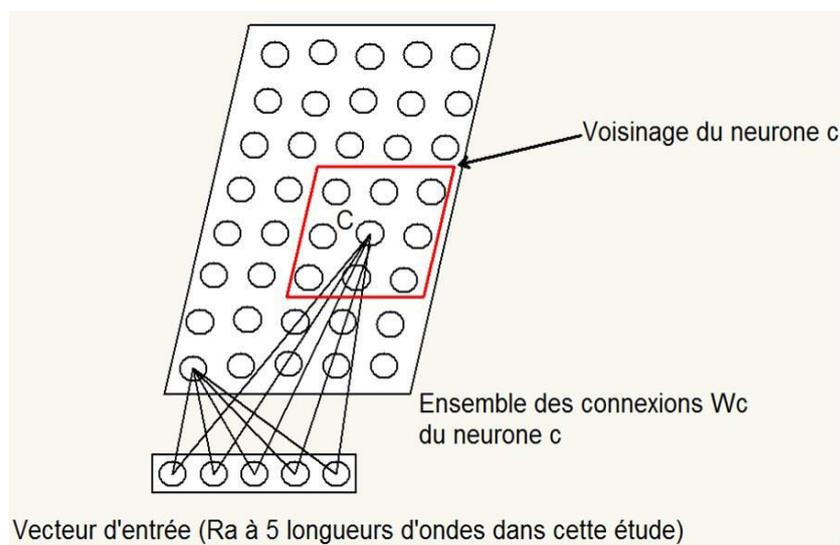


Figure 3.1. Structure de la carte topologique : treillis régulier de dimension 2, muni d'un système de voisinage et entièrement connecté à la couche d'entrée

Pour chaque neurone c appartenant à la carte topologique, cette distance permet de définir la notion de voisinage qui introduit les contraintes topologiques dans la représentation finale. Cette notion de voisinage est fondamentale dans l'algorithme des cartes topologiques, elle résulte de la forme de la carte sélectionnée et permet de définir la zone d'influence de chaque neurone.

Afin de gérer la taille du voisinage, on utilise une fonction noyau K^T qui est régie par un paramètre, appelé « température » et noté « T », que nous faisons décroître à chaque cycle de l'apprentissage (un cycle d'apprentissage représente chaque itération durant laquelle les données sont présentées au réseau de neurones), et par la distance d . Pour deux neurones c et r , on a :

$$K^T(d(c,r)) = \exp\left(\frac{d(c,r)^2}{-T^2}\right) \quad (\text{Eq. 3.1})$$

Le voisinage est d'autant plus faible que la distance entre les deux neurones est plus grande et l'influence exercée par le voisinage diminue avec la température. Les fonctions noyaux usuelles peuvent être de type gaussien ou à seuil (Figure 3.2). Dans la fonction de voisinage à seuil, les neurones du voisinage ont la même influence. En dehors, ils n'en ont aucune. Dans la fonction de voisinage de type gaussien, l'influence entre deux neurones dépend de la distance qui les sépare. C'est la fonction noyau de type gaussien qui a été utilisée dans le cadre de notre étude car elle assure une certaine souplesse quant à la conservation de la topologie de l'espace des données, l'interaction du neurone ne se limitant pas à son voisinage direct (appelé aussi voisinage d'ordre 1 car il désigne les neurones directement connectés).

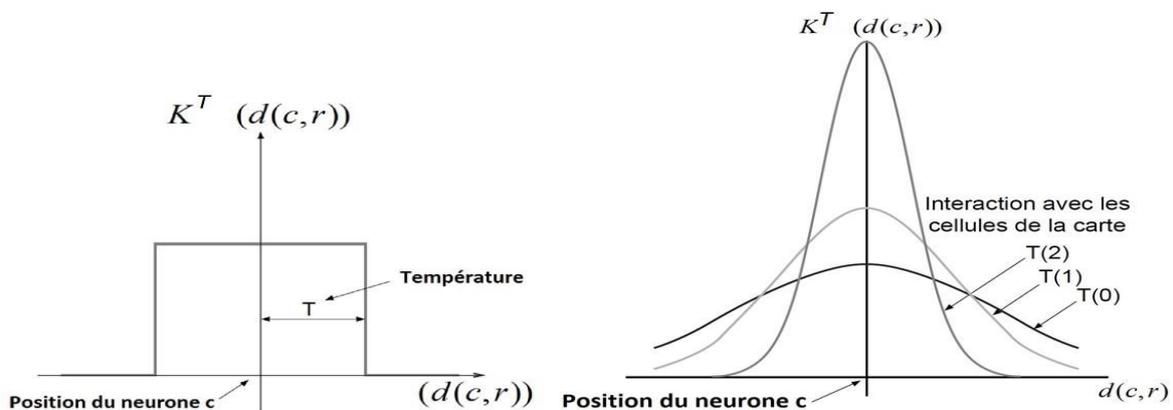


Figure 3.2. Les fonctions de voisinages les plus usuelles : à seuil (à gauche) et gaussienne (à droite) (D'après Dreyfus et al., 2002)

La structure topologique de la grille bidimensionnelle a un rôle fondamental sur la notion de voisinage. Les structures les plus usuelles sont la grille hexagonale et la grille rectangulaire (Figure 3.3). Le choix de la structure de la carte va directement impacter le voisinage d'ordre d d'un neurone c , défini par :

$$Vc(d) = \{r \in C, d(c, r) \leq d\} \quad (\text{Eq. 3.2})$$

Pour toute paire de neurones (c, r) de C , $d(c, r)$ est définie comme le plus court chemin entre c et r sur C .

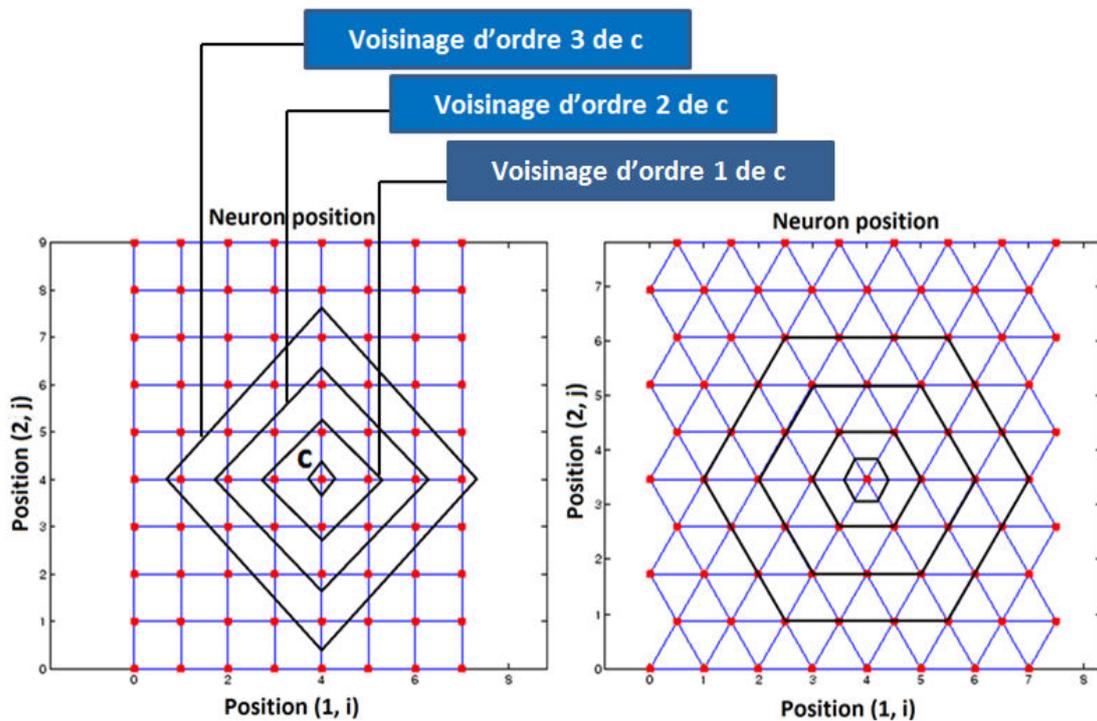


Figure 3.3. Le voisinage dans une carte topologique avec une structure rectangulaire (à gauche) et hexagonale (à droite) (D'après Dreyfus et al., 2002)

3.2.3.2. Apprentissage

Le but de la phase d'apprentissage est de générer un système de « vecteurs référents ». A la fin de cette phase, chaque neurone de la carte est représenté par un vecteur référent qui résume l'information classée par ce neurone. Les vecteurs référents doivent représenter l'ensemble de la base de données d'apprentissage (ils constituent un résumé des observations) tout en conservant les propriétés topologiques des données : deux neurones voisins dans l'espace de la carte doivent représenter deux observations similaires dans l'espace des données d'apprentissage.

Il existe plusieurs versions de l'algorithme d'apprentissage des cartes topologiques. Nous présentons ici la forme « Batch » utilisée dans cette thèse. A la différence de l'algorithme « séquentiel » qui procède à l'apprentissage de la carte topologique en mettant à jour les poids après chaque présentation d'une séquence de données (c'est un algorithme incrémentiel), l'algorithme « Batch » sélectionne l'ensemble des données et les ajuste au mieux dans la grille bidimensionnelle, les poids étant ajustés à la fin de la présentation de l'ensemble des données d'apprentissage.

L'apprentissage se fait en un nombre d'itérations prédéfini, une itération correspondant à la présentation de l'ensemble des observations au réseau. L'apprentissage se fait en 3 principales étapes. La première est une étape d'initialisation durant laquelle l'ensemble des référents initiaux w_0 sont définis (en général, d'une manière aléatoire). Le choix de la structure de la carte ainsi que des valeurs des paramètres d'apprentissage (e.g. les valeurs de température minimale et maximale, T_{\min} et T_{\max} , ainsi que le nombre d'itérations N_{iter}) est aussi fait durant cette phase. A étant l'ensemble des observations et c un neurone particulier de la carte C affecté à l'observation z_i par l'intermédiaire de la fonction d'affectation $\chi(z_i)$, l'apprentissage cherche à minimiser une fonction de coût. Cette fonction doit être minimisée par rapport aux deux composantes χ et W . Elle est exprimée par :

$$J_{som}^T(\chi, W) = \sum_{z \in A} \sum_{c \in C} K^T(d(c, \chi(z_i))) \|z_i - w_c\|^2 \quad (\text{Eq. 3.3})$$

où W désigne l'ensemble des vecteurs référents estimés durant la phase d'apprentissage, w_c le vecteur référent reliant la couche d'entrée au neurone c . Pour un choix d'un ensemble de référents W et d'une observation z_i , la fonction d'affectation qui minimise J_{som}^T (pour W fixé) est donnée par la relation suivante :

$$\chi(z_i) = \arg \min_c \|z_i - w_c\|^2 \quad (\text{Eq. 3.4})$$

où w_c désigne le vecteur poids ou vecteur référent reliant la couche d'entrée au neurone c du treillis (Figure 3.1). Durant cette étape d'affectation, chaque observation z_i présentée en entrée va être affectée à un neurone de la carte. Ce neurone sera appelé neurone gagnant ou BMU (Best Matching Unit).

Une fois le neurone gagnant déterminé, l'observation lui sera associée. La troisième étape est l'étape d'adaptation pendant laquelle les neurones modifient leur poids. L'auto-organisation permet l'adaptation des poids au fur et à mesure des itérations et en fonction de la valeur de la température à l'instant t . A chaque présentation d'une observation z_i , les nouveaux référents sont calculés pour tous les neurones de la carte en fonction de leur distance par rapport au neurone sélectionné lors de l'étape d'affectation (neurone gagnant) selon l'équation suivante :

$$w_c^t = w_c^{t-1} + \alpha(t) K^T (d(c, \chi_t(z_i))) (z_i - w_c^{t-1}) \quad (\text{Eq. 3.5})$$

où w_c^{t-1} désigne l'ancien référent, $\alpha(t)$, le pas d'apprentissage (valeur fixée initialement) qui décroît à chaque itération et $\chi_t(z_i)$, la fonction d'affectation d'une observation z_i au neurone gagnant à l'instant t .

1. Phase d'initialisation

- Choisir la structure et la taille de la carte et l'ensemble des référents initiaux (en général, d'une manière aléatoire) ;
- Fixer les valeurs de température T_{\max} , T_{\min} et le nombre d'itérations N_{iter} à $t=0$.

2. Etape itérative : l'ensemble des référents w^{t-1} de l'étape précédente étant connus :

- Choisir une observation z_i (en général d'une manière aléatoire) ;
- Calculer la nouvelle valeur de T en appliquant la formule :

$$T = T_{\max} * \left(\frac{T_{\min}}{T_{\max}} \right)^{\frac{t}{N_{\text{iter}} - 1}} \quad \text{Eq. 3.6}$$

Pour cette valeur du paramètre T , effectuer les deux phases suivantes :

- Phase d'affectation : On suppose w^{t-1} connu, on affecte l'observation z_i au neurone $\chi_t(z_i)$ défini à partir de la fonction d'affectation (Eq. 3.4) ;
- Phase de minimisation : calcul de l'ensemble des nouveaux référents w^t ; les vecteurs référents sont modifiés selon l'équation 3.5 en fonction de leur distance par rapport au neurone sélectionné à l'étape d'affectation.

3. Répéter l'étape itérative en faisant décroître la valeur du paramètre T jusqu'à $t=N_{\text{iter}}$

En fin d'apprentissage, les poids de chaque neurone convergent vers des valeurs telles qu'un neurone ne sera plus actif que pour un sous-ensemble bien déterminé d'observations de la base d'apprentissage. Chaque neurone c a « capté » un sous-ensemble de données représenté à l'aide d'un vecteur référent. Le vecteur référent n'est autre que le vecteur poids w_c arrivant au neurone c en fin d'apprentissage.

L'apprentissage effectué par les cartes auto-organisatrices vise à ce que les vecteurs référents captent au mieux la densité de probabilité sous-jacente aux observations. Le réseau va approximer les relations de voisinage existant entre les différents sous-ensembles des données d'apprentissage. A la fin de cette phase, la carte auto-organisatrice est un résumé ordonné des données étudiées et on peut parler de « quantification vectorielle » de l'espace des données d'entrée puisque l'algorithme de SOM compresse les valeurs d'un espace vectoriel multidimensionnel continu vers un nombre discret d'éléments de même dimension. Le choix de la forme et de la dimension du treillis se fait a priori et dépend de la taille du jeu de données d'apprentissage, de sa complexité ainsi que d'autres paramètres, dont les valeurs servent à évaluer et optimiser la classification faite par SOM. Ces paramètres concernent la conservation de la topologie et les erreurs de quantification vectorielle à la fin de l'apprentissage. Ils seront définis en détail plus loin dans ce chapitre.

L'apprentissage se termine lorsque la phase d'affectation des spectres de l'espace d'entrée s'est faite de façon assez parcimonieuse (tous les référents ont capté un nombre significatif de données et s'organisent de façon cohérente sur la carte à la fin de l'apprentissage), que l'erreur de quantification est relativement faible et que l'espace des données est bien représenté par les référents (chaque partition associée à un référent suit une loi normale locale).

Les cartes auto-organisatrices permettent de faire une représentation simplifiée des données décrites dans un espace de grande dimension. Cette représentation se fait dans un espace discret C de petite dimension qui est défini par une structure de graphe. Le problème est donc de choisir une structure de carte bien adaptée au problème particulier qui est traité. Pour ce faire, il faut définir les critères qui permettent de juger l'adéquation d'une carte à un problème donné. L'affectation d'une observation à un vecteur référent d'un neurone de la carte topologique doit vérifier la conservation de la topologie.

Sur le plan pratique :

- La durée de l'apprentissage varie en fonction de la nature et de la dimension des données. Elle peut être également sensible aux conditions initiales dans la mesure où il s'agit d'une minimisation itérative d'une fonction de coût. Différentes méthodes d'initialisation sont disponibles dans la littérature pour le choix des valeurs initiales des prototypes : de nombreuses variantes existent (Kohonen, 1995, 1997, 2001; Thiria et al., 1997 ; Dreyfus et al., 2002).
- La méthode est particulièrement robuste : elle permet une gestion efficace des grandes bases de données bruitées et le résultat n'est pas grandement modifié par l'ajout d'un nouvel élément ou par des données manquantes. Cette propriété n'est par exemple pas partagée par la classification ascendante hiérarchique, dont le résultat peut être remis en cause par la suppression ou l'apport d'une donnée supplémentaire. Avec les cartes de Kohonen, pour « classer » un vecteur avec des données manquantes, il suffit de trouver le neurone gagnant en utilisant les composantes du vecteur référent pour lequel on a des données. Si on est dans la phase de création de la carte, on ne doit modifier que les composantes du vecteur référent correspondant aux composantes non manquantes.
- Les logiciels et programmes utilisés permettent d'avoir un contrôle fin sur tous les paramètres décrits précédemment. La « SOM toolbox » (<http://www.cis.hut.fi/projects/somtoolbox/>) a été utilisée. C'est une boîte à outils comprenant plusieurs programmes qui fonctionnent sous le langage de programmation « MATLAB[®] ». Plusieurs fonctions sont disponibles pour la création de structures de données. Ces structures peuvent inclure des informations concernant les données d'apprentissage, les propriétés de la carte auto-organisatrice, les paramètres d'initialisation et d'entraînement utilisés, etc. La boîte à outil rassemble aussi des fonctions qui permettent le prétraitement des données (e.g. normalisation), l'évaluation de la qualité de l'apprentissage ainsi qu'un ensemble d'outils graphiques permettant de pouvoir interpréter les groupements formés à la fin de l'apprentissage de façon aisée.

Il est essentiel de souligner que la capacité de généralisation de la carte topologique entraînée, c'est-à-dire à bien classer des données nouvelles, dépend de la qualité de l'apprentissage (quantification vectorielle et préservation de la topologie des observations) et

de la base de données utilisée pour l'apprentissage. Cette dernière doit être représentative de la variabilité du phénomène observé. Si ce n'est pas le cas, une observation parmi l'ensemble des données qu'on souhaite étudier pourrait être mal ou non classée. Dans ce cas, l'exploitation de données nouvelles pourrait être significativement limitée ou biaisée.

Dans le cas où un échantillonnage des données est nécessaire (réduction des temps de calcul et de la mémoire nécessaires), il est primordial que ce dernier soit pertinent et que sa fréquence soit adéquate, de telle façon à ce que le jeu de données d'apprentissage soit représentatif de la variabilité du (des) paramètre (s) étudié (s). Une connaissance globale de la variabilité du (des) paramètre (s) qu'on veut étudier est donc nécessaire afin de procéder à un échantillonnage adéquat.

3.2.4. Evaluation de la qualité de l'apprentissage

La qualité de la classification faite avec SOM passe par une bonne discrétisation du paramètre étudié. Avant de procéder à l'exploitation de la classification, il est important d'évaluer la qualité de la classification obtenue en fin d'apprentissage. Cette évaluation peut se faire selon deux principaux critères : la qualité de la quantification vectorielle et celle de la préservation de la topologie de l'espace des observations.

3.2.4.1. L'erreur de quantification vectorielle

La qualité de la quantification vectorielle donne une information sur l'adéquation entre chaque sous-ensemble de données et le vecteur référent auquel il a été affecté. Cette information peut être quantifiée grâce à la mesure de l'erreur de quantification vectorielle, notée Q.E (pour Quantization error). Cette dernière représente la distance euclidienne moyenne entre chaque observation et son plus proche vecteur référent. Soit n le nombre total d'échantillons à classer, z_i une observation donnée et w_c son référent le plus proche, l'erreur de quantification vectorielle est définie par :

$$Q.E = \frac{1}{n} \sum_{i=1}^n \|z_i - w_c\| \quad (\text{Eq. 3.7})$$

La qualité de la quantification vectorielle est donc d'autant plus élevée que cette distance est faible. Si la valeur de ce paramètre est relativement élevée, ceci révèle une faible adéquation (ou ressemblance) entre l'observation et son vecteur référent. L'erreur de quantification vectorielle atteint un minimum après un certain nombre d'itérations. La valeur

finale de cette erreur ainsi que le nombre d'itérations nécessaires pour l'atteindre dépendent de plusieurs facteurs, tels que le nombre et la distribution des données, la structure et la taille de la carte (nombre de neurones) ainsi que l'adéquation entre le nombre de données à classer et la taille de la carte auto-organisatrice.

3.2.4.2. L'erreur topologique

L'erreur topologique quantifie la préservation de la topologie de l'espace des observations. Elle représente la proportion des observations pour lesquelles les deux neurones les plus proches (correspondant aux deux plus proches vecteurs référents au sens de la distance euclidienne) ne sont pas adjacents au niveau de la carte topologique :

$$T.E = \frac{1}{n} \sum_{i=1}^n u(z_i) \quad \text{(Eq. 3.8)}$$

où $u(z_i)$ représente une fonction dont le résultat est égal à 1 si les deux neurones les plus proches (le premier BMU et le suivant) de l'observation z_i sont adjacents au niveau de la carte topologique et 0 sinon.

Ce paramètre évalue donc la fiabilité de la conservation topologique induite par l'algorithme SOM. La préservation de la topologie de l'espace des observations permet la conservation de la notion de similitude des différents sous-ensembles de données et une structuration claire de ces partitions. Cette structuration des vecteurs référents facilite potentiellement la tâche de classification. L'erreur topologique obtenue en fin d'apprentissage peut varier significativement en fonction de plusieurs facteurs, dont la structure du voisinage (hexagonale ou rectangulaire) ainsi que la forme du treillis (carré ou rectangulaire).

Afin de minimiser les erreurs topologiques, il convient de choisir une structure de carte bien adaptée aux données qu'on cherche à classer et au problème qu'on cherche à résoudre. Il faut définir les critères qui permettent de juger de l'adéquation d'une carte au problème donné. Si la dimension de la carte n'est pas en adéquation avec la dimension sous-jacente aux données (variété des classes engendrées par les observations), deux observations proches pourraient être classées par des neurones éloignés de la carte. Or, la propriété de conservation de la topologie est intéressante car elle permet de réaliser une réduction de la dimension des données en les projetant dans un espace plus petit, tout en conservant la notion de similitude.

3.2.4.3. La distorsion

La distorsion représente la somme des erreurs de quantification vectorielle, des erreurs topologiques ainsi que la contrainte qui existe entre ces deux paramètres (Vesanto et al., 1999). Cette contrainte, appelée aussi «biais de voisinage» établie le lien entre l'erreur topologique et celle de la quantification vectorielle (Vesanto, 1999). L'expression générale de la distorsion est la suivante :

$$D = \sum_{i=1}^n \sum_{j=1}^m K^T_{(j, \chi(z_i))} \|z_i - w_j\|^2 \quad (\text{Eq. 3.9})$$

où $K^T_{(j, \chi(z_i))}$ représente la fonction de voisinage entre un neurone j et le neurone le plus proche d'une observation z_i déterminé par la fonction d'affectation $\chi(z_i)$ et w_j représente le vecteur référent du neurone j .

La mesure de la distorsion moyenne de tous les neurones doit être minimisée et permet de quantifier la qualité globale de la classification faite par SOM, en tenant compte à la fois de la qualité de la quantification vectorielle et de celle de la préservation de la topologie. La valeur de la distorsion à la fin de l'apprentissage va donc aussi dépendre des paramètres qui influent sur les erreurs topologiques et de quantification vectorielle. Un choix pertinent du nombre de neurones et de la forme de la carte est nécessaire pour minimiser la distorsion de la carte topologique. Par ailleurs, l'optimisation du nombre d'itérations contribue aussi à la diminution de la distorsion en fin de phase d'apprentissage. La figure 3.4 (exemple d'un apprentissage réalisé sur une base de données générée aléatoirement) montre qu'on peut faire diminuer l'erreur de quantification vectorielle jusqu'à une valeur de 0,208 avec 150 itérations. Cependant, un grand nombre d'itérations pourrait induire un «surapprentissage» de SOM et causer une augmentation des erreurs topologiques. Le surapprentissage ou sur-ajustement est, d'une manière générale un problème pouvant survenir dans les méthodes mathématiques et informatiques de classification, à l'instar des réseaux de neurones. Il est en général provoqué par un mauvais dimensionnement de la structure utilisée pour classifier. De par sa trop grande capacité à stocker des informations, une structure dans une situation de surapprentissage aura de la peine à généraliser les caractéristiques des données. Elle se comporte alors comme une table contenant tous les échantillons utilisés lors de l'apprentissage et perd ses pouvoirs de prédiction sur de nouveaux échantillons. Le nombre d'itération doit par conséquent être choisi de telle manière à optimiser la distorsion en trouvant un équilibre entre les erreurs

topologiques et de quantification vectorielle. La distorsion peut aussi être diminuée en augmentant le degré de flexibilité de la carte auto-organisatrice. Pour ce faire, on agit sur la taille du voisinage en faisant décroître la température. Ce paramètre contrôle la fonction de voisinage de SOM et aura un impact déterminant sur la qualité de la classification neuronale. En règle générale, des valeurs élevées de température entraînent toutes les données dans leur centre de gravité et engendrent des contraintes excessives sur l'ordre topologique aux dépens de la qualité de quantification vectorielle.

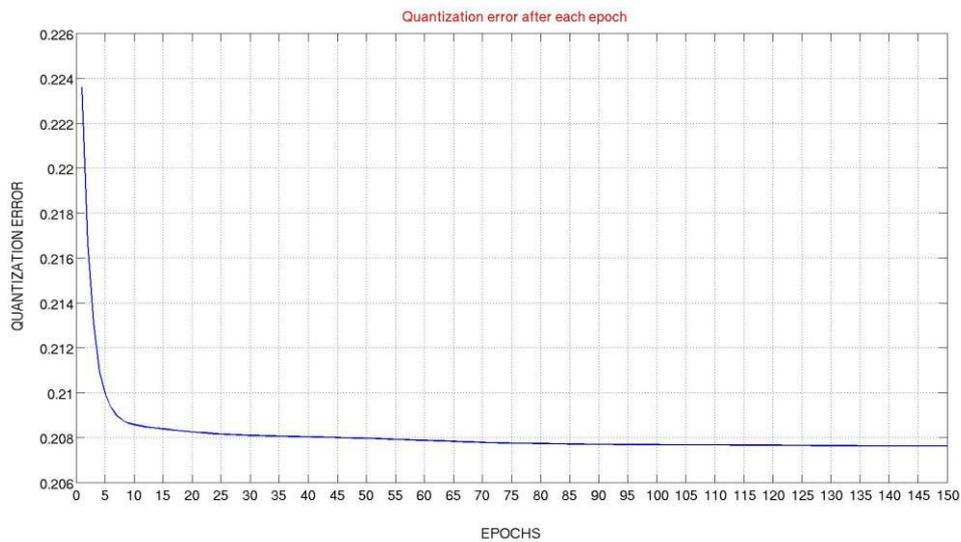


Figure 3.4. Evolution de l'erreur de quantification vectorielle en fonction du nombre d'itérations (exemple d'un apprentissage réalisé sur une matrice de données à 5 dimensions générée aléatoirement)

3.3. Exploitation de la carte topologique en fin d'apprentissage

Une fois l'apprentissage de la carte topologique accompli, il est possible de l'exploiter en «**généralisant**» cette classification à d'autres données. Cette généralisation s'effectue en procédant à une «**projection**» de l'ensemble de données qu'on veut exploiter (pour des fins de classification par exemple) sur les vecteurs référents générés en fin d'apprentissage de la carte auto-organisatrice. Cette projection est faite par l'intermédiaire de la fonction d'affectation (équation 3.4). Chaque observation va être affectée à son neurone gagnant (BMU) correspondant au vecteur référent le plus proche au sens de la distance euclidienne. Ainsi, l'ensemble des données qu'on veut exploiter va être réparti sur les neurones de la carte. Pour un neurone donné, le vecteur référent résume l'information sous-jacente à une portion (classe) de l'ensemble d'observations.

L'auto-organisation, telle que nous venons de la décrire, ne permet pas de résoudre directement les tâches de classification: le résultat de l'apprentissage non supervisé permet d'affecter une observation à un sous-ensemble d'une partition, indépendamment de toute notion de classe. On considère que l'on dispose d'un très grand nombre d'observations bruitées dont on ne connaît pas exactement la classe. La partition proposée dépend de la densité de probabilité qui est sous-jacente à l'ensemble des observations de l'ensemble d'apprentissage. Les régions ayant une forte densité d'observations vont être décrites par une sous-partition fine, celles de faible densité par une sous-partition plus grossière (Figure 3.5). La grande quantité de données disponibles dans les régions de forte densité permet d'obtenir une information plus précise pour ces régions.

L'idée principale de l'algorithme est de faire apparaître des groupements (les sous-ensembles de la partition) en relation avec l'application traitée. Si l'application envisagée consiste en un problème de classification en n classes, les différents sous-ensembles doivent s'intégrer au mieux avec celle-ci. On cherche alors à affecter chaque sous-ensemble de la partition à l'une des n classes. Puisque chaque sous-ensemble est associé à un neurone de la carte, le problème de classification se résume à celui de l'«étiquetage» ou la «labellisation» de chaque neurone de la carte.

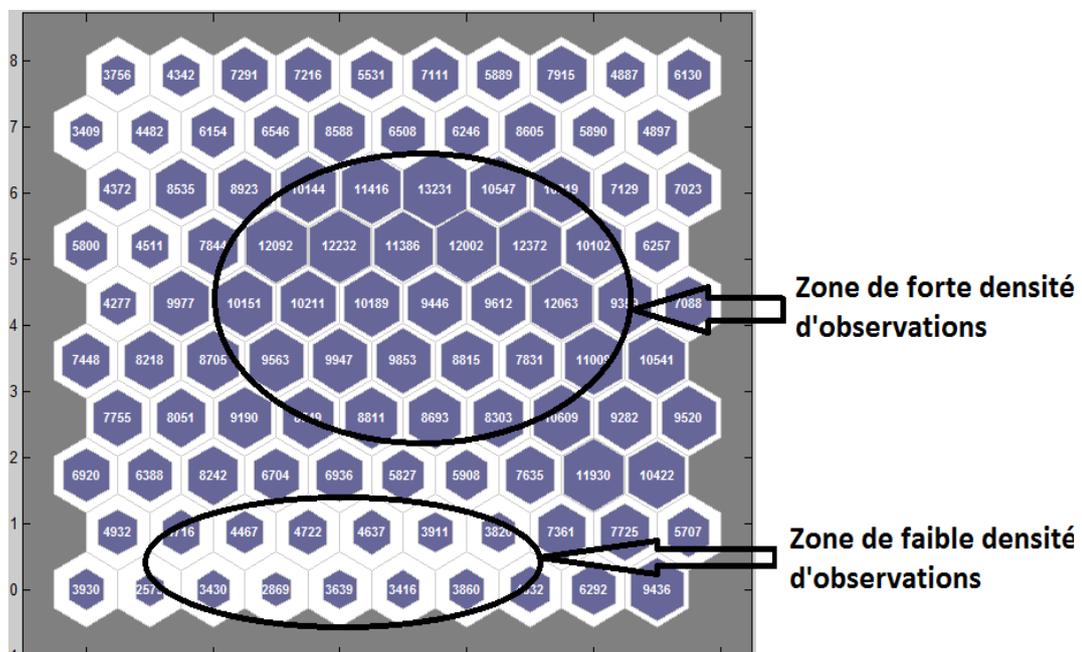


Figure 3.5. Représentation topologique d'une carte auto-organisatrice à 2 dimensions en fin d'apprentissage avec le nombre d'observations affecté à chaque neurone, montrant des zones à forte densité d'observations (décrites par une sous-partition fine) et d'autres avec une faible densité d'observations (décrites par une sous-partition plus grossière)

La labellisation peut se faire de 3 manières différentes :

1. Puisque chaque sous-ensemble d'observations est représenté par un vecteur référent et que celui-ci est une observation moyenne, il est possible de demander à un expert du domaine d'application de reconnaître, grâce aux caractéristiques de cette observation, la classe à laquelle elle appartient.
2. En regroupant les neurones de la carte d'une manière statistique (avec des méthodes de classification multivariée par exemple), le recours à l'expertise pour l'étiquetage des différentes classes ne se faisant qu'à l'issue de cette phase. L'analyse ascendante hiérarchique représente un moyen efficace pour regrouper les neurones de la carte en se basant sur un critère de similarité (ou de dissimilarité) entre les sous-ensembles de données classées par chaque neurone. L'ordre topologique permet de supposer qu'il existe une structure dans les données qui est sous-jacente au problème de classification. Il est possible avec l'ordre topologique de la carte d'exhiber cette structure : deux sous-ensembles représentés par des neurones voisins ont une forte probabilité de représenter des observations appartenant à la même classe. Cependant, cette hypothèse est forte et sous-entend qu'un codage adéquat des données a été utilisé pour effectuer la classification. Cela suppose qu'une étude préalable a été faite dans le but d'obtenir une bonne représentation des observations, et donc, une sélection adéquate des variables et un codage pertinent pour le problème de classification traité.
3. En utilisant des données expertisées. Ce sont des données qui font partie d'une base dont le contenu a déjà été classé soit par une expertise directe, soit en utilisant des critères de classification. Les différentes classes sous-jacentes à ces données étant connues, il est possible d'utiliser ces classes pour étiqueter les différents neurones de la carte auto-organisatrice.

La 3^{ème} possibilité est celle qui a été utilisée dans le cadre de cette thèse pour labelliser les anomalies spectrales automatiquement classées avec SOM. L'approche de labellisation sera détaillée dans le chapitre 4.

3.4. Application de SOM à la classification automatique des anomalies spectrales

3.4.1. Echantillonnage des données d'entrée et apprentissage de SOM

Le jeu de données d'apprentissage utilisé pour classer les anomalies spectrales à l'échelle globale comprend 714 264 spectres de R_a , tels que définis par Alvain et al. (2005), obtenus par le capteur SeaWiFS, à 5 longueurs d'onde dans le domaine du visible : 412, 443, 490, 510 et 555 nm. Ce jeu de données d'apprentissage a été obtenu après avoir procédé à un échantillonnage des spectres de R_a calculés pour toute la période d'observation du capteur SeaWiFS en prenant 2 jours par mois et un pixel sur 10. Les 2 jours ont été sélectionnés d'une manière régulière avec un intervalle de 15 jours. Cet échantillonnage a été fait pour chaque année entre 1998 et 2010. La normalisation des données n'a pas été nécessaire car les anomalies spectrales résultent déjà d'une normalisation de la luminance marine par un signal moyen de chlorophylle, ce qui réduit préalablement certains écarts dans les mesures satellitaires.

Différents tests d'apprentissage de SOM ont été effectués en utilisant séparément les sous-ensembles de spectres d'anomalies de luminance de toutes les années entre 1998 et 2010 (chaque sous-ensemble correspondant à une année échantillonnée tel que précisé ci-dessus). Le but de ces tests est de voir s'il existe des différences significatives en termes de formes et d'amplitudes entre les 100 vecteurs référents générés par l'apprentissage de SOM avec les différents sous-ensembles. Ceci n'ayant pas été le cas (voir en annexe la comparaison entre les vecteurs référents des différentes années testées), la phase d'apprentissage a été réalisée avec les données d'une seule année (2002, année centrale) qui sont donc représentatives de la diversité des spectres de R_a existante sur l'archive satellite traitée. Les critères sur l'épaisseur optique des aérosols et la concentration en chlorophylle a , définis dans le chapitre 2, ont de même été appliqués afin d'être en accord avec les valeurs utilisés pour le développement de la version 2008 de PHYSAT (Alvain et al., 2005 ; Alvain et al., 2008). Cela permet de faciliter la comparaison des deux approches.

Pour cette étude, nous avons choisi d'utiliser une carte auto-organisatrice carrée avec 100 neurones (10x10 neurones). Les spectres de R_a sont donc classés en 100 groupes distincts selon leur forme et amplitude, permettant ainsi d'avoir une discrétisation très précise de l'ensemble des observations. Le choix de la forme de la carte (carrée ou rectangulaire) ainsi

que du nombre de neurones a été établi de sorte à minimiser les erreurs topologiques, de quantification vectorielle ainsi que la distorsion moyenne à la fin de l'apprentissage de SOM. Ainsi, afin de déterminer la carte SOM la plus optimale à notre problème, plusieurs tests ont été établis en faisant varier la taille de la carte (nombre de neurones) en gardant la même forme (carrée ou rectangulaire), puis, en faisant varier la forme de la carte tout en maintenant le même nombre de neurones. Pour chaque configuration de SOM, les erreurs topologiques, de quantification vectorielle ainsi que la distorsion moyenne de tous les neurones ont été calculés. Le tableau 3.1 rassemble les valeurs de ces paramètres pour chaque configuration de SOM testée. La distorsion moyenne est un indicateur à prendre particulièrement en considération car il résume la qualité globale de SOM en prenant en compte les erreurs topologiques et de quantification vectorielle ainsi que le lien entre ces deux paramètres. En considérant ce paramètre, la carte SOM optimale est celle ayant une forme rectangulaire et une taille de 10x10 neurones. Cette configuration est aussi celle qui donne les erreurs topologiques et de quantification vectorielle (respectivement 0,012 et 0,254) les plus faibles comparées aux autres configurations.

Tableau 3.1. Valeurs des paramètres d'erreur (Q.E pour erreur de quantification vectorielle, T.E pour erreur topologique et A.D pour distorsion moyenne) pour les différentes configurations de SOM testées

	4x4	8x2	8x8	16x4	20x5	10x10
Q.E	0,313	0,326	0,308	0,321	0,267	0,254
T.E	0,027	0,047	0,032	0,059	0,067	0,012
A.D	0,9443	0,8132	0,8039	0,7985	0,7585	0,73

De même, la carte de 10x10 neurones représente le choix idéal parmi les différentes configurations testées étant donné qu'on cherche à avoir une classification fine des anomalies spectrales indépendamment du nombre de groupes de phytoplancton qui sera détecté (ce dernier étant relié aux données *in situ* et susceptibles d'évoluer vers un plus grand nombre à l'avenir). Les cartes auto-organisatrices sont utilisées, en premier lieu, dans le but d'avoir une classification précise des anomalies spectrales, dont les formes et amplitudes dépendent de plusieurs facteurs et qui représentent une information spectrale potentiellement utilisable (Alvain et al., 2012). La labellisation des neurones en termes de groupes de phytoplancton

représente une application finale de cette classification. Le choix d'un grand nombre de neurones permet ainsi de garder une certaine flexibilité quand il s'agira d'exploiter de grandes bases de données de pigments ou autres types d'informations *in situ*.

3.4.2. Variabilité des anomalies spectrales à l'échelle globale

Avec la configuration de SOM choisie, la classification des spectres de Ra donne une répartition équilibrée des données sur la carte auto-organisatrice, ce qui dénote d'une bonne qualité d'apprentissage. Les vecteurs référents classent un maximum de 14245 spectres (neurone 100) et un minimum de 2941 spectres (neurone 19). Le nombre de spectres associés à chaque neurone à la fin de l'apprentissage est affiché dans le tableau 3.2. Les spectres référents sont affichés sur la figure 3.6 qui montre l'ordre topologique induit par l'algorithme SOM à la fin de la phase d'apprentissage. La fonction de voisinage impose que les spectres voisins au niveau de la carte auto-organisatrice soient similaires en termes de forme et d'amplitude et vice-versa. Chaque spectre référent représente une synthèse de la variabilité de tous les spectres de Ra classés par le neurone correspondant.

Tableau 3.2. Nombre de spectres de Ra associés à chaque neurone à la fin de la phase d'apprentissage, le numéro des neurones est indiqué entre parenthèses

(1) 9824	(11)7341	(21)9987	(31)6845	(41)5059	(51)5223	(61)8513	(71)13950	(81)9672	(91)12431
(2) 5495	(12)4360	(22)6428	(32)6460	(42)6337	(52)6404	(62)6857	(72)7377	(82)4600	(92)6307
(3) 7970	(13)6344	(23)8922	(33)8585	(43)8490	(53)8556	(63)9128	(73)10223	(83)7328	(93)10066
(4) 6075	(14)5708	(24)8021	(34)7864	(44)7860	(54)7978	(64)8610	(74)8891	(84)6295	(94)7819
(5) 5368	(15)5177	(25)7579	(35)7668	(45)7541	(55)7784	(65)8153	(75)8351	(85)5771	(95)6439
(6) 5588	(16)5177	(26)7411	(36)7594	(46)7324	(56)7641	(66)7577	(76)8240	(86)5612	(96)6602
(7) 7586	(17)5261	(27)7703	(37)7417	(47)6764	(57)6667	(67)6756	(77)8294	(87)6320	(97)8845
(8) 8964	(18)5192	(28)8195	(38)7434	(48)5888	(58)5679	(68)5820	(78)8410	(88)7133	(98)10262
(9) 4663	(19)2941	(29)5454	(39)5483	(49)4040	(59)4506	(69)4582	(79)5944	(89)4730	(99)5121
(10)8988	(20)4485	(30)8397	(40)5205	(50)3003	(60)7666	(70)8132	(80)7993	(90)5296	(100)14245

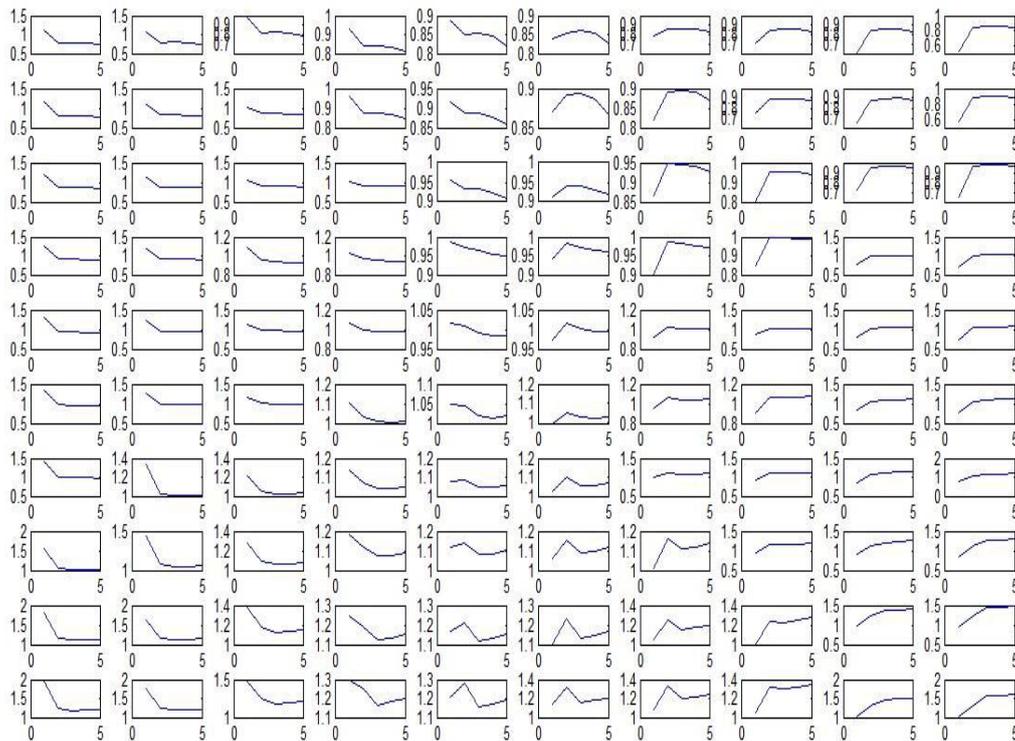


Figure 3.6. Représentation des spectres référents correspondant aux Ra associés à chaque neurone à la fin de la phase d'apprentissage. Pour chaque unité, l'axe des abscisses représente les longueurs d'onde et l'axe des ordonnées représente les Ra. Les axes des ordonnées n'ont pas été mis à la même échelle afin de mettre en évidence les différentes formes des spectres

La figure 3.7 représente la projection des composantes des spectres référents comparées entre chacune des 5 longueurs d'onde prises en compte deux à deux. Puisque ces spectres référents (figure 3.6) représentent une synthèse (ou une compression) de l'information contenue dans le jeu de données d'apprentissage, les relations présentés dans la figure 3.7 sont représentatives de la variabilité des anomalies spectrales à l'échelle globale.

La figure montre que Ra (490), Ra (510) et Ra (555) sont globalement corrélés alors qu'elles sont significativement différentes de Ra (412) et Ra (443). Les valeurs des anomalies spectrales à 412 nm sont particulièrement différentes de toutes les autres longueurs d'onde. Cette observation est aussi mise en évidence par les spectres référents affichés dans la figure 3.6 qui montre que, pour un spectre donné, les plus grandes variations de Ra se situent entre les courtes longueurs d'onde (412 et 443 nm) et les plus grandes longueurs d'ondes (où les valeurs de Ra demeurent relativement constantes).

Comme cela a été expliqué dans le chapitre 2, la forme et l'amplitude des spectres de Ra dépendent principalement de l'environnement bio-optique. D'après l'étude théorique

d'Alvain et al. (2012), certaines situations spécifiques de dominances montrent des formes de Ra sans corrélations semblables à celles retrouvées globalement. C'est-à-dire que les valeurs de Ra à 490, 510 et 555 nm ne présentent pas forcément la même variabilité pour certains groupes de phytoplancton (Alvain et al., 2012). Garder toute l'information spectrale est par conséquent nécessaire d'autant plus que les corrélations montrées par la figure 3.7 concernent les spectres référents et ne représentent donc qu'une synthèse globale de l'information contenue dans la base de données d'anomalies spectrales. Ainsi, même si globalement on retrouve des corrélations fortes entre certaines longueurs d'onde, le choix a été fait de toutes les conserver en raison des cas spécifiques possibles qui pourraient être labellisé à partir de nouvelles mesures *in situ* à l'avenir. Ces analyses montrent que la classification faite par SOM en utilisant 100 neurones donne une répartition équilibrée des spectres de Ra sur la carte auto-organisatrice ainsi qu'une discrimination précise de la variabilité spectrale en termes de formes et d'amplitudes. La prochaine étape de la classification est supervisée. Elle consiste à établir le lien entre les spectres référents mis en évidence précédemment à la fin de la phase d'apprentissage et les groupes de phytoplancton identifiés à partir des données *in situ* de pigments. Cette étape, appelée labellisation, permettra par la suite d'exploiter les images satellitaires en associant chaque spectre de Ra à son vecteur référent le plus proche et générer ainsi des cartes de distribution globale des groupes phytoplanctoniques.

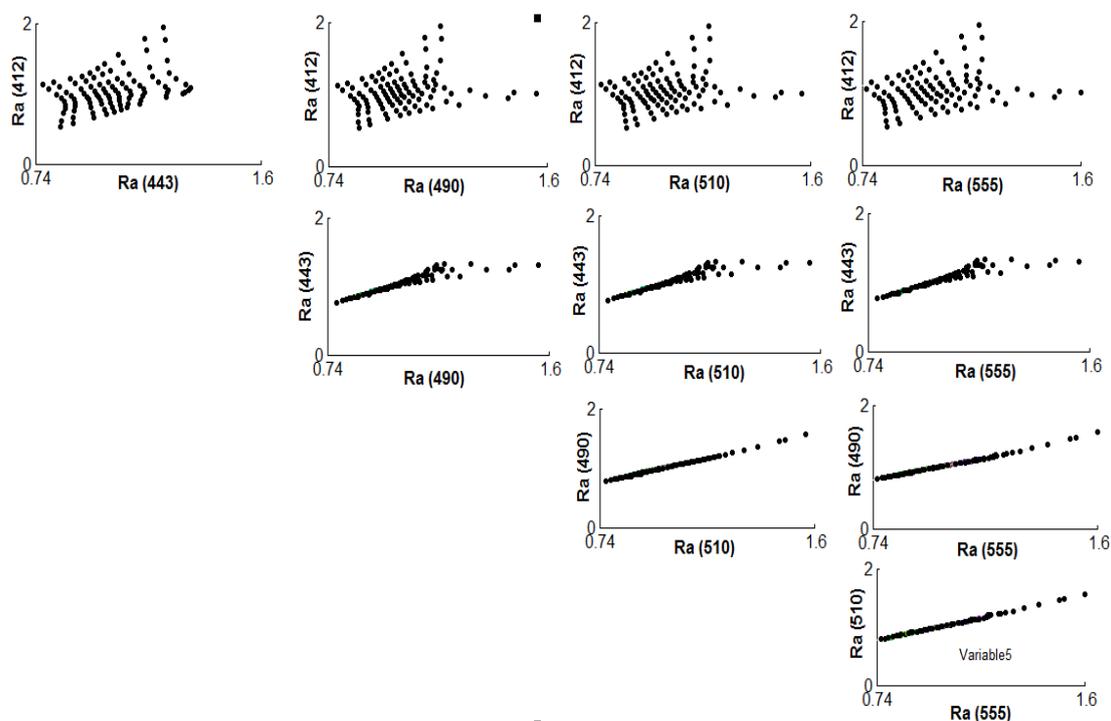


Figure 3.7. Projection des composantes des spectres référents comparées entre chacune des 5 longueurs d'onde prises en compte deux à deux

3.4.3. Amélioration de la caractérisation de la variabilité des anomalies spectrales avec SOM par rapport à l'approche actuelle de PHYSAT

Pour démontrer, d'un point de vue méthodologique, l'avantage d'utiliser les cartes auto-organisatrices par rapport à l'ancienne approche de PHYSAT (en termes de caractérisation potentielle de la variabilité des anomalies spectrales), nous nous sommes basées sur l'ensemble des 41 spectres de Ra utilisés par Alvain et al. (2005) pour définir la signature spectrale des groupes de phytoplancton. Avec l'ancienne approche décrite au chapitre 2, la moitié des spectres associés à une mesure *in situ* n'avait pas pu être valorisée lors de l'étape de labellisation en raison de la définition de tubes (figure 2.6 du chapitre 2). Pour évaluer l'amélioration apportée par l'approche par cartes auto-organisatrices, les 41 spectres initiaux de Ra ont été projetés sur la carte auto-organisatrice entraînée. Chaque spectre a été affecté à un neurone spécifique correspondant au vecteur référent le plus proche au sens de la distance euclidienne (figure 3.8). Les spectres de Ra qui n'ont précédemment pas été pris en compte par l'ancienne approche sont maintenant classés par SOM et il sera désormais possible de les rechercher plus aisément dans les cartes journalières d'anomalies spectrales. Avec cet outil il est donc possible de tirer profit de toute l'information spectrale contenue dans le même jeu de données utilisé pour développer l'ancienne approche.

L'erreur de quantification vectorielle associée à ces 41 spectres a été calculée dans le but d'évaluer leur classification par SOM. En effet, même si tous les spectres sont affectés à un (ou plusieurs) neurones, une grande divergence entre un spectre donné et son vecteur référent le plus proche augmente l'erreur de quantification et indique donc une mauvaise classification. Le tableau 3.3 montre que tous les spectres de Ra sont statistiquement bien classés car ils ont une Q.E (erreur de quantification vectorielle) inférieure à la Q.E maximale associée au neurone correspondant à la fin de la phase d'apprentissage (erreurs issues des spectres de la base de données d'apprentissage), dont 34 spectres avec une Q.E inférieure à la Q.E moyenne du neurone correspondant. Ils correspondent à ceux qui se rapprochent le plus des spectres référents.

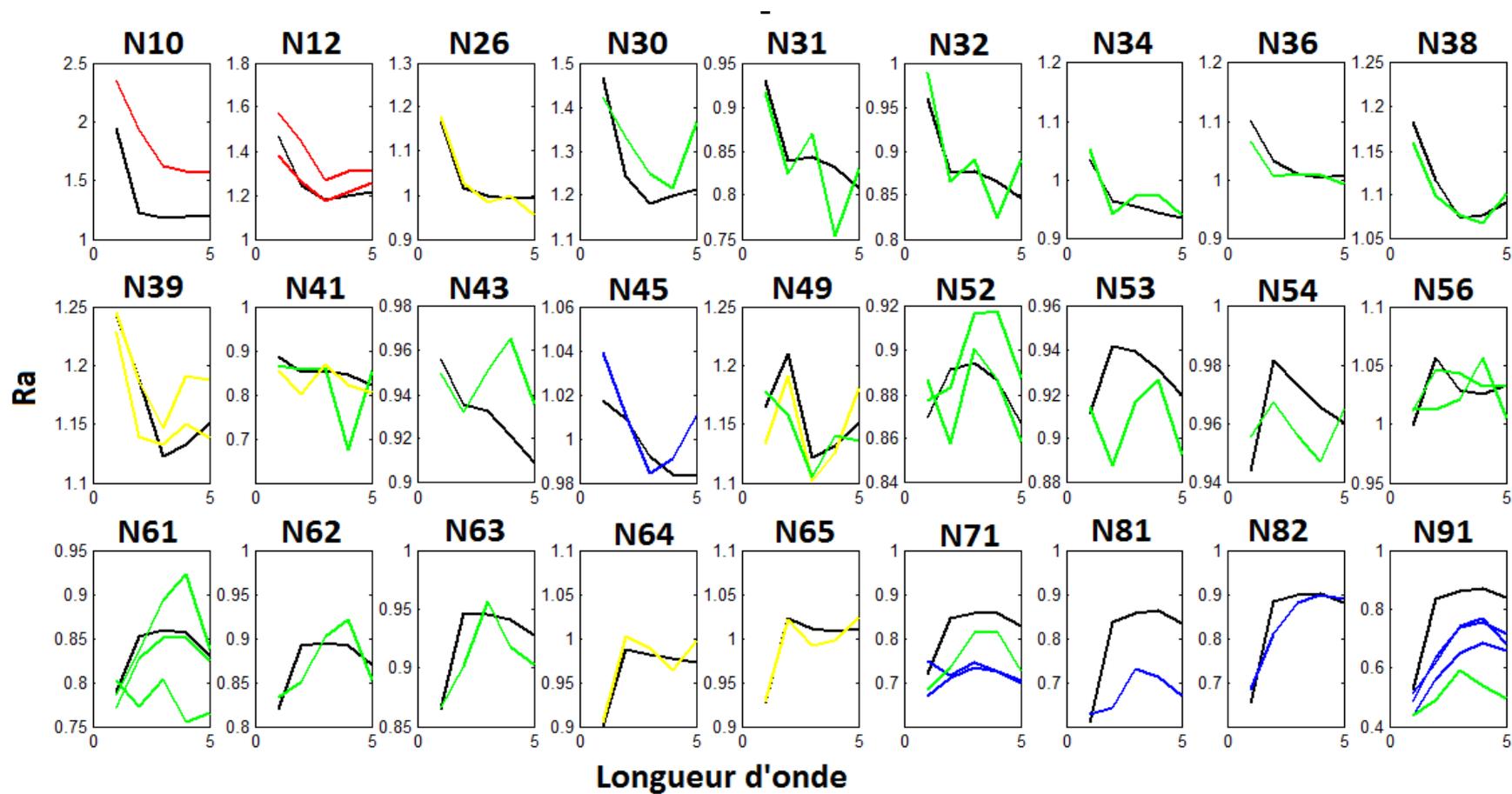


Figure 3.8. Projection des 41 spectres de Ra sur la carte auto-organisatrice. Seuls les neurones ayant « captés » des spectres sont représentés avec, en noir, leur spectre référent et en code couleur les spectres des 4 groupes de phytoplancton suivant : les nanoeucaryotes (bleu), *Prochlorococcus* (vert), SLC (jaune) et diatomées (rouge)

Tableau 3.3. Erreurs de quantification vectorielle minimales, maximales et moyennes associées aux neurones (à la fin de l'apprentissage) sur lesquels les 41 spectres de Ra ont été projetés. La Q.E associée à chaque spectre projeté est affichée dans la 5^{ème} colonne.

Neurone	Q.E min	Q.E max	Q.E moyenne	Q.E spectre
Nanoecaryotes				
45	0,01	1,42	0,16	0,04
71	0,02	1,67	0,22	0,26
				0,26
81	0,01	1,34	0,26	0,32
82	0,01	1,23	0,18	0,08
91	0,02	1,92	0,38	0,44
				0,31
				0,29
Prochlorococcus				
30	0,02	1,94	0,34	0,19
31	0,02	1,06	0,17	0,09
32	0,01	1,30	0,15	0,07
34	0,01	1,25	0,16	0,05
36	0,01	1,39	0,17	0,05
38	0,01	1,25	0,20	0,03
41	0,01	1,14	0,16	0,17
43	0,01	1,35	0,15	0,06
49	0,01	1,24	0,20	0,06
52	0,01	1,28	0,15	0,05
				0,04
53	0,01	1,00	0,15	0,07
54	0,01	1,19	0,16	0,03
56	0,01	1,01	0,18	0,06
				0,02
61	0,02	1,22	0,17	0,08
				0,03
				0,16
				0,16
62	0,01	1,15	0,15	0,06
63	0,01	1,23	0,15	0,06
71	0,02	1,67	0,22	0,17
91	0,02	1,92	0,38	0,65
SLC				
26	0,01	1,42	0,17	0,05
39	0,02	1,57	0,22	0,07
				0,06
41	0,01	1,14	0,16	0,07
49	0,01	1,24	0,20	0,05
64	0,01	1,83	0,16	0,03
65	0,01	1,20	0,17	0,03
Diatomées				
10	0,05	4,38	0,85	1,07
30	0,02	1,94	0,34	0,29
				0,10

Il est toutefois préférable de définir un seuil sur la distance euclidienne maximale prise en compte pour affecter un spectre à un neurone donné, et ce, dans le but d'éviter que des spectres soient affectés à un spectre référent différent (éloigné en termes de distance euclidienne) au risque d'être mal classés en termes de groupes de phytoplancton. Ce seuil peut être introduit, par exemple, sur l'écart-type de la distribution des spectres d'anomalies affectés à chaque neurone à la fin de la phase d'apprentissage. Ce seuil sera appliqué lors de la projection des cartes satellitaires journalières d'anomalies spectrales (chapitre 4). Il est important de souligner que le critère en question n'introduit pas une limitation dans la classification neuronale mais, plutôt, une assurance quant à la bonne qualité de la projection des spectres de Ra sur la carte entraînée, qualité qui sera déterminante pour l'application de cette classification à la détection des groupes de phytoplancton.

Cette comparaison montre l'un des avantages méthodologiques de la classification automatique des anomalies spectrales avec SOM par rapport à l'approche de PHYSAT version 2008, sans omettre la fiabilité statistique de la nouvelle méthode ainsi que sa flexibilité, donnant la possibilité d'exploiter efficacement de larges bases de données.

CHAPITRE 4

Application de SOM à la télédétection des groupes phytoplanctoniques à l'échelle globale : la version PHYSAT-SOM

On a précédemment montré que la méthode de classification automatique pouvait caractériser l'entière variabilité spectrale des Ra à partir du même jeu de données de spectres d'anomalies utilisé pour le développement de PHYSAT-v2008. Il est donc intéressant d'exploiter SOM pour la caractérisation des groupes de phytoplancton à partir d'une plus grande base de données de pigments et étudier son apport en termes de précision, de couverture spatiale. Ce chapitre a pour but la présentation des résultats de l'application de SOM à la télédétection des groupes phytoplanctoniques à l'échelle globale à partir des données du capteur SeaWiFS. Cette méthode a été appelée : PHYSAT-SOM.

Après la validation des cartes journalières produites par PHYSAT-SOM, certaines améliorations apportées par l'utilisation de la classification automatique, permettant d'avoir une prise en compte de l'entière variabilité des anomalies spectrales d'une part, et donnant la possibilité d'exploiter efficacement une large base de données *in situ* de pigments, d'autre part, seront analysées.

4.1. Description de la base d'inventaires pigmentaires utilisée

Afin de lier les anomalies spectrales, précédemment classées avec les cartes auto-organisatrices, à la présence de groupes de phytoplancton dominants à l'échelle globale, une expertise de terrain doit être utilisée. Une base de données *in situ*, comprenant 8410 échantillons d'eau collectés entre 1999 et 2009 dans le cadre de plusieurs campagnes océanographiques, a été exploitée dans cette partie de l'étude. Le nom des différentes campagnes ainsi que les zones couvertes sont résumés dans le tableau 4.1.

Selon les critères de qualité des données de PHYSAT détaillées précédemment (concentration en chlorophylle *a* entre 0,04 et 3 mg.m⁻³ et épaisseur optique des aérosols à 865 nm < 0,15), les mesures *in situ* coïncidentes à celles du capteur SeaWiFS sont au nombre de 1068 : c'est la base de données expertisées. Ces mesures ont été reliées aux produits satellites journaliers à 9 km de résolution (procédure de « match-up ») en les faisant correspondre au jour de la mesure et sur une fenêtre carrée de 3x3 pixels autour du pixel concerné. Bien que moins précise que la fenêtre de 3 heures standardisée pour la validation des données spatiales et des algorithmes bio-optiques (Bailey et Werdell, 2006, Jamet et al., 2011), cette fenêtre a été utilisée pour maximiser le nombre de mesures coïncidentes, de la même manière que dans Alvain et al. (2008).

Tableau 4.1. Origine des données *in situ* exploitées, zones couvertes et études associées

Campagne	Zone couvertes	Travaux associés
ICOTA (Ichtyologie Côtière en Terre Adélie)	Océan Austral-Antarctique	A. Goffart et D. Vincent, communication personnelle; Zorn et al., 2008
OOMPH (Organics over the Ocean Modifying Particles in the both Hemispheres)	Océan Austral	E. Peeken, communication personnelle
HALIFAX	Côte Atlantique Canadienne	E. Devred, communication personnelle
NOMAD (NASA bio-Optical Marine Algorithm Data set)	Global	Werdell et Bailey, 2005
OISO (Service d'Observation de l'Océan Indien)	Océan Indien	Metzl et al., 2006
GeP&CO (Geochemistry, Phytoplankton and Color of the Ocean)	Atlantique Nord-Mer des Caraïbes-Pacifique Equatorial et Sud	Dandonneau et al., 2004

La distribution géographique de ces mesures est représentée sur la figure 4.1. Bien que ces mesures coïncidentes ne représentent que 12,7 % de la base de données totale disponible, elles couvrent une large variabilité saisonnière ainsi que plusieurs régions de l'océan global, en particulier au niveau de l'Atlantique Nord et le Pacifique Equatorial grâce aux données des campagnes GeP&CO. Ceci permet la prise en compte de la diversité du phytoplancton sur de larges échelles spatio-temporelles, associés à des conditions environnementale diverses. En revanche, de larges zones demeurent faiblement représentées, telles que la mer Méditerranée ainsi que le Pacific Nord, l'Océan Austral ou encore l'Océan Indien. La base utilisée dans le cadre de cette étude est cependant plus complète que celle utilisée par la version 2008 de

PHYSAT. Cette dernière ne comprenait par exemple pas certaines mesures issues de l’Océan Austral, telles que les campagnes ICOTA et OOMPH.

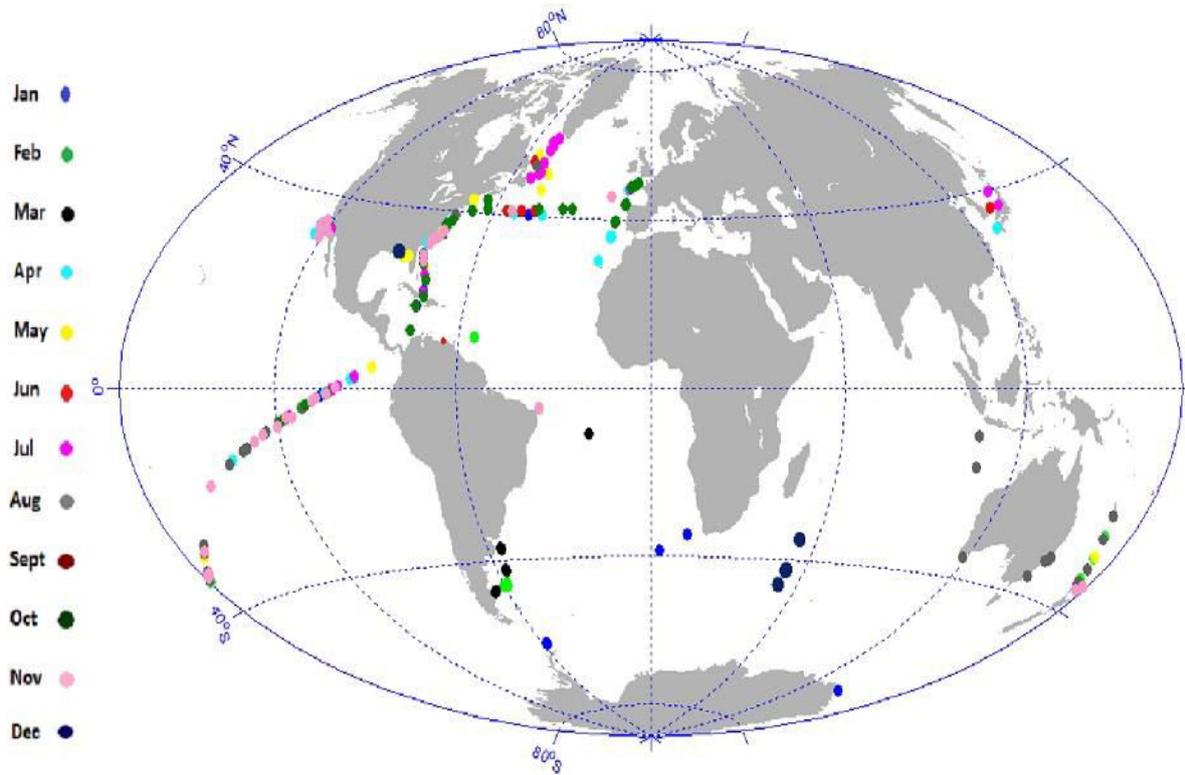


Figure 4.1. Distribution géographique des 1068 mesures in situ coïncidentes à une mesure du capteur SeaWiFS. Chaque mois est représenté par une couleur différente pour montrer la variabilité saisonnière couverte par les données in situ

4.2. Identification des groupes de phytoplancton et procédure de labellisation des anomalies spectrales

4.2.1. Identification des groupes de phytoplancton à partir des données *in situ*

Afin d’identifier les groupes de phytoplancton dans les inventaires pigmentaires, l’approche basée sur les seuils de la concentration relative en pigments biomarqueurs (Alvain et al., 2005 et 2008), décrite dans le chapitre 2, a été appliquée aux données précédemment présentées. Selon les seuils sur les pigments utilisés, 759 inventaires parmi les 1068 mesures *in situ* (soit 71 % du total des données coïncidentes à une mesure satellite) ont pu être associés à 4 groupes de phytoplancton dominants (tableau 4.2). Le reste des échantillons (309) correspondent soit à des compositions mixtes, soit à des inventaires non identifiés avec les critères utilisés. Dans le cadre de cette thèse, on a uniquement considéré les cas de dominance afin d’être le plus proche possible des hypothèses de développement de la méthode PHYSAT

originale de façon à simplifier la comparaison des résultats. Parmi les 759 inventaires identifiés, 20 % des échantillons ont été sélectionnés d'une manière aléatoire et séparément pour chacun des 4 groupes de phytoplancton, ces derniers seront utilisés pour l'étape de validation des cartes journalières de distribution des groupes phytoplanctoniques produites par PHYSAT-SOM. Les 609 échantillons restant (correspondant à 609 spectres de Ra coïncidents) ont été utilisés pour affecter un groupe de phytoplancton aux différents neurones.

Tableau 4.2. Nombre d'échantillons associés à chaque groupe de phytoplancton identifié à partir des inventaires de pigments

Groupe	Nombre d'échantillons/Total
Nanoecaryotes	216 / 759 (28,45 %)
Prochlorococcus	213 / 759 (28,06 %)
SLC	278 / 759 (36,62 %)
Diatomées	52 / 759 (6,87 %)

4.2.2. Procédure de labellisation des spectres de Ra classés avec SOM

La procédure de labellisation (ou étiquetage) des anomalies spectrales est basée sur la projection des spectres de Ra, associés empiriquement à des groupes dominants, sur la carte auto-organisatrice entraînée. Cette projection est réalisée par l'intermédiaire de la fonction d'affectation de SOM (voir chapitre 3). Puisque chaque spectre est associé à un groupe de phytoplancton particulier, cette procédure va nous permettre d'associer chaque groupe à un ou plusieurs neurones (neurones gagnants ou BMU), dont les vecteurs référents sont les plus proches de ce spectre au sens de la distance euclidienne, tel que ça a été expliqué dans le chapitre 3. Une fois cette projection effectuée, les spectres de Ra se trouvent distribués sur les 100 neurones de la carte auto-organisatrice. Chaque neurone capte une quantité d'information spécifique et peut donc être associé à :

- Un ou plusieurs spectres ayant le même label (cas 1),
- Plusieurs spectres associés à différents labels (cas 2),
- Aucun spectre (cas 3).

Pour chaque neurone de SOM, on calcule ensuite la fréquence relative des spectres de Ra affectés lors de la projection de la base de données *in situ*. Cette fréquence donne une information sur la proportion de données *in situ* associée à chaque unité de la carte auto-

organisatrice. Elle est calculée séparément pour chaque groupe de phytoplancton en divisant, pour chaque neurone, le nombre de spectres relatifs à un groupe donné et affectés à ce neurone par le nombre total de spectres correspondant à ce groupe de phytoplancton :

$$F.R(Neurone) = \frac{N_{groupe}^{Neurone}}{N_{groupe}^{Total}} \quad (\text{Eq. 4.1})$$

Ce choix de calcul a été fait dans le but de tenir compte du label plutôt que du nombre de spectres qui diffère d'un groupe à un autre, pouvant introduire un biais dans la labellisation des neurones (Les SLC sont par exemples associés à 37 % des mesures *in situ* contre seulement 7 % pour les diatomées en raison de la dominance relative des cyanobactéries sur de larges échelles spatiales, contrairement aux diatomées qui représentent un groupe aux floraisons intenses mais plutôt locales).

Dans le premier cas (un ou plusieurs spectres ayant le même label), le neurone prend le label du spectre affecté. Pour le deuxième cas (plusieurs spectres associés à différents labels), une étape supplémentaire est nécessaire pour labelliser ce type de neurones : si, au moins 50 % de la fréquence relative des spectres de Ra captés ont le même label, le neurone prend ce label ; sinon il prend l'étiquette « pas de dominance spécifique ». Pour le cas 3 (aucun spectre), les vecteurs référents de ces neurones ne sont associés à aucun spectre de Ra. Ces vecteurs référents correspondent à des spectres de Ra qui ne sont pas représentés dans ce jeu de données *in situ*. Les neurones correspondant ne sont, par conséquent, pas labellisés mais pourraient l'être si des données *in situ* supplémentaires sont disponibles à l'avenir. Par ailleurs, seuls les neurones ayant une fréquence relative de spectres affectés supérieure à 0,025 (ce qui représente 3 spectres en moyenne, tous groupes confondus) ont été pris en compte dans le but de garder uniquement les vecteurs référents associés à une part significative de données *in situ*.

Les fréquences relatives des spectres labellisés associés aux 100 neurones de SOM sont représentées sur la figure 4.2. Notons que tous les spectres coïncidents pris en compte pour la labellisation de la carte auto-organisatrice (609) ont été capturés par au moins un neurone. Tous les spectres de la base de données *in situ* projetés sur la carte auto-organisatrice ont une erreur de quantification vectorielle inférieure à l'erreur maximale associée au neurone gagnant (Q.E maximale calculée d'après les données d'apprentissage). De ce fait, aucun seuil

sur l'écart-type n'a été appliqué pour cette phase de labellisation. Toute l'information *in situ* disponible pour l'étiquetage des neurones a donc été conservée.

Tel que le montre la figure 4.2, 14 neurones n'ont été affectés à aucun spectre (cas 3). Les spectres associés à un seul groupe phytoplanctonique (1^{er} cas) ont été affectés à 12 neurones. Enfin, les neurones dont le vecteur référent a capté des spectres associés à différents groupes de phytoplancton (2^{ème} cas) sont les plus fréquents (74 %).

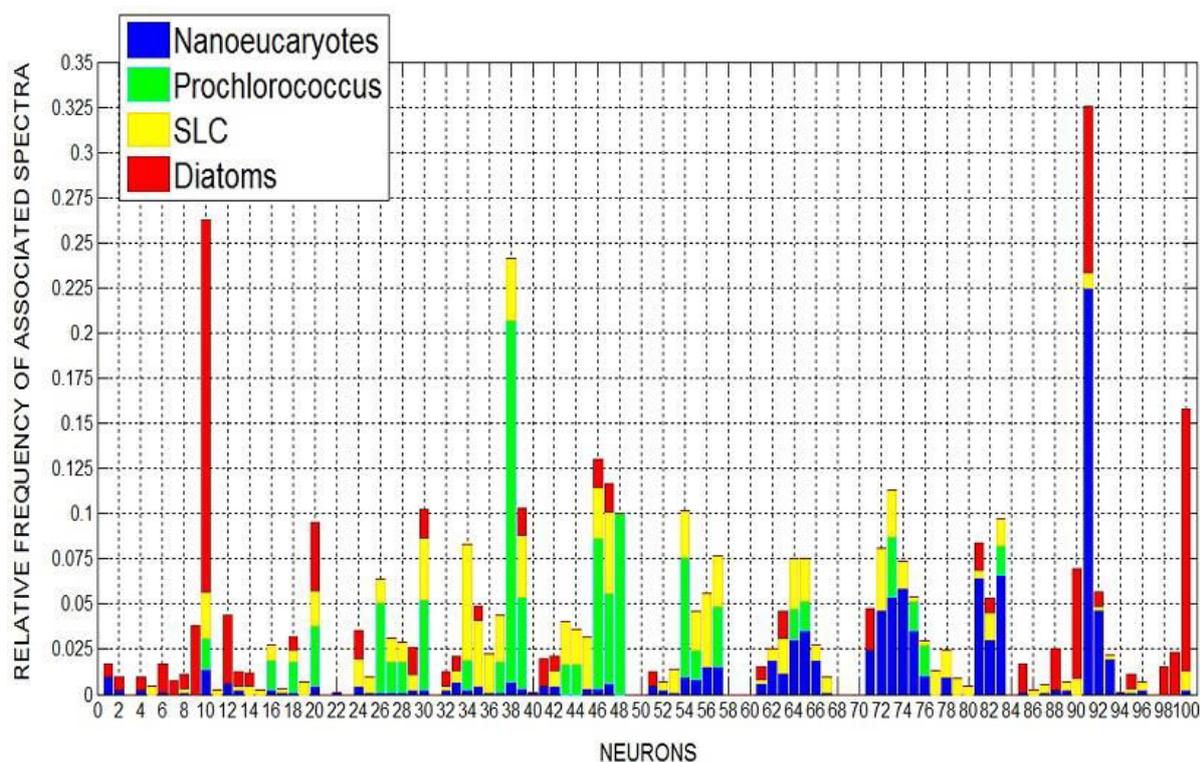


Figure 4.2. Fréquence relative des labels associés aux spectres de *Ra* répartis sur les 100 neurones de la carte auto-organisatrice

Une fois que tous les critères de labellisation décrits précédemment sont appliqués, 44 neurones sont affectés à l'un des 4 groupes de phytoplancton ou à l'étiquette « pas de dominance spécifique », les 56 neurones restants ayant une fréquence relative de spectres associés inférieure à 0,025 ou appartenant au troisième cas. Ces neurones n'ont pas été affectés à un groupe phytoplanctonique dans le cadre de cette étude, cependant, ils représentent une information spectrale potentiellement exploitable dès lors que des données *in situ* supplémentaires seraient disponibles ou lors d'études non focalisés sur les cas de dominances seuls. En outre, tel que précisé dans le chapitre 3, le choix d'un grand nombre de neurones vise en premier lieu à avoir une discrétisation très fine des spectres d'anomalies. Leur labellisation dépendra des données *in situ* utilisées. D'après la figure 4.2, cinq neurones

sont associés aux diatomées et neuf aux nanoeucaryotes. Les SLC se retrouvent plus souvent affectés à des neurones sans dominance spécifique, mélangés avec les *Prochlorococcus*. Ceci est à mettre en relation avec le fait que ces deux groupes soient associés à des environnements bio-optiques proches (Alvain et al., 2012) et dominent souvent les mêmes zones géographiques. Les assemblages dominants de SLC ont été affectés à six neurones.

La figure 4.3 représente la carte bidimensionnelle des spectres référents avec les fréquences relatives des groupes de phytoplancton pour les neurones portant un label de dominance (les compositions mixtes ne sont pas représentées par un camembert). L'ordre topologique sur la forme et l'amplitude des spectres de Ra induit par l'algorithme SOM est en concordance avec la position des différents labels sur la carte bidimensionnelle. En effet, les groupes de phytoplancton occupent une zone particulière de la carte et se retrouvent affectés à des neurones adjacents (qui sont donc similaires en termes de forme et d'amplitude), à l'exception des diatomées qui se retrouvent affectés à des zones différentes de la carte, avec, par conséquent, des signatures spectrales différentes pour un même groupe phytoplanctonique, ce qui représente une nouveauté par rapport aux résultats de la méthode PHYSAT dans sa version de 2008.

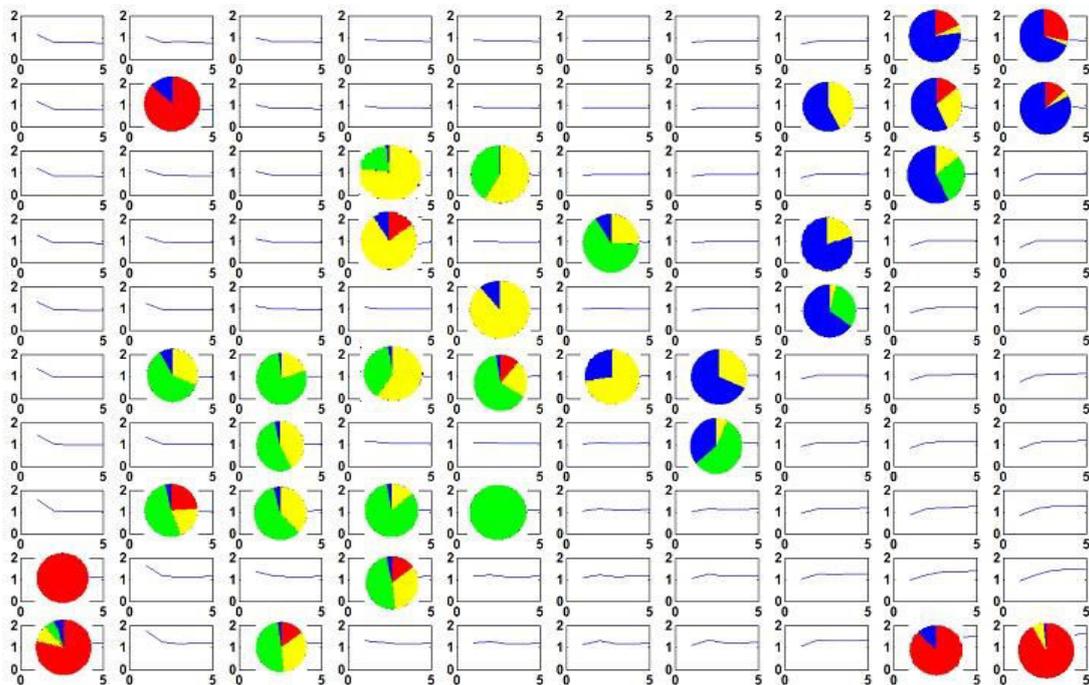


Figure 4.3. Représentation des fréquences relatives des groupes de phytoplancton dominants sur la carte bidimensionnelle des spectres référents. Sont représentés en bleu les nanoeucaryotes, en vert les *Prochlorococcus*, en jaune les SLC et en rouge les diatomées (les labels des neurones portant l'étiquette « de dominance spécifique » ne sont pas représentés sur cette figure). On note la concordance entre l'ordre topologique des spectres et les labels affectés.

D'après l'étude théorique d'Alvain et al. (2012), il est possible que des mêmes groupes de phytoplancton soient associés à des spectres de Ra différents. La forme et l'amplitude de ces spectres étant sous l'influence de plusieurs paramètres optiques (voir chapitre 2), ces différentes signatures ne correspondraient nécessairement pas à différentes espèces, mais pourraient être associées à des environnements optiquement différents. Les différentes signatures spectrales associées aux diatomées seront analysées plus en détails dans le prochain chapitre.

En plus des quatre groupes de phytoplancton identifiés à partir des inventaires pigmentaires, les cartes précédemment produites par la méthode PHYSAT (Alvain et al., 2008) ont permis de détecter un groupe supplémentaire, appelé *Phaeocystis-like*, en raison de sa labellisation basée sur des études bibliographique et non pas à partir de mesures *in situ*. Les spectres de Ra correspondant aux pixels appartenant à ce groupe dans la version de PHYSAT-2008 (principalement le mois de Janvier au niveau de l'Océan Austral) ont été projetés sur les vecteurs référents de la carte auto-organisatrice entraînée. Ces spectres ont été « captés » par deux neurones adjacents, labellisés à leur tour « *Phaeocystis-like* ».

4.3. Analyse des formes et amplitudes des spectres associés aux groupes phytoplanctoniques

Les signatures spectrales de Ra correspondant aux quatre groupes de phytoplancton dominants identifiés à partir des inventaires pigmentaires sont affichées dans la figure 4.4. Cette figure représente les spectres référents des neurones affectés à un groupe phytoplanctonique dominant selon les critères définis précédemment dans la procédure de labellisation. Les valeurs des spectres affichées dans cette figure sont des valeurs moyennes synthétisant les ordres de grandeur de Ra pouvant être trouvés pour chaque groupe de phytoplancton.

Les nanoeucaryotes (figure 4.4 a) sont caractérisés par des amplitudes d'anomalies spectrales relativement faibles par rapport aux autres groupes avec des valeurs moyennes inférieures à 1 pour toutes les longueurs d'onde et une légère augmentation entre 412 et 443 nm. Les *Prochlorococcus* et SLC (respectivement représentés par les volets (b) et (c) de la figure) sont associés à des spectres plus neutres avec des valeurs maximales et minimales de Ra autour de 1.

Les diatomées sont caractérisées par des valeurs d'anomalies spectrales relativement élevées par rapport aux trois autres groupes phytoplanctoniques. Cependant, deux signatures spectrales distinctes caractérisent les diatomées. En effet, la figure 4.4 montre que ce groupe a été associé à deux types de vecteurs référents différents en termes de forme et d'amplitude. Le premier type de spectres (figure 4.4 d) est caractérisé par des valeurs maximales de Ra autour de 412 nm avec une décroissance en allant des courtes vers les grandes longueurs d'onde. Le second type de spectres (figure 4.4 e) est, en revanche, caractérisé par des valeurs minimales de Ra autour de 412 nm et des valeurs plus élevées entre 443 et 555 nm. Ce second type de spectre ne correspond pas nécessairement à une espèce différente de diatomées, mais pourrait être aussi lié à la même espèce localisée dans un environnement bio-optique différent avec des valeurs distinctes de propriétés optiques inhérentes (Alvain et al., 2012). Cette différence pourrait par exemple être associée à différents stades de croissance ou à la localisation de l'espèce dans des environnements caractérisés par la présence de substances optiquement actives, telles que les matières organiques dissoutes (CDOM). Ce volet sera abordé en détails dans le prochain chapitre, dédié aux diatomées.

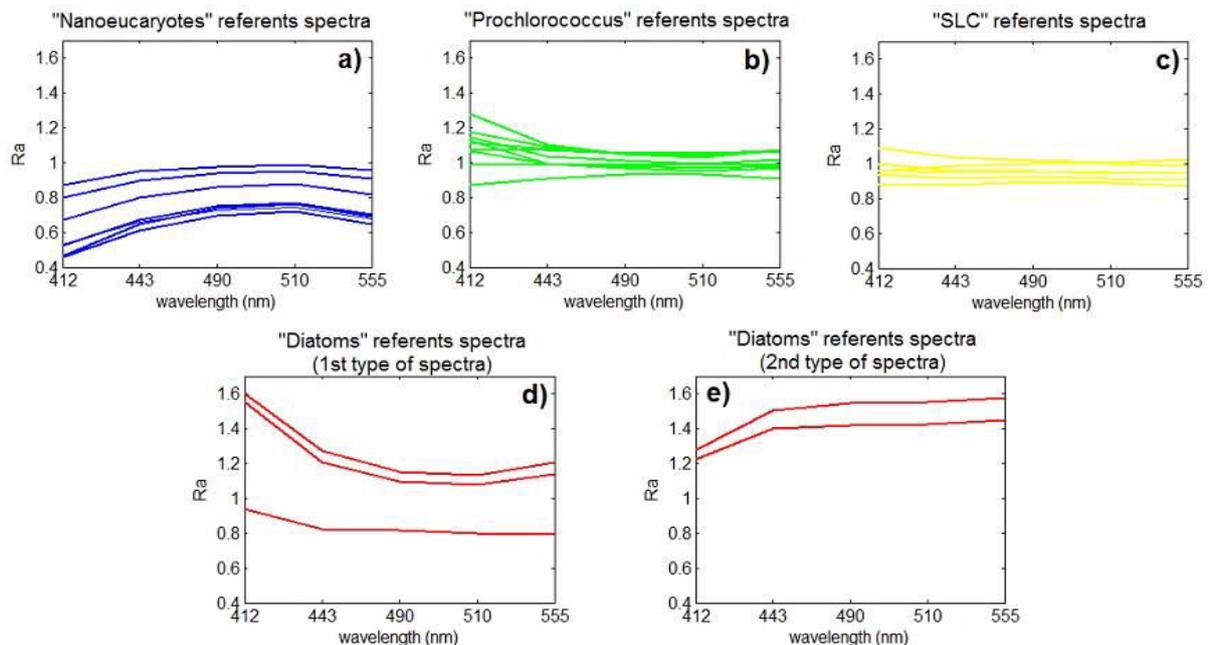


Figure 4.4. Signatures spectrales des groupes de phytoplancton identifiés à partir des inventaires pigmentaires, via la projection sur la carte SOM. Chaque groupe est représenté par les spectres référents des neurones auxquels il a été affecté.

Si les signatures spectrales moyennes (spectres référents) des 4 groupes de phytoplancton sont relativement proches de celles obtenues par la version 2008 de PHYSAT

(figure 2.2 du chapitre 2), les gammes spectrales correspondant à chaque groupe, c'est-à-dire les minima et les maxima de R_a , ne sont pas les mêmes entre les deux méthodes. La figure 4.5 affiche les spectres de R_a minimaux et maximaux de PHYSAT-SOM (spectres associées aux neurones labellisés à la fin de la phase d'apprentissage) et PHYSAT version 2008 correspondant à chaque groupe de phytoplancton identifié à partir des inventaires de pigments. A première vue, on observe que les intervalles de R_a de chaque groupe définis par PHYSAT-SOM sont plus larges que ceux de l'ancienne méthode. A titre d'exemple, les nanoeucaryotes sont caractérisés par des valeurs minimales de R_a plus faibles chez PHYSAT-SOM et le spectre minimal défini par PHYSAT version 2008 pour ce groupe correspond au spectre moyen trouvé avec PHYSAT-SOM (entre 0,4 et 0,5). Les *Prochlorococcus* et SLC se trouvent dans des gammes spectrales plus ou moins proches pour les deux méthodes (entre 0,9 et 1,2) avec tout de même des valeurs de R_a maximales légèrement plus élevées avec PHYSAT-SOM (autour de 1,2 pour PHYSAT-SOM et 1 pour PHYSAT) et vice-versa pour les SLC. Les diatomées présentent les plus grandes différences avec des valeurs de R_a maximales significativement plus élevées avec PHYSAT-SOM (légèrement supérieures à 2 avec PHYSAT à 412 nm alors qu'elles atteignent 4 avec PHYSAT-SOM sur la même longueur d'onde). Notons aussi que ce groupe est caractérisé par deux signatures spectrales différentes en termes de forme avec PHYSAT-SOM, ce qui représente une nouveauté par rapport à l'ancienne méthode.

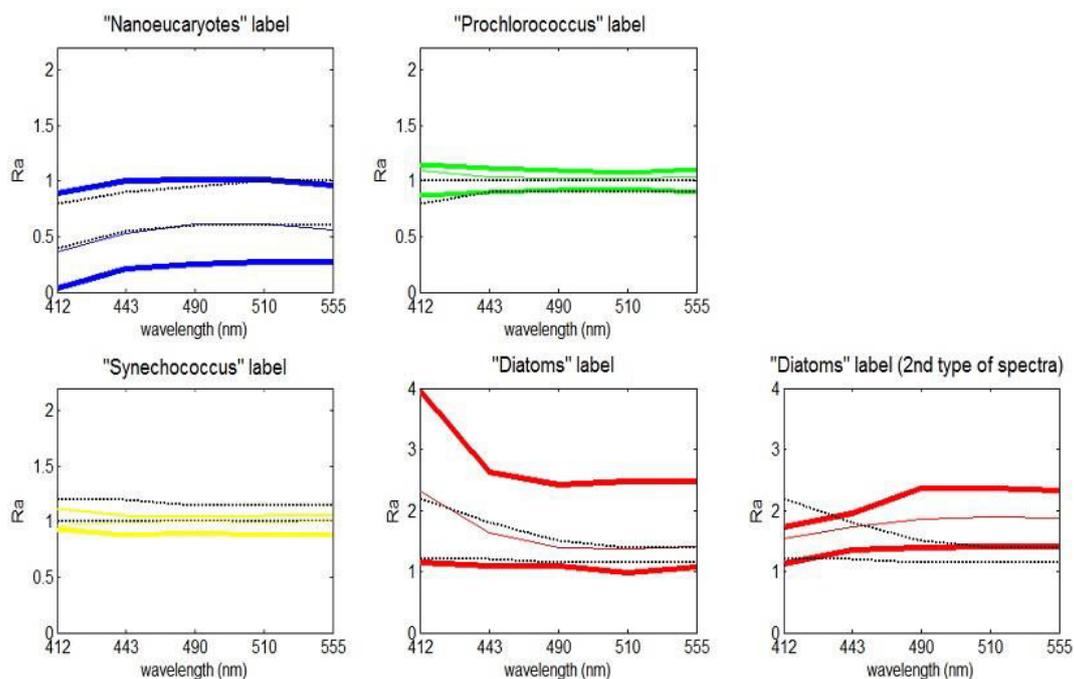


Figure 4.5. Gammes spectrales de R_a pour les 4 groupes de phytoplancton identifiés à partir des inventaires pigmentaires: les lignes fines désignent le spectre moyen et les lignes en gras l'écart type. Les lignes noires entrecoupées désignent les tubes de spectres de R_a définies par la version 2008 de PHYSAT.

Ces différences sont toutefois à relativiser car elles sont associées, d'une part, à un changement de méthode (utilisation de SOM pour la classification automatique des Ra) et d'autre part, à l'exploitation de nouvelles informations *in situ*. L'apport de la méthode seule consiste au fait qu'une plus grande variabilité spectrale est prise en compte au sein du même label par rapport à la version 2008 de PHYSAT, ceci a été montré dans le chapitre précédent, en se basant sur le même jeu de 41 spectres de Ra (paragraphe 3.4.3 du chapitre 3). En effet, la méthode présentée dans ce manuscrit a été développée à partir d'une très large base de données satellite comprenant une vaste variabilité de spectres de Ra en termes de formes et d'amplitudes, ce qui représente en soi un avantage considérable des réseaux de neurones puisque c'est l'entière variabilité des Ra qui est prise en compte et non celle associée uniquement à une information *in situ*.

L'apport de nouvelles données de terrain va, quant à lui, avoir un impact sur le nombre total de neurones labellisés ainsi que sur le label associé à ces neurones, tel que le montre la figure 4.6 qui représente les fréquences relatives de spectres associés aux neurones de SOM labellisés uniquement à partir de 41 spectres de Ra. En effet, bien que les labels affectés aux neurones à partir des 609 spectres (figure 4.2) et ceux affectés aux neurones à partir des 41 spectres uniquement (figure 4.6) restent consistants (distribution relativement semblable des labels sur les neurones), il apparaît qu'en utilisant uniquement cette base de données réduite, seuls 26 neurones sont affectés à un ou plusieurs labels, contre 86 en utilisant l'ensemble des 609 mesures coïncidentes. A titre d'exemple, seuls deux neurones (9 et 10) sont affectés au label « diatomées » en utilisant uniquement la base des 41 spectres de Ra. Dans ce cas, les diatomées n'auraient pu être associés à la nouvelle signature spectrale (figure 4.4 e) car cette dernière correspond aux spectres référents des neurones 90 et 100, qui, cette fois, n'ont classé aucun spectre de Ra. Ceci met en évidence le caractère indispensable de la poursuite des mesures *in situ* dès que cela est possible pour augmenter les bases de données disponibles.

Les différentes signatures spectrales, précédemment analysées, vont à présent être exploitées pour étudier la distribution spatio-temporelle des groupes de phytoplancton à l'échelle globale à partir des cartes journalières de Ra (calculées à partir de l'archive d'images disponibles du capteur SeaWiFS de 1997 à 2010). Une fois la labellisation de la carte auto-organisatrice accomplie, chaque pixel des cartes satellite journalières d'anomalies spectrales peut être affecté à son neurone le plus proche (vecteur référent le plus proche) de SOM. Chaque pixel classé prend par conséquent le label de son neurone le plus proche pour

produire des cartes journalières de distribution des groupes de phytoplancton à l'échelle globale.

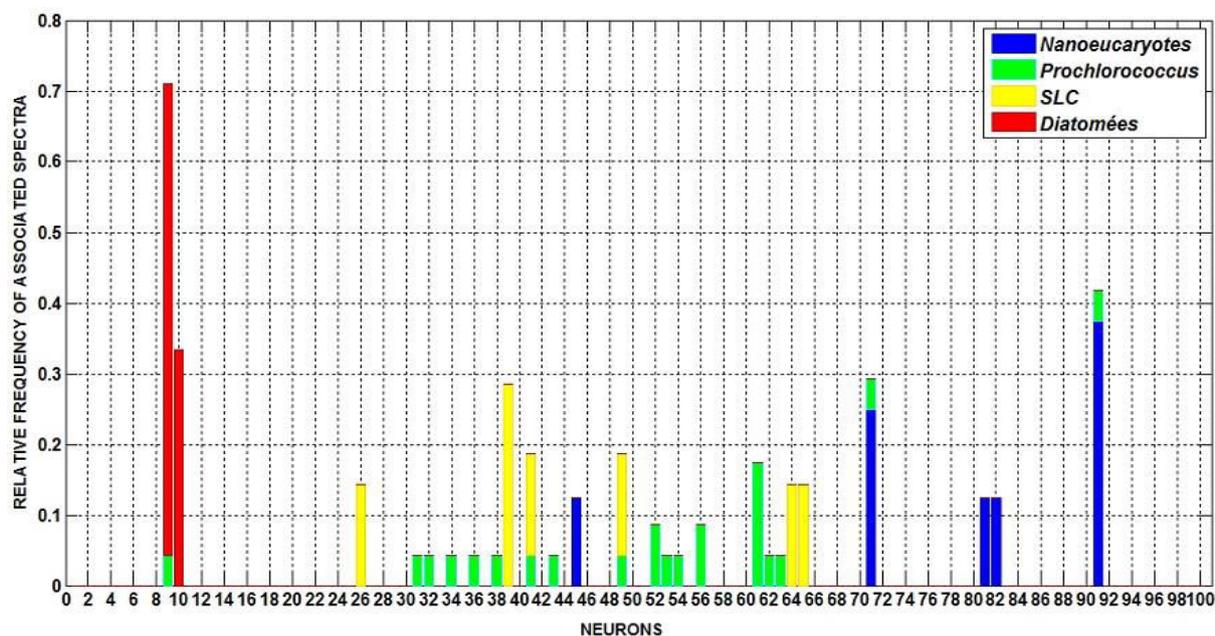


Figure 4.6. Fréquence relative des labels associés aux spectres de Ra répartis sur les 100 neurones (labellisation à partir des 41 spectres de Ra utilisés pour le développement de PHYSAT version 2008)

4.4. Validation des cartes journalières de distribution des groupes de phytoplancton obtenues avec PHYSAT-SOM

La validation des cartes journalières de distribution globale des quatre groupes de phytoplancton identifiés à partir des inventaires pigmentaires a été faite en utilisant les 20 % des échantillons qui ont été sélectionnés aléatoirement avant l'étape de labellisation, séparément pour chacun des 4 groupes de phytoplancton. Ce jeu de données de validation comprend 42 échantillons dominés par les nanoeucaryotes, 44 par *Prochlorococcus*, 54 par SLC et 10 par les diatomées (figure 4.7). Les mesures *in situ* relatives à chaque groupe ont été reliées aux pixels des cartes satellite journalières produites par PHYSAT-SOM (procédure de match-up) en les faisant correspondre à travers une fenêtre carrée de 3x3 pixels autour du pixel concerné. Précisons que chaque mesure coïncidente n'a été prise en compte pour la validation que si, au moins, 60 % des pixels valides de la fenêtre prise en compte portent le même label (c'est à dire, au moins 2/3, 3/4, 3/5, 4/6, 5/7, 5/8, ou 6/9 pixels portant le même label). Ce critère d'homogénéité des labels permet d'assurer une comparaison stricte entre les mesures *in situ* et les mesures satellite et, surtout, d'éviter les cas de bordures des structures (ou la présence de nuages), donnant une fiabilité aux résultats de l'exercice de validation. En

définitive, 40 échantillons dominés par les nanoeucaryotes, 43 par *Prochlorococcus*, 52 par SLC et 10 par les diatomées ont été gardés.

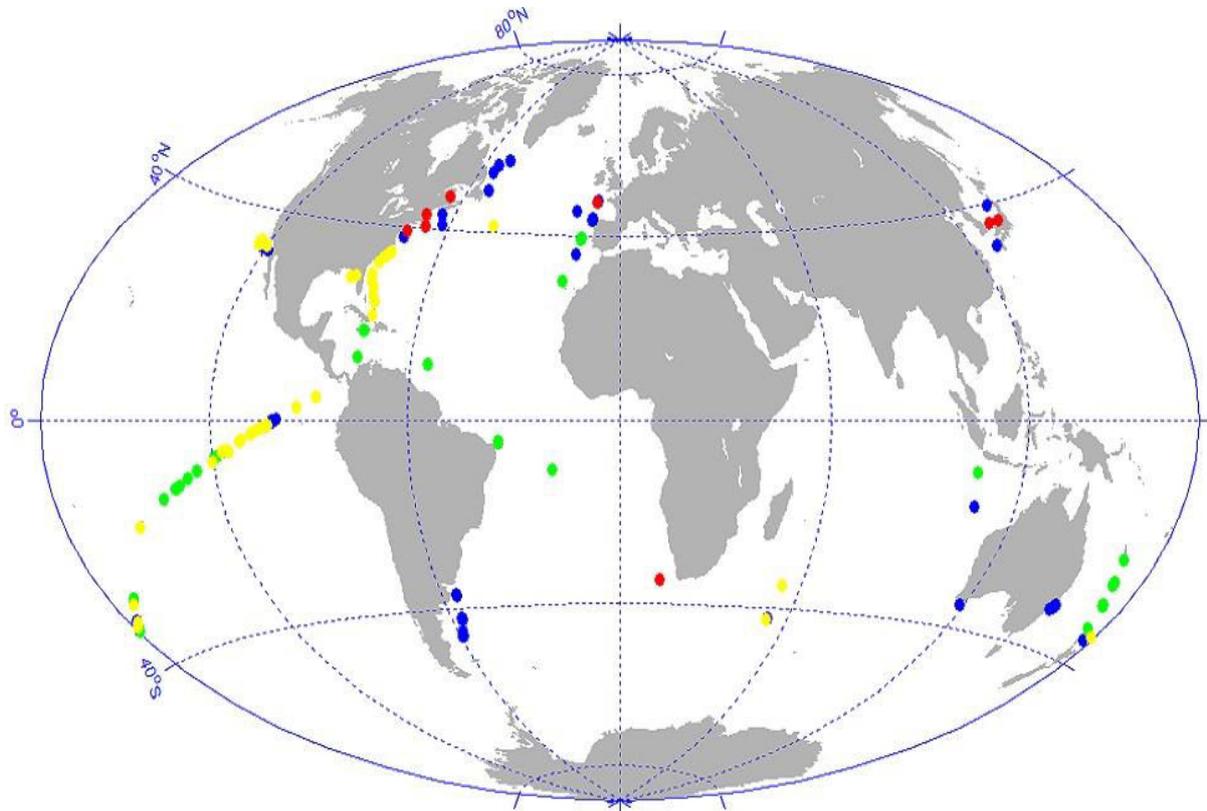


Figure 4.7. Distribution géographique des échantillons gardés pour la validation des résultats (en bleu les nanoeucaryotes, en vert *Prochlorococcus*, en Jaune SLC et en rouge les diatomées)

Le tableau 4.3, appelé matrice de confusion, montre le pourcentage de bonnes classifications pour chaque groupe de phytoplancton. On note que cette validation ne prend pas en compte les pixels caractérisés par une dominance de *Phaeocystis-like* puisqu'ils ont été identifiés en se basant sur la littérature, tel que fait par Alvain et al., (2008). Le tableau 4.3 montre que 66,7 % des inventaires pigmentaires dominés par les nanoeucaryotes sont associés à ce même groupe dans les cartes journalières produites par PHYSAT-SOM. Les identifications erronées de ce groupe sont principalement associées à des échantillons prélevés dans l'hémisphère Nord et le Pacifique Equatorial. Dans ces cas-là, les nanoeucaryotes sont associés de façon erronée à des spectres *Prochlorococcus* (19,05 %) et SLC (9,52 %).

Prochlorococcus et SLC sont correctement identifiés dans, respectivement, 58,07 % et 66,67 % des mesures *in situ* dominées par ces groupes. Il est intéressant de noter que la majeure partie des identifications erronées de *Prochlorococcus* soient associées à des échantillons dominés par SLC (dans 22,58% des mesures associées à *Prochlorococcus*). Ce

résultat n'est pas surprenant compte tenu des caractéristiques proches de ces cyanobactéries en termes de distribution géographique, taille et forme des cellules (Partensky et al., 1999 ; Zubkhov et al., 2000 ; Dandonneau et al., 2004 ; Longhurst et al., 2007 ; Flombaum et al., 2013) ainsi que l'amplitude et la forme de leurs signatures spectrales respectives (voir figure 4.5 et Alvain et al., 2012).

Une partie significative des inventaires pigmentaires dominés par les diatomées est correctement identifiée dans les cartes journalières de PHYSAT-SOM à 9 km de résolution (83 %). Les indentifications erronées des diatomées sont associées aux nanoeucaryotes. Ces échantillons sont localisés dans des eaux alternativement dominés par les diatomées et les nanoeucaryotes, essentiellement dans l'Atlantique Nord.

Tableau 4.3. Résultats de l'exercice de validation de la méthode PHYSAT-SOM. Pour chaque groupe de phytoplancton, le pourcentage des identifications valides (en rouge) et erronées (en noir) est indiqué. Les identifications erronées sont séparées par groupe (le nombre d'échantillons pris en compte pour chaque groupe est indiqué entre parenthèses)

PHYSAT-SOM / IN SITU	Nanoeucaryotes (40)	<i>Prochlorococcus</i> (43)	SLC (52)	Diatomées (10)
Nanoeucaryotes	66,67	19,05	9,52	4,76
<i>Prochlorococcus</i>	19,35	58,07	22,58	0
SLC	20	13,33	66,67	0
Diatomées	16,67	0	0	83,33

Les résultats de la dernière validation de PHYSAT version 2008 (Alvain et al., 2008 et Alvain et al., 2012) sont affichés dans la tableau 4.4 pour comparaison. Notons que nous comparons ici les résultats de validation des deux méthodes uniquement à titre indicatif car la méthode développée dans ce manuscrit est basée sur un apprentissage à partir d'une plus large base de données de spectres (chapitre 3). Par ailleurs, la base de données *in situ* utilisée pour labelliser les neurones de la carte auto-organisatrice est également plus fournie que celle exploitée pour développer la méthode de 2008.

En comparant les tableaux 4.3 et 4.4, on constate que les nanoeucaryotes sont associés à plus d'identifications correctes avec la version 2008 de PHYSAT (82%) qu'avec la méthode développée durant cette thèse (PHYSAT-SOM) (66,67 %). Ceci dit, en comparant en détail

les résultats pour ce groupe, on observe que, pour les deux méthodes, les identifications erronées de nanoeucaryotes sont majoritairement associées à *Prochlorococcus* (19,05 % et 10 % respectivement pour PHYSAT-SOM et PHYSAT) et SLC (9,52 % et 5 % respectivement pour PHYSAT-SOM et PHYSAT). Ces mesures correspondent à des échantillons prélevés dans l'hémisphère Nord et le Pacifique Equatorial (Dans ce cas, les nanoeucaryotes sont identifiés SLC par les deux méthodes) (Alvain et al., 2008). Par ailleurs, les mesures correctement identifiées comme nanoeucaryotes par PHYSAT-SOM (par exemple autour des côtes australiennes de même que quelques points dans l'Atlantique Nord) sont associées au même groupe au niveau de ces zones avec la version 2008 de PHYSAT.

La plus grande partie des identifications erronées de *Prochlorococcus* et SLC est majoritairement associée à l'un ou l'autre groupe par les deux méthodes. De même, pour les diatomées, les fausses détections sont majoritairement associées aux nanoeucaryotes par les deux méthodes (dans respectivement 16,67 % et 24 % des cas avec PHYSAT-SOM et PHYSAT version 2008). Pour les deux approches, ces classifications erronées trouvent dans des zones alternativement dominées par les diatomées et les nanoeucaryotes (e.g. Atlantique Nord).

Tableau 4.4. Résultats de l'exercice de validation de la méthode PHYSAT version 2008 (en pourcentages)

PHYSAT / IN SITU	Nanoeucaryotes	<i>Prochlorococcus</i>	SLC	Diatomées
Nanoeucaryotes	82	10	5	3
<i>Prochlorococcus</i>	14	61	25	0
SLC	20	23	57	0
Diatomées	24	0	3	73

4.5. Observations globales et cycles saisonniers

Des cartes climatologiques mensuelles de la distribution globale des groupes de phytoplancton ont été générées à 9 km de résolution à partir des cartes journalières de niveau 3 entre Septembre 1997 et Décembre 2010 en sélectionnant, pour chaque mois de la période d'observation de SeaWiFS (13 ans), le groupe le plus fréquemment détecté dans chaque pixel, selon la méthode Alvain et al. (2008). Les cartes générées avec PHYSAT-SOM sont

représentées dans les figures 4.8, 4.9, 4.10 et 4.11 pour les 12 mois de l'année avec celles générées par PHYSAT (version de 2008). Chacune d'elles représente le groupe de phytoplancton dominant dans chaque pixel sur une synthèse de chaque mois de l'archive satellite SeaWiFS entière. Rappelons que les bases de données de Ra et *in situ* utilisées pour développer PHYSAT-SOM et PHYSAT sont différentes, comme expliqué précédemment.

Les résultats globaux montrent des structures caractérisées par la dominance des *Prochlorococcus* et *Synechococcus* dans les eaux oligotrophes tropicales. La taille de ces cyanobactéries picoplanctoniques (0.2-2 μm) les rend moins vulnérables à la limitation en nutriments de ces environnements que les cellules phytoplanctoniques plus larges (Longhurst, 2007). Le cycle saisonnier est nettement marqué dans les eaux des hautes latitudes riches en nutriments où les nanoeucaryotes dominent en hiver (Décembre-Janvier-Février dans l'hémisphère Nord et Juin-Juillet-Aout dans l'hémisphère Sud) et où de larges régions de floraisons de diatomées apparaissent en été (Décembre-Janvier-Février dans l'hémisphère Sud et Juin-Juillet-Aout dans l'hémisphère Nord). L'extension de ces floraisons de diatomées est particulièrement importante dans l'Océan Austral durant le mois de Janvier. La distribution géographique des groupes phytoplanctoniques est cohérente avec de précédentes études (Dandonneau et al., 2004; Zubkov et al., 2000 ; Marty et Chiavérini, 2002; DuRand et al., 2001; Steinberg et al., 2001; Longhurst, 2007), tel que ça a été souligné par Alvain et al. (2005). Dandonneau et al. 2004 montrent que les *Prochlorococcus* et les cyanobactéries sont les groupes dominants tout au long de l'année. Même s'ils dominent généralement les mêmes environnements oligotrophes, Partensky et al. (1999) montrent que *Prochlorococcus* est plus important en termes d'abondance (100 fois plus abondant selon l'étude précitée) dans les eaux océaniques oligotrophes (gyres subtropicaux) avec des concentrations d'environ 1 à 4 $\times 10^5$ cellules/ml. Le nombre de cellules chute d'une manière exponentielle à partir des latitudes $> 45^\circ$ (Partensky et al., 1999 ; Dandonneau et al., 2004). La distribution de ces cyanobactéries picoplanctoniques n'est toutefois pas restreinte aux eaux oligotrophes, elles se trouvent aussi dans les eaux mésotrophes. Par exemple, dans le Pacifique Equatorial, il n'y a pas un changement drastique dans la distribution de ces cyanobactéries entre les eaux chaudes oligotrophes et les zones où la concentration en chlorophylle est faible relativement à celle des nutriments (zones appelées HLNC pour High Nutrient Low Chlorophyll).

Des structures caractérisées par des floraisons de *Phaeocystis-like* sont également visibles, particulièrement sur la carte générée par PHYSAT-SOM durant le mois de Mars

(figure 4.8 c) au niveau de la Manche et de la côte atlantique française. Bien que des blooms de *Phaeocystis* aient été régulièrement observés dans les eaux de la Manche et de la Mer du Nord (Brunet et al., 1996; Breton et al., 2000; Houliez et al., 2012; Lancelot et Mathot, 1987; Rousseau et al., 2006), ce groupe a été identifié en se basant uniquement sur la littérature et, par conséquent, toutes les observations le concernant doivent être considérés avec prudence en attendant une évaluation avec des données *in situ* supplémentaires.

Les synthèses mensuelles montrent de larges structures qui sont globalement en accord avec la répartition globale des groupes de phytoplancton dans les cartes produites avec la version 2008 de PHYSAT (Alvain et al., 2008). La distribution des groupes est cependant significativement différente entre les deux méthodes dans certaines zones, en particulier au niveau des hautes latitudes de l'hémisphère sud en Décembre et Janvier et des upwellings côtiers, caractérisés par de fortes floraisons de diatomées. Par exemple, les blooms de diatomées sont plus clairement visibles dans les upwellings des côtes péruviennes et équatoriennes ainsi qu'au niveau de l'upwelling du Benguela, le long des côtes Sud-Africaines, de Namibie et de l'Angola (zones encerclées noir sur la figure 4.8 a). Les différences apportées par PHYSAT-SOM quant à la détection des blooms de diatomées dans les upwellings seront discutées dans le prochain chapitre.

Par ailleurs, on observe des zones caractérisées par la dominance des nanoeucaryotes sur les cartes générées par PHYSAT version 2008 au niveau de l'Atlantique Sud au large des côtes de Patagonie vers les côtes sud-africaines au mois de Janvier (figure 4.8 a de droite) alors que ces zones sont caractérisés par la dominance des cyanobactéries (*Prochlorococcus* ou SLC) ou des diatomées sur les cartes de PHYSAT-SOM (figure 4.8 a de gauche). L'exercice de validation a montré que les résultats des deux méthodes, bien que différents en termes de pourcentages de détection, conservent néanmoins les mêmes tendances d'identification (les mesures erronées sont, par exemple, associées aux mêmes groupes pour les deux méthodes). Une explication possible aux différences observées réside dans le fait que les inventaires pigmentaires des campagnes GeP&CO utilisés pour développer la version 2008 n'incluaient pas d'échantillons prélevés dans l'Atlantique Sud. Les eaux au large des côtes de Patagonie sont par exemple connues pour être caractérisées par des floraisons intenses de diatomées durant la période estivale, en particulier dans les zones où le fer n'est pas en quantité limitée telles que les eaux du plateau continental de Patagonie ainsi que près de la Géorgie du Sud et les Îles Sandwich du Sud (Borrione et Schlitzer., 2012; Korb et al.,

2004). Par ailleurs, la dominance de *Prochlorococcus* ou SLC conjointement aux diatomées dans les zones de convergences entre 40° et 50°S a déjà été constatée (Zubkov et al., 2000). Ces organismes sont connus pour être dominants dans les zones où les conditions ne sont pas favorables aux diatomées. Ceci peut arriver quand, par exemple, la concentration en silicates, qui représentent un nutriment essentiel pour la croissance des diatomées, sont en quantité insuffisante (Peeken, 1997 ; Selph et al., 2001). Les blooms de diatomées débutent vers le mois d'Octobre dans la zone de convergence autour de 40°S (Tremblay et al., 2002) et couvrent la majeure partie de l'Océan Austral en Décembre et Janvier (Smetacek et al., 1990). Par ailleurs, la contribution des nanoeucaryotes à la concentration totale en chlorophylle *a* baisse significativement (moins de 40 %) dans l'Océan Austral lorsque les diatomées et *Phaeocystis* dominant (Alvain et al., 2008). Ces groupes sont connus pour être largement dominants en hiver dans les eaux riches en nutriments de l'Océan Austral (Longhurst, 2007).

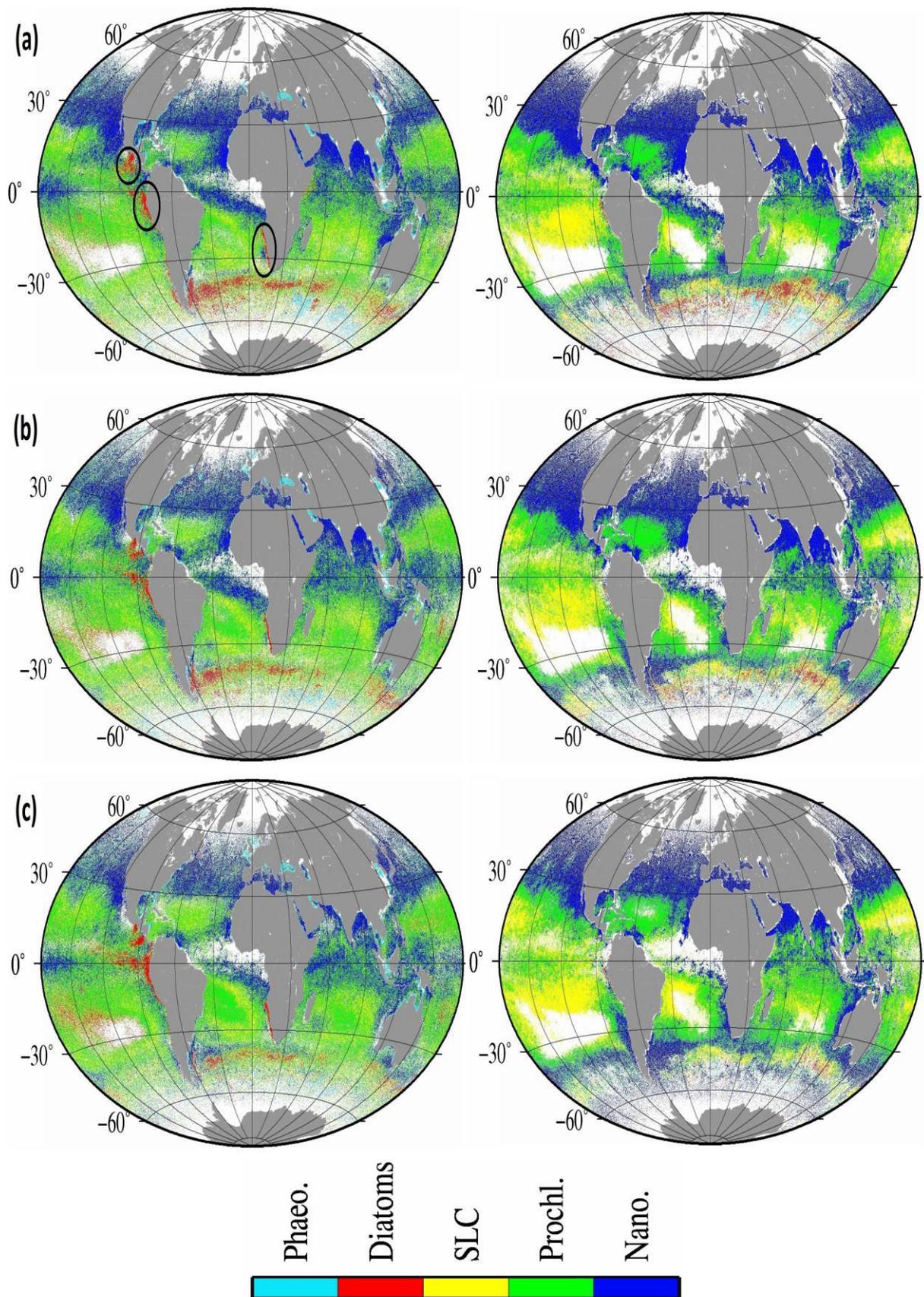


Figure 4.8. Cartes représentant des synthèses mensuelles de la distribution des 5 groupes de phytoplancton entre 1997 et 2010 générées par PHYSAT-SOM (à gauche) et par la version de 2008 de PHYSAT (à droite) pour les mois de Janvier (a), Février (b) et Mars (c)

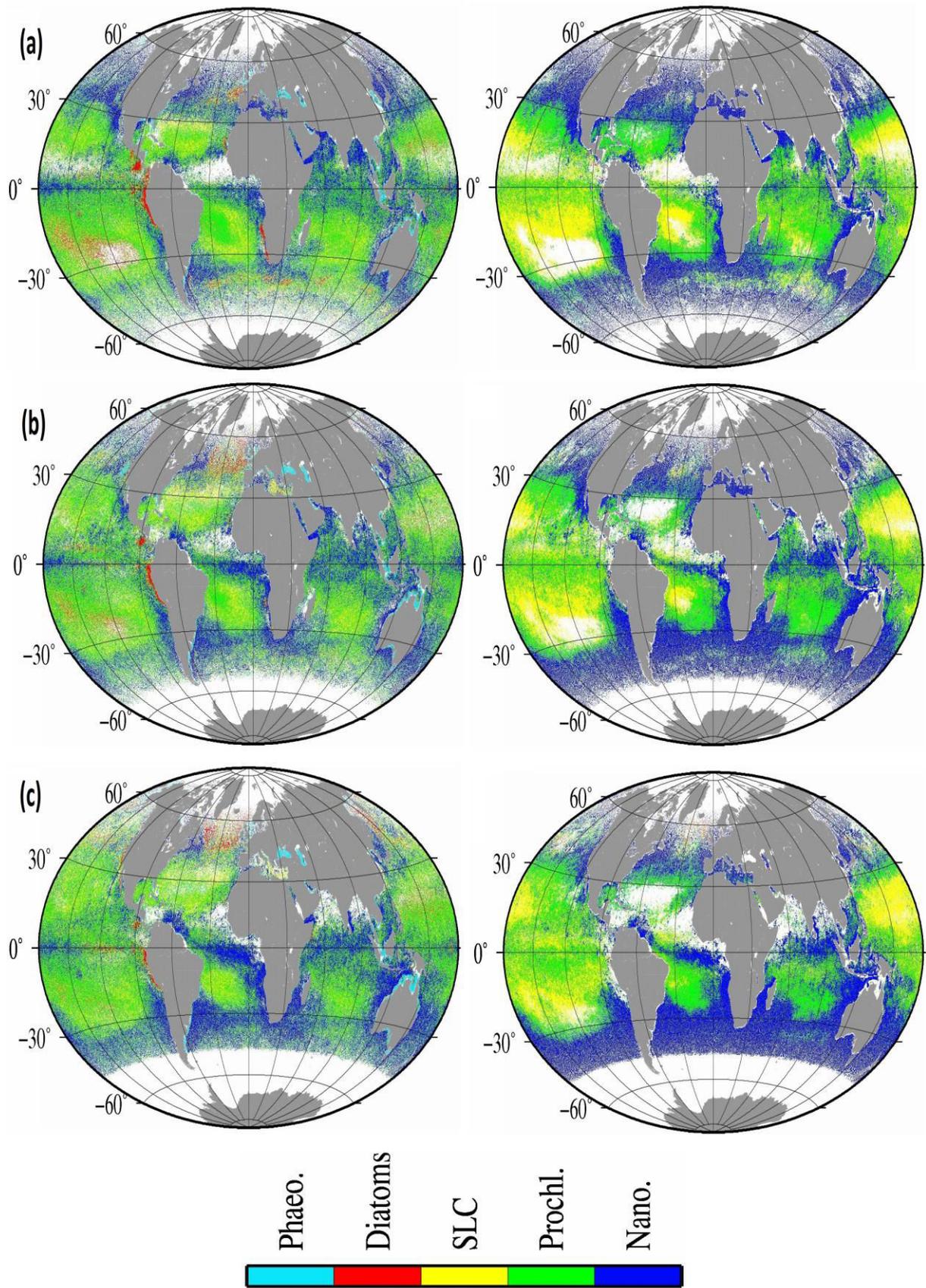


Figure 4.9. Cartes représentant des synthèses mensuelles de la distribution des 5 groupes de phytoplancton entre 1997 et 2010 générées par PHYSAT-SOM (à gauche) et par la version de 2008 de PHYSAT (à droite) pour les mois de Avril (a), Mai (b) et Juin (c)

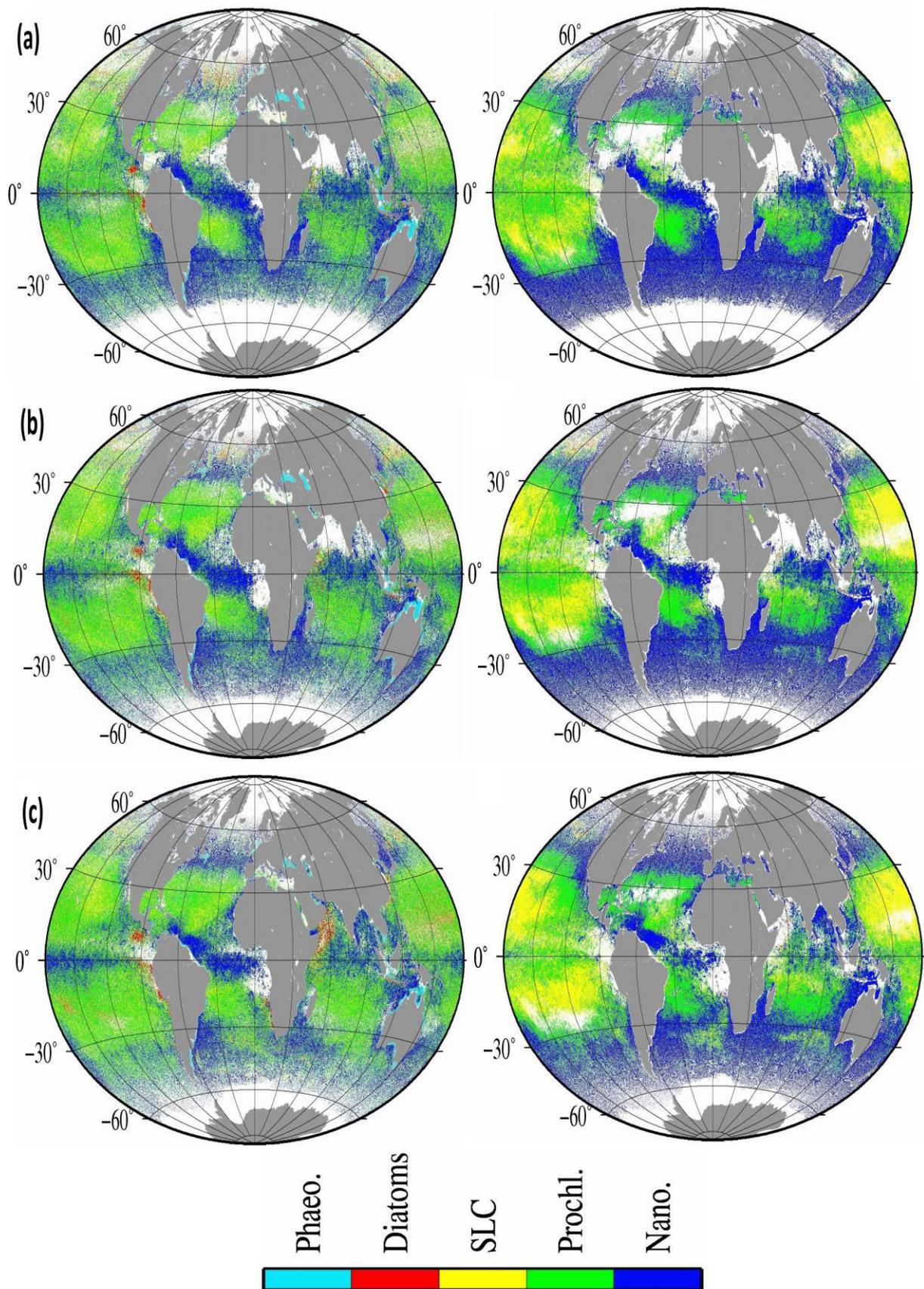


Figure 4.10. Cartes représentant des synthèses mensuelles de la distribution des 5 groupes de phytoplancton entre 1997 et 2010 générées par PHYSAT-SOM (à gauche) et par la version de 2008 de PHYSAT (à droite) pour les mois de Juillet (a), Aout (b) et Septembre (c)

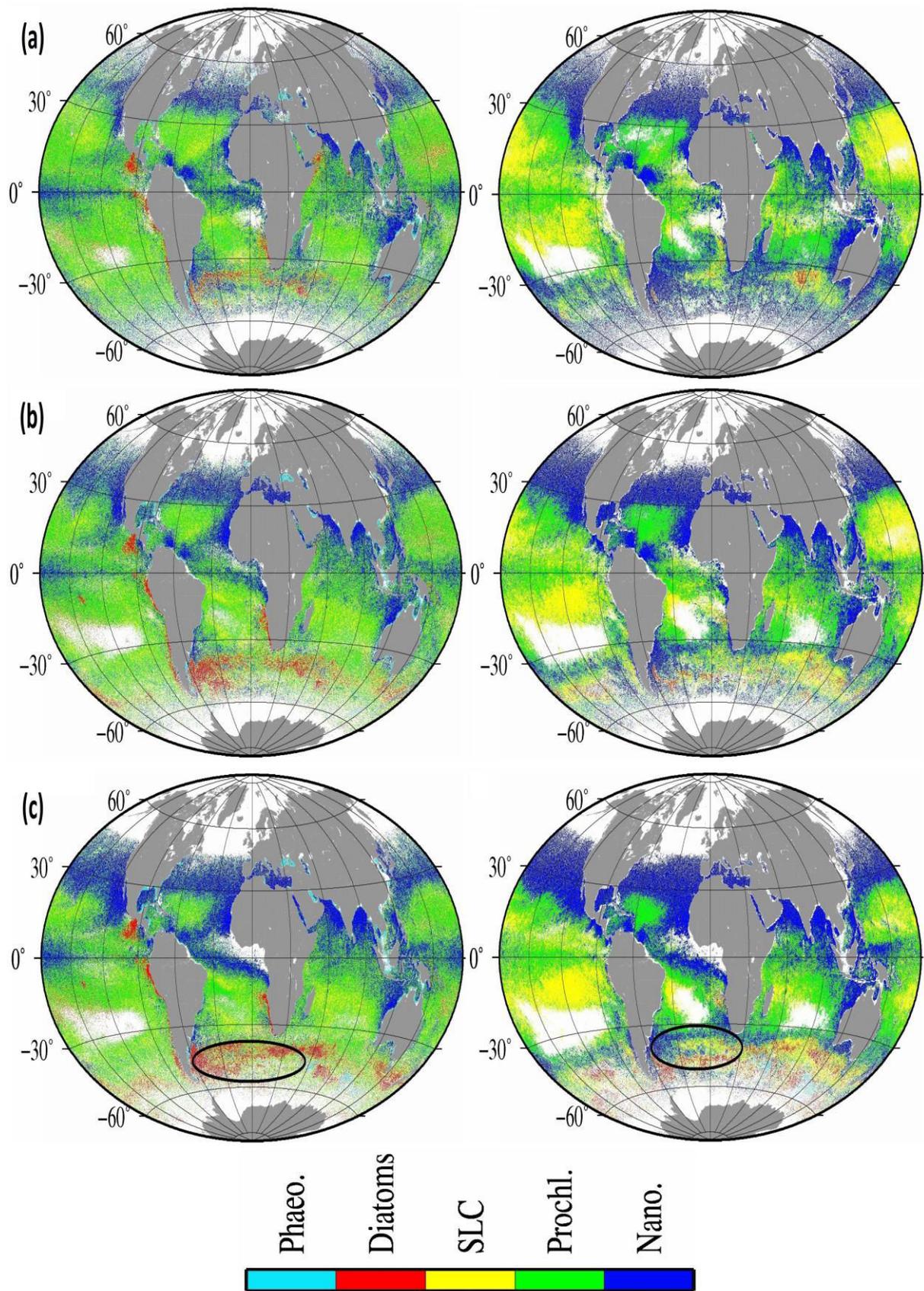


Figure 4.11. Cartes représentant des synthèses mensuelles de la distribution des 5 groupes de phytoplancton entre 1997 et 2010 générées par PHYSAT-SOM (à gauche) et par la version de 2008 de PHYSAT (à droite) pour les mois d'Octobre (a), Novembre (b) et Décembre (c)

4.6. Augmentation de la couverture spatiale avec SOM

Un des avantages majeurs de l'utilisation des cartes auto-organisatrices pour la classification automatique des anomalies spectrales par rapport à l'ancienne approche de PHYSAT, réside dans l'augmentation de la couverture spatiale en pixels labellisés. Le ratio des pixels labellisés par les deux méthodes sur toute la période d'observation du capteur SeaWiFS (climatologies mensuelles entre 1997-2010) est présenté dans le tableau 4.5. Ce ratio constitue le rapport entre le nombre de pixels affectés à chacun des 4 groupes de phytoplancton identifiés à partir des inventaires pigmentaires par PHYSAT-SOM et PHYSAT version 2008 (*Phaeocystis-like* n'est pas inclus étant identifié à partir de la bibliographie seulement).

Tableau 4.5. Rapport entre le nombre de pixels affectés à chacun des 4 groupes de phytoplancton identifiés à partir des inventaires pigmentaires par PHYSAT-SOM et PHYSAT version 2008 (rapport calculé à partir climatologies mensuelles entre 1997-2010)

	Jan	Fév	Mar	Avr	Mai	Jun	Jui	Aou	Sep	Oct	Nov	Dec	Moyenne/groupe
Nanoecaryotes	1,3	1	1,3	1,6	1,8	2	1,5	1,5	1,4	1,5	1,4	1,5	1,48
<i>Prochlorococcus</i>	1,8	1,8	1,8	1,7	1,6	2	2,5	2,2	1,7	2,1	2	1,9	1,92
SLC	2	2	1,9	2,5	2,4	1,7	2,5	2,1	3,2	2,3	2,2	2,1	2,25
Diatomées	2,3	3,5	5,1	5	10	4,1	4,6	5,4	8,1	5,5	5,1	5,5	5,35
Moyenne tous les pixels	1,8	2,1	2,5	2,7	4	2,5	2,8	2,8	3,6	2,8	2,7	2,8	2,75

Les résultats montrent qu'avec PHYSAT-SOM, on labellise, en moyenne et pour les 4 groupes de phytoplancton, entre 1,48 et 5,35 fois plus de pixels qu'avec la version de 2008 de PHYSAT. Les diatomées sont ainsi caractérisées par les plus fortes valeurs, atteignant 10 fois plus de pixels au mois de Mai avec un minimum de 2,3 au mois de Janvier. *Prochlorococcus* et SLC sont détectés deux fois plus en moyenne avec PHYSAT-SOM avec un ratio maximal de 3,21 pour SLC en Septembre. Cependant, il pourrait sembler que les SLC soient sous-estimés par PHYSAT-SOM dans les synthèses mensuelles montrées dans les figures précédentes. Toutefois, la répartition des SLC et des *Prochlorococcus* n'est pas la même entre les deux méthodes et cette apparence est due au fait que les pixels associés à ces deux groupes soient mélangés d'une façon homogène sur les cartes générées par PHYSAT-SOM alors que

pour la version 2008, ces deux groupes se trouvent plus souvent dominants dans des régions distinctes. Ceci est illustré par la figure 4.12 qui montre un zoom sur la distribution des deux groupes de cyanobactéries dans le gyre subtropical de l'Atlantique Sud (synthèse climatologique du mois d'Avril) d'après PHYSAT-SOM (à gauche) et PHYSAT version 2008 (à droite).

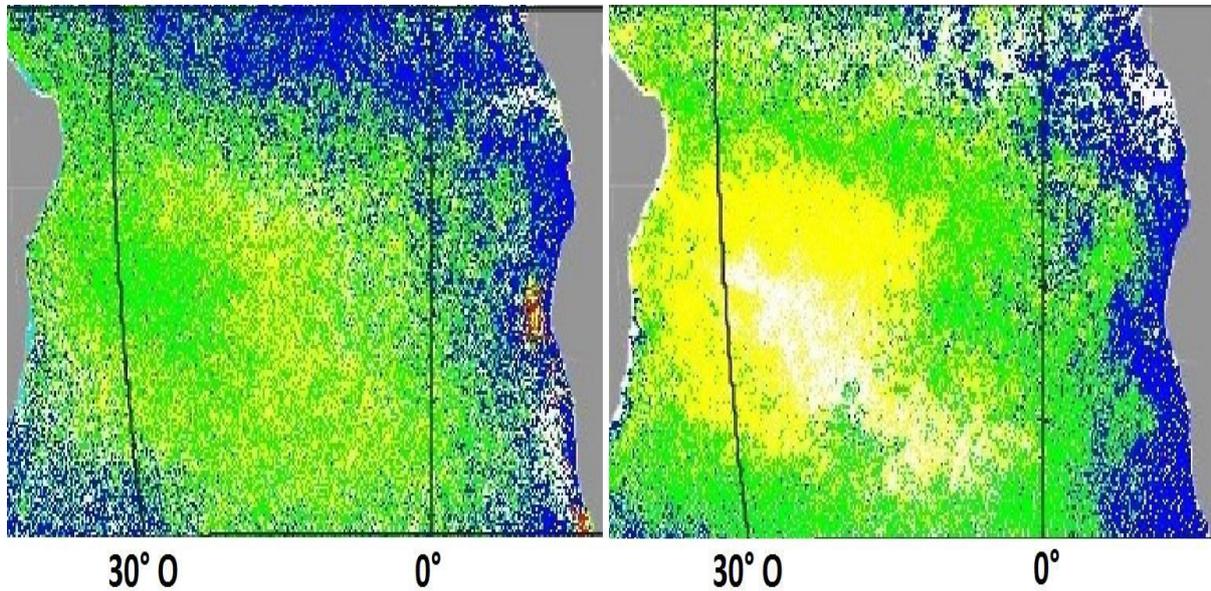


Figure 4.12. Distribution des Prochlorococcus (en vert) et SLC (en jaune) dans le gyre subtropical de l'Atlantique Sud (synthèse climatologique du mois d'Avril sur toute la période d'observation SeaWiFS) d'après PHYSAT-SOM (à gauche) et PHYSAT version 2008 (à droite).

Les nanoeucaryotes sont caractérisés par les plus faibles valeurs de ratio de labellisation relativement aux autres groupes, avec un minimum en Février de 1 et une valeur moyenne de 1,48 fois plus de pixels labellisés sur toute la période d'observation.

Cependant, il est à préciser que les valeurs présentées dans la tableau 4.5 sont le résultat d'un changement de méthode mais aussi de la prise en compte d'une base de données *in situ* plus large (1068 mesures coïncidentes) que celle qui a finalement été utilisée pour déterminer les signatures spectrales des groupes de phytoplancton par la version 2008 de PHYSAT. Rappelons que cette dernière était composée de 41 inventaires pigmentaires coïncidents avec des mesures spatiales. Dans le but de quantifier uniquement l'apport de la nouvelle base de données utilisée dans cette étude (en gardant la même méthode), une labellisation de SOM a été faite en se basant uniquement sur les 41 spectres de Ra correspondant aux inventaires pigmentaires utilisés par la version 2008 de PHYSAT. Le ratio représentant le quotient du nombre de pixels labellisés par PHYSAT-SOM à partir de toute la

base de données *in situ* par le nombre de pixels labellisés par PHYSAT-SOM uniquement avec 41 mesures a été calculé pour les 4 groupes de phytoplancton. Les résultats sont présentés dans le tableau 4.6. Le ratio moyen du total des pixels des 4 groupes sur les synthèses mensuelles de 1997 à 2010 montre qu'on labellise 1,35 fois plus de pixels en utilisant l'entière base de données (en moyenne pour les 4 groupes identifiés à partir des inventaires pigmentaires). On remarque toutefois que ce ratio est très variable en fonction du groupe et du mois. En effet, les nanoeucaryotes sont détectés 2,24 fois plus en utilisant l'entière base de données alors que *Prochlorococcus* est détecté, en moyenne, deux fois moins. Les SLC affichent un ratio de 1,09 et les diatomées sont, en moyenne, détectées 1,5 fois plus en considérant toute la base de données.

Tableau 4.6. Le ratio représentant le quotient du nombre de pixels labellisés par PHYSAT-SOM à partir de toute la base de données *in situ* par le nombre de pixels labellisés par PHYSAT-SOM uniquement avec 41 mesures (rapport calculé à partir climatologies mensuelles entre 1997-2010)

	Jan	Fév	Mar	Avr	Mai	Jun	Jui	Aou	Sep	Oct	Nov	Dec	Moyenne/groupe
Nanoeucaryotes	2,15	2,10	1,89	2,08	2,32	2,58	2,34	2,12	2,10	2,28	2,44	2,53	2,24
<i>Prochlorococcus</i>	0,53	0,54	0,55	0,51	0,46	0,45	0,44	0,47	0,51	0,55	0,56	0,56	0,51
SLC	0,90	0,88	0,87	1,09	1,34	1,39	1,32	1,20	1,09	1,07	1,01	0,91	1,09
Diatomées	1,20	1,32	1,59	1,78	1,65	1,44	1,53	1,66	1,75	1,90	1,48	1,18	1,54
Moyenne tous les pixels	1,20	1,21	1,23	1,37	1,44	1,47	1,41	1,36	1,36	1,45	1,37	1,30	1,35

En comparant les résultats des tableaux 4.5 et 4.6, on peut donc déduire que l'exploitation d'une information *in situ* complémentaire a un impact significatif sur l'identification des groupes de phytoplancton (sauf pour SLC dont le ratio de détection est égal à 1,09), de même que sur le ratio global des pixels labellisés (1,35 fois plus de pixels en utilisant PHYSAT-SOM avec la base de données *in situ* élargie).

Si le rajout de données *in situ* supplémentaires augmente la représentativité spatiale de l'information prise en compte pour labelliser le réseau de neurones et donc, logiquement, la fiabilité des résultats sur une large échelle spatiale, il impacte fortement la détection des groupes de phytoplancton. La labellisation des neurones de SOM étant basée sur le vote majoritaire (c'est le label qui est présent à plus de 50% qui détermine l'étiquette du neurone),

c'est la présence, l'absence et la proportion des signaux spécifiques à chaque groupe phytoplanctonique qui vont déterminer à la fois si le neurone est labellisé et, si c'est le cas, quel sera son label. Cette conclusion est logique puisque les neurones sont labellisés selon l'information qu'ils « captent ». Par conséquent, une information « manquante » ou « additionnelle » (selon si on a une information *in situ* ou pas sur une zone donnée) peut significativement changer l'information globale classée par chaque neurone, ce qui va modifier son étiquette (label).

4.7. Conclusion

La carte auto-organisatrice, précédemment entraînée avec un jeu de spectres d'anomalies de luminances, a été labellisée avec une base de données *in situ* de pigments et exploitée pour détecter les principaux groupes de phytoplancton dominants à l'échelle globale. Quatre de ces groupes (Nanoeucaryotes, *Prochlorococcus*, SLC et diatomées) ont été identifiés à partir des inventaires pigmentaires. Les différentes étapes de cette méthode, basée sur la classification automatique des anomalies spectrales et leur labellisation avec SOM, ont été regroupées sous le nom de PHYSAT-SOM.

La nouvelle méthode a été appliquée aux données journalières du capteur SeaWiFS sur la période s'étendant de Septembre 1997 à Décembre 2010. L'utilisation de SOM a permis de faire une analyse du signal d'anomalies spectrales relié à la dominance des groupes de phytoplancton. Les observations de PHYSAT-SOM ont été validées avec un jeu de données *in situ* indépendant. Même si la répartition de ces groupes reste globalement en accord avec celle trouvée lors de l'application de la version 2008 de PHYSAT, des différences notables existent pour tous les groupes de phytoplancton étudiés. Ainsi, bien que les cyanobactéries (*Prochlorococcus* et SLC) dominent globalement les eaux oligotrophes équatoriales et subéquatoriales d'après les deux approches, des différences significatives existent entre leurs répartitions relatives et ils ne dominent pas de la même façon ces zones. Les blooms de diatomées dans certaines régions telles que les upwellings côtiers et l'Océan Austral sont plus clairement détectés avec PHYSAT-SOM. L'étude des signatures spectrales associées aux diatomées et l'amélioration de la détection de leurs blooms dans certaines régions seront détaillés dans le prochain chapitre.

L'utilisation d'une classification automatique permet d'établir une discrétisation fine des Ra, indépendamment de toute notion de labellisation reliée à la disponibilité de données *in situ*. C'est alors l'entière variabilité des anomalies spectrales qui est prise en compte. Par ailleurs, il est apparu que l'utilisation de SOM permettait d'augmenter le nombre de pixels labellisés. Cette amélioration est due, à la fois à la meilleure prise en compte de la variabilité spectrale des Ra par rapport à la version 2008 de PHYSAT (utilisation des tubes et introduction de seuils sur la forme et l'amplitude des spectres, voir paragraphe 3.4.3 du chapitre 3 et le 2.5 du chapitre 2) et à la contribution d'une large base de données *in situ*, couvrant une grande variabilité de conditions environnementales et bio-optiques et efficacement exploitable avec SOM. L'utilisation de données *in situ* complémentaires impacte les proportions relatives des groupes détectés, de même que le ratio total moyen des pixels labellisés.

CHAPITRE 5

Distribution spatio-temporelle des signatures spectrales associées aux diatomées et amélioration de la détection de leurs blooms avec PHYSAT-SOM

5.1. Introduction

Dans le précédent chapitre, l'analyse des signatures spectrales moyennes (spectres référents des neurones labellisés) des différents groupes de phytoplancton identifiés à partir des inventaires de pigments a montré qu'elles étaient globalement proches en termes de formes et d'amplitudes de celles trouvées par Alvain et al. (2005), lors du développement de PHYSAT. L'exploitation de la base de données élargie avec SOM a cependant permis d'associer les diatomées à deux différentes signatures spectrales (figure 4.4 d et e du chapitre 4). En effet, les spectres de Ra coïncidents avec des mesures *in situ* de pigments caractérisant les diatomées ont été affectés à des neurones éloignés dans la carte auto-organisatrice (spectres référents de formes et amplitudes différentes). La question qui se pose est donc: quelles seraient les différentes hypothèses susceptibles d'expliquer le fait que les diatomées soient associées à différentes signatures spectrales de Ra ?

Les liens entre les anomalies spectrales et l'environnement bio-optique (Alvain et al., 2012), présentés dans le chapitre 2 (paragraphe 2.4.2) ont montré que la forme et l'amplitude des spectres de Ra étaient étroitement liés aux propriétés optiques inhérentes (IOP) du groupe considéré et du milieu (b_{bp} , a_{cdom} et a_{phy}). Les valeurs de ces paramètres varient selon plusieurs facteurs (e.g. taille des particules, présence de matières organiques dissoutes, produits de dégradation des cellules phytoplanctoniques). Dans ce qui suit, nous nous focalisons sur les diatomées, groupe de phytoplancton aux floraisons particulièrement intenses et qui représente un cas d'étude intéressant du fait de son rôle majeur dans l'exportation de la matière organique vers le fond des océans et dans la production primaire (Nelson et Brzezinski, 1997 ; Brzezinski et al., 1998). Les diatomées sont aussi connues pour constituer un bon indicateur de la pollution organique des eaux (Sudhakar et al., 1994 ; Birkett et Gardiner, 2005 ; Sladeczek, 1986). Cependant, la question peut de même se poser pour *Prochlorococcus* qui

affiche aussi des différences entre les formes de spectres de Ra associés (figure 4.4 b). Les nanoeucaryotes (figure 4.4 a) et SLC (figure 4.4 c) présentent, quant à eux, des signatures spectrales relativement homogènes en terme de formes.

Pour examiner les différentes hypothèses susceptibles d'expliquer les différences des Ra associés aux diatomées, on a choisi d'étudier leur distribution spatio-temporelle dans certains environnements ainsi que les valeurs des IOP associées. L'objectif est de voir si ces différents signaux pourraient, par exemple, être associés à différentes formes de diatomées, localisées dans des environnements différents ou à la présence de différentes formes de diatomées à des moments différents. Ces microalgues unicellulaires, caractérisés par la présence d'un squelette externe siliceux (appelé frustule), présentent une très grande variété de formes et de tailles (figure 5.1). Elles sont usuellement divisées en deux ordres selon la forme de leur frustule : l'ordre des *Biddulphiales* (ou diatomées pennées), dont le squelette possède une symétrie bilatérale (forme allongée) et l'ordre des *Bacillariales* (ou diatomées centrales), dont le squelette est caractérisé par une symétrie axiale (forme circulaire).

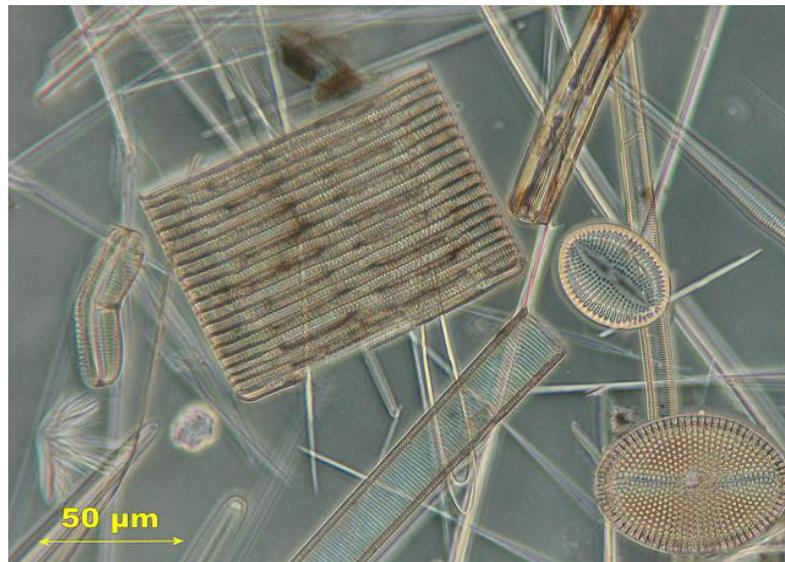


Figure 5.1. Diversité des formes et tailles de différentes diatomées

<http://www.micromagus.net>

A l'échelle globale, la littérature montre que plus de 900 genres de diatomées sont cosmopolites (Fourtanier et Kociolek, 2003). Il y a cependant quelques exceptions. Par exemple, une étude, faite sur le genre *Thalassiosira*, a révélé que, parmi les quelques espèces se trouvant dans les eaux tempérées à froides, une espèce est cosmopolite, une est restreinte à

l'hémisphère Nord et deux autres à l'Océan Austral. Les autres espèces sont endémiques aux eaux chaudes (Vanormelingen et al., 2008).

A travers la compilation d'une base de données à l'échelle globale comprenant l'abondance, la biomasse et le biovolume de nombreuses espèces de diatomées, Leblanc et al. (2012) ont montré, qu'en termes d'abondance, les genres dominants sont *Rhizosolenia* (17,4 %), *Chaetoceros* (14,5 %) et *Thalassiosira* (12,6 %). Les diatomées centriques ont été reportées comme étant celles qui contribuent majoritairement à la biomasse totale (86 %) et la forme cylindrique est globalement dominante. Cependant, peu de preuves existent actuellement quant à l'endémisme de genres ou familles de diatomées, qui sont généralement connus pour constituer un groupe de phytoplancton cosmopolite (Vanormelingen et al., 2008).

Dans la première partie du présent chapitre, nous analysons la variabilité spatio-temporelle, dans certains environnements, des différentes signatures spectrales de Ra caractérisant les diatomées (mises en évidence par PHYSAT-SOM dans le chapitre 4). Dans sa seconde partie, nous montrons les améliorations apportées par PHYSAT-SOM (par rapport à PHYSAT version 2008) quant à la détection des blooms de diatomées dans des conditions spécifiques (i.e. upwellings, blooms succédant à un évènement climatique).

5.2. Distribution spatio-temporelle des différentes signatures spectrales des diatomées

5.2.1. Distribution spatiale des Ra et relations avec la composition phytoplanctonique et les propriétés optiques inhérentes du milieu

Dans le chapitre 4, les figures 4.8 à 4.11 ont montré que, même sur des climatologies mensuelles représentant des synthèses sur toute la période d'observation du satellite, des blooms de diatomées étaient clairement visibles au niveau des zones d'upwellings (e.g. Pérou, Benguela) ainsi qu'au niveau de l'Océan Austral durant l'été. L'objectif de cette analyse est de voir si les différentes signatures spectrales de Ra caractérisant les diatomées dans ces régions correspondent à des zones géographiques distinctes ou pas. Si c'est le cas, il serait question de voir si les différences entre ces spectres de Ra pourraient être expliquées par la présence de diatomées aux tailles et formes non identiques ou par leur présence dans des

stades de développements différents. Il pourrait par exemple s'agir de différents ordres/formes de diatomées (diatomées pennées ou centrales) ou même de différentes espèces (coloniales ou non par exemple).

Les fréquences relatives des pixels associés aux 5 neurones labellisés « diatomées » durant le mois de Janvier 2002 (synthèse mensuelle à partir des cartes journalières) sont représentées dans la figure 5.2 au niveau de deux zones géographiques caractérisées par des floraisons de diatomées : le Pacifique Equatorial au niveau de l'upwelling du Pérou (à gauche) et l'Océan Atlantique Sud (à droite). Pour un pixel valide donné (pour lequel on a une information), la fréquence mensuelle relative à chacun des 5 neurones représente, dans ce cas d'étude, le total des occurrences de ce neurone, divisé par le total des occurrences de tous les autres neurones labellisés « diatomées » au niveau de ce pixel. Les spectres référents de chaque neurone sont représentés à la droite de la figure. La figure 5.2 montre que les différents neurones ne sont pas uniformément présents dans les mêmes environnements. Par exemple, le neurone 90 (représenté par le second type de spectre référent) est plus fréquemment localisé près des côtes ouest sud-américaines au niveau de l'upwelling du Pérou, alors que le neurone 10 (représenté par le premier type de spectre) peut également être observé au large des côtes sud-américaines ainsi qu'au niveau de l'Atlantique Sud. On remarque aussi que le neurone 100 est présent dans des endroits distincts de ceux des autres neurones.

L'analyse des échantillons d'eau prélevés au niveau de l'upwelling du Pérou (Hutchins et al., 2002) a montré que la population de phytoplancton était globalement caractérisée par la dominance de diatomées pennées de petite taille (exemple montré dans la figure 5.3 de gauche) ainsi que par la présence d'haptophytes (tels que coccolithophores et *Phaeocystis globosa*) (Bruland et al., 2005). En revanche, au niveau du plateau continental péruvien, il existe des zones (appelées « eaux brunes du Pérou»), où la remontée d'eaux riches en fer conduit à l'efflorescence massive de larges cellules de diatomées centrales (e.g. *Coscinodiscus*, exemple montré dans la figure 5.3 de droite) avec des concentrations en chlorophylle *a* supérieures à 45 mg/m³ (DiTullio et al., 2005). L'abondance de ces larges cellules de diatomées augmente proportionnellement à l'élévation de la concentration en fer (Bruland et al., 2005). A l'inverse, au large du plateau continental péruvien (zones appelées « eaux bleues », il existe des endroits où la concentration en chlorophylle *a* est faible relativement à celle de nutriments (zones appelées HNLC pour « High Nutrient-Low

Chlorophyll). Ceci est dû au fait que la limitation en fer ralentit la croissance du phytoplancton, en particulier des diatomées qui exigent une concentration élevée en fer (Bruland et al., 2005).

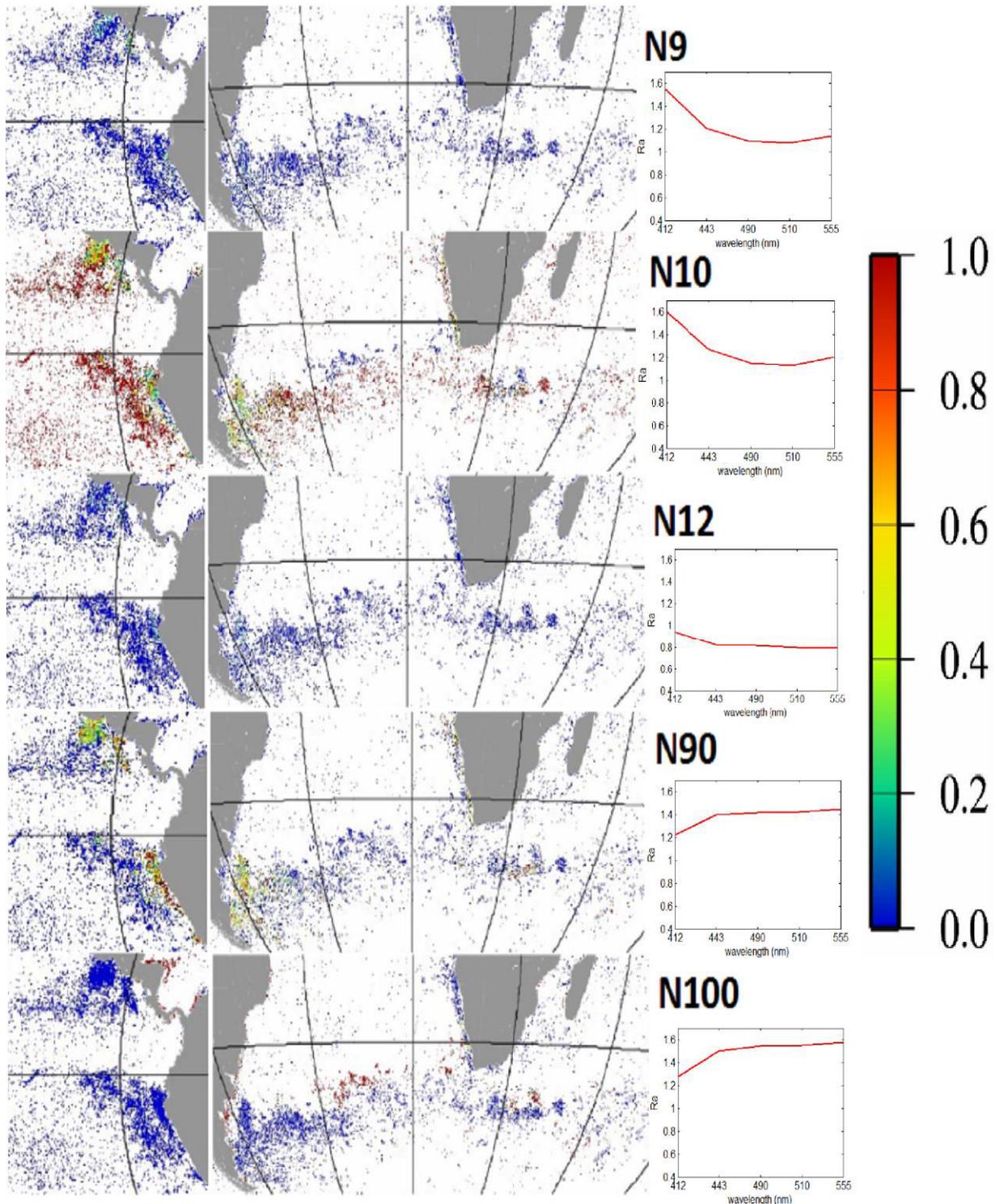
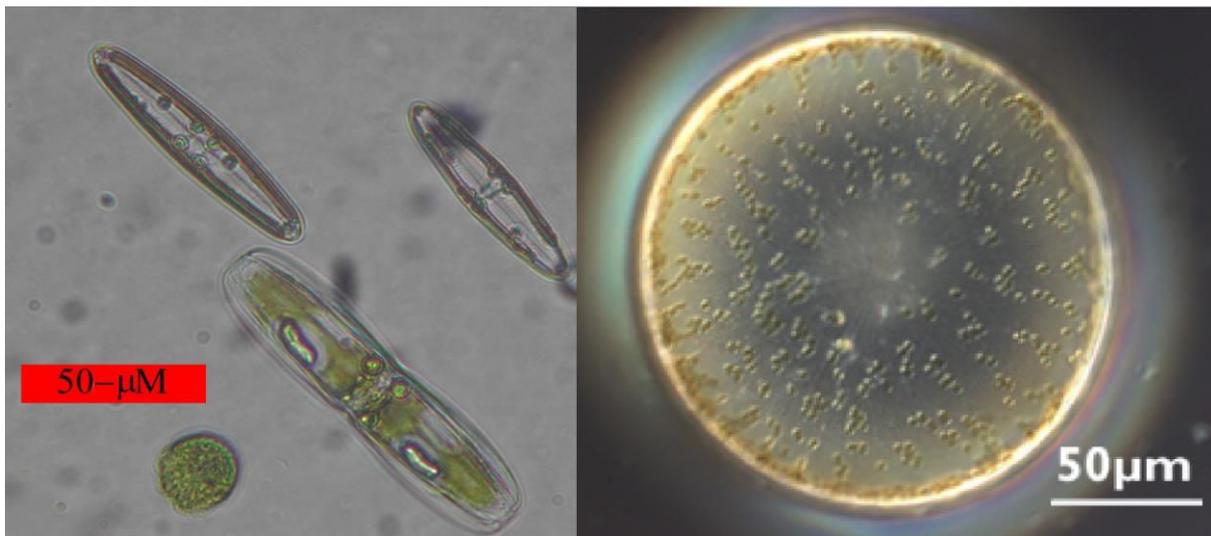


Figure 5.2. Fréquences relatives d'apparition des cinq neurones labellisés « diatomées » dans le Pacifique Equatorial (à gauche) et l'Atlantique Sud (à droite) pour le mois de Janvier 2002. Chaque neurone est représenté avec son vecteur référent. A noter que dans ce cas d'étude, la fréquence relative est calculée uniquement par rapport aux occurrences des neurones labellisés « diatomées ».

Ce contraste « eaux bleues / eaux brunes » est dû à l'assimilation massive de fer qui représente un facteur limitant dans les eaux du large, aux processus de lessivage ainsi qu'au transport d'Ekman. Ces différents facteurs conduisent à une décroissance de la concentration en fer, et par conséquent, à un changement radical des communautés phytoplanctoniques. DiTullio et al. (2005) montre ainsi que la zone frontalière est dominée par des cryptophytes et, au large, la population phytoplanctonique devient caractérisée par la dominance de diatomées pennées de petite taille.



*Figure 5.3. Différentes formes de diatomées pennées (à gauche) et une diatomée centrale du genre *Coscinodiscus* (à droite) (<http://www.exploratorium.edu>)*

La figure 5.2 montre que le spectre référent du neurone 90 est présent uniquement près des côtes alors que celui du neurone 10 est plus fréquemment détecté au large et, dans une moindre mesure, dans la zone du plateau continental. Ce gradient entre « eaux bleues » du large et « eaux brunes » du plateau continental plus chargées en chlorophylle (un exemple, illustrant ce gradient, est montré dans la figure 5.4), où des espèces plus larges dominent, pourrait être à l'origine des différents spectres de Ra associés aux diatomées dans cette région. Toutefois, ceci reste une hypothèse car les auteurs montrent aussi que les deux formes de diatomées peuvent être simultanément présentes dans de mêmes environnements et que leur dominance relative dépend de la concentration en fer et des autres nutriments, de l'efficacité photosynthétique (rapport entre la production photosynthétique et l'énergie consommée) ainsi que du site d'échantillonnage (Hutchins et al., 2002 ; Bruland et al., 2005 ; DiTullio et al., 2005).

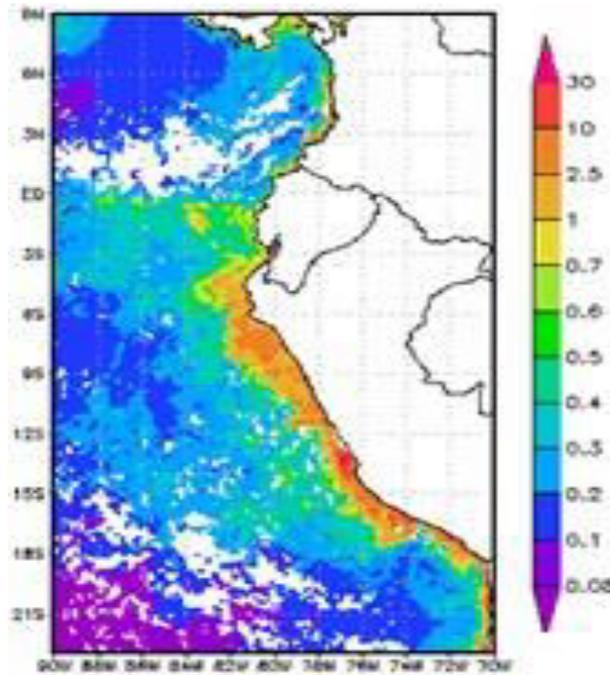


Figure 5.4. Carte illustrant le gradient de concentration en chlorophylle a (mg/m^3) au niveau de l'upwelling du Pérou en Décembre 2004 (gradient entre « eaux bleues » du large et « eaux brunes » du plateau continental plus chargées en chlorophylle) (<http://disc.sci.gsfc.nasa.gov>)

Cependant, au-delà des différences en termes de composition spécifique du groupe diatomées (variabilité intra-groupe), d'autres facteurs peuvent être à l'origine des différentes signatures spectrales observées, tels que la présence plus ou moins importante de matières organiques dissoutes (CDOM). Ces dernières sont en effet connues pour avoir un impact significatif sur la forme des spectres de R_a , particulièrement dans les courtes longueurs d'onde (Alvain et al., 2012).

Dans le but d'analyser la variabilité bio-optique sous-jacente aux différentes signatures spectrales observées pour les diatomées, les principales propriétés optiques inhérentes constituant l'environnement bio-optique influant sur la forme et l'amplitude des spectres de R_a (Alvain et al., 2012) ont été calculées. Les valeurs d'absorption par les matières organiques dissoutes (a_{cdom}), de la diffusion par les particules (b_p) et de l'absorption par le phytoplancton (a_{phy}) ont été calculées en utilisant des relations théoriques de référence.

Ainsi, pour les 5 longueurs d'onde, a_{phy} a été calculé selon l'équation suivante :

$$a_{\text{phy}}(\lambda) = A [\text{Chl } a]^E \quad (\text{Eq. 5.1})$$

où A et E représentent les coefficients calculés par Bricaud et al. (1998) en utilisant une large base de données de mesures *in situ* collectées dans des eaux oligotrophes, mésotrophes et eutrophes.

L'absorption par les matières organiques dissoutes (a_{cdom}) a été calculée d'après les méthodes de Bricaud et al., 1998 et Morel, 2009, suivant la formule suivante :

$$a_{\text{cdom}}(\lambda) = 0,065 [\text{Chl } a]^{0,75} \exp(-0,014(\lambda - 443)) \quad (\text{Eq. 5.2})$$

Enfin, la diffusion par les particules (b_p) et sa dépendance spectrale ont été calculées d'après les méthodes de Morel et Maritorena (2001) et Loisel et al. (2006) :

$$b_p(\lambda) = 0,30 [\text{Chl } a]^{0,62} \left(\frac{\lambda}{660}\right)^{-\gamma} \quad (\text{Eq. 5.3})$$

où γ représente la pente spectrale calculée en fonction de la concentration en chlorophylle a :

$$\gamma = -0,55 \log[\text{Chl } a] + 1,6 \quad (\text{Eq. 5.4})$$

Ces 3 paramètres ont été calculés pour le mois de Janvier 2002, au niveau des zones correspondant à la localisation des 3 neurones « diatomées » les plus fréquemment observées (neurone 10 : premier type de spectre référent ; neurones 90 et 100 : second type de spectre référent). Pour chaque neurone, les intervalles de valeurs de ces paramètres sont montrés dans la figure 5.5. Des différences significatives entre les valeurs des 3 paramètres relatives aux trois neurones sont observées sur toutes les longueurs d'onde (p-value < 0,0001 au seuil de signification de 0,05). Les valeurs d' a_{cdom} sont plus élevées pour les neurones 90 et 100 que le neurone 10, en particulier à 412 nm (figure 5.5 a). Les amplitudes des anomalies spectrales dans les courtes longueurs d'onde (dans le bleu, autour de 412 nm) étant principalement déterminées par les valeurs d' a_{cdom} (Alvain et al., 2012), ce paramètre pourrait être à l'origine des valeurs de Ra relativement faibles à 412 nm pour les neurones 90 et 100 par rapport au neurone 10. Ce dernier étant associé à de plus faibles valeurs d' a_{cdom} .

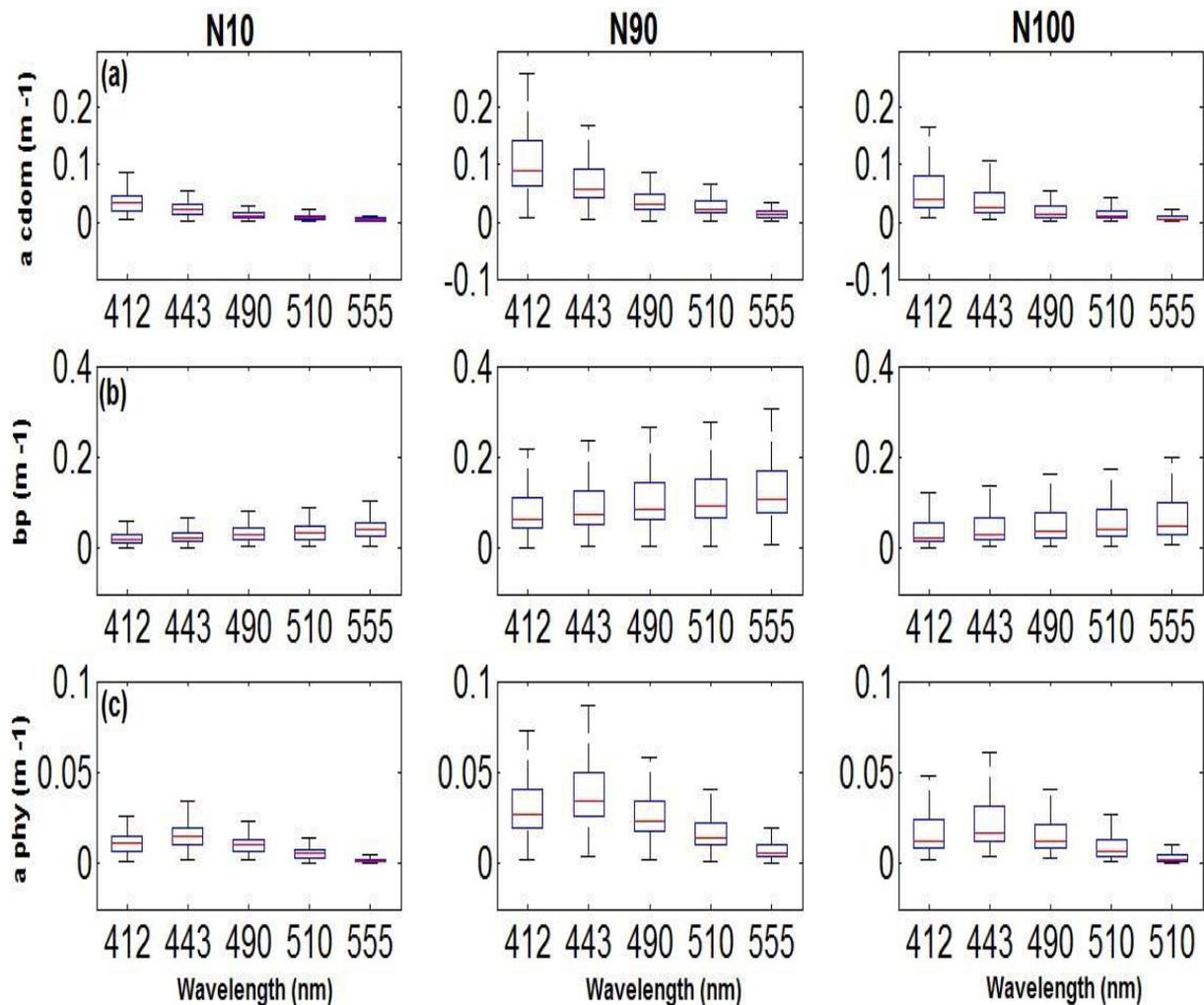


Figure 5.5. Intervalles de valeurs des propriétés optiques inhérentes correspondant aux neurones 10 (premier type de spectre référent), 90 et 100 (deuxième type de spectre référent) pour le mois de Janvier 2002. Les barres verticales représentent les écarts-type et le trait rouge horizontal les valeurs médianes.

Selon l'exemple de localisation géographique des différents neurones (figure 5.2) et les analyses bibliographiques détaillées précédemment, les valeurs relativement fortes d' a_{cdom} associées au neurone 90 pourraient alors être associées à:

- La présence plus fréquente près des côtes de ce neurone par rapport au neurone 10, ce qui pourrait indiquer une influence côtière plus importante.
- La prédominance d'espèces de diatomées plus larges (e.g. *Coscinodiscus*) au niveau des eaux du plateau continental (e.g. contraste « eaux brunes » riches en chlorophylle *a* du plateau continental Péruvien par rapport aux « eaux bleues » du large), dont les périodes de fin de blooms peuvent être caractérisées par de fortes concentrations en matières organiques dissoutes (dégradation des cellules).

La rétrodiffusion par les particules détermine, en majeure partie, les amplitudes des anomalies spectrales sur toutes les longueurs d'onde (Alvain et al., 2012). Les spectres référents des neurones 90 et 100 étant associés à des valeurs de rétrodiffusion plus fortes que celles associées au neurone 10 (figure 5.5 b), ceci pourrait expliquer les plus fortes valeurs de R_a caractérisant ces spectres entre 443 nm et 555 nm. La rétrodiffusion dépend de plusieurs facteurs tels que la taille des cellules ainsi que l'effet d'agrégation de plusieurs cellules (appelé aussi effet de « packaging »). La formation des agrégats peut survenir durant ou après les efflorescences des diatomées à cause de l'effet de cohésion (adhérence) des cellules (Kiørboe et al., 1998 ; Burd et Jackson 2009). Selon l'analyse bibliographique précédente, les valeurs de b_p plus élevées pour les neurones 90 (et dans une moindre mesure le neurone 100) par rapport au neurone 10 pourraient alors être associées à la prédominance de diatomées plus larges au niveau des zones où ces neurones sont fréquemment observés.

Les valeurs d' a_{phy} (figure 5.5 c) sont plus élevées pour les neurones 90 et 100 que le neurone 10. Pour les 3 neurones, les valeurs les plus fortes d' a_{phy} s'observent à 443 nm, longueur d'onde à laquelle le phytoplancton présente son maximum d'absorption.

Les différentes signatures spectrales de R_a sont déterminées par des proportions différentes de valeurs d'IOP, créant l'équilibre de l'environnement bio-optique sous-jacent (Alvain et al., 2012). Ces observations montrent que les différents neurones, associés au même groupe de phytoplancton en se basant sur les données pigmentaires, se trouvent dans des environnements où les valeurs des IOP sont significativement différentes. Ces paramètres représentant le principal facteur impactant les valeurs des anomalies spectrales (Alvain et al., 2012), les différences observées expliquent directement l'existence de plusieurs signatures spectrales pour les diatomées.

Dans l'Océan Austral, la majorité des études ayant analysé la biogéographie des diatomées ont été réalisées à partir de données paléo-océanographiques (Crosta et al., 2005 ; Romero et al., 2005 ; Armand et al., 2005, 2008). Ces différentes méthodes consistent en l'analyse des différents genres et espèces de diatomées présentes dans les carottes sédimentaires ou glaciaires et leur mise en relation avec les conditions environnementales et météorologiques passées (principalement la disponibilité en nutriments, la température de surface et la surface couverte par la glace).

Par ailleurs, durant la campagne PESO (Pilot Expedition to the Southern Ocean) conduite au niveau de l'océan Austral, Mohan et al. (2011) ont constaté que, parmi toutes les espèces de diatomées identifiées (e.g. *Chaetoceros* spp., *Navicula* spp.), *Fragilariopsis kerguelensis* (figure 5.6) était l'espèce dominante et qu'elle constituait plus de 90 % de l'abondance totale des diatomées à partir de 41° Sud vers les latitudes plus élevées (vers le pôle) le long du tracé. Toutes les autres espèces étaient caractérisées par une faible abondance et une répartition clairsemée. Les données récoltées durant cette campagne ont toutefois montré une grande variabilité de taille des cellules. Celle-ci augmentait avec la chute de la température de surface et l'élévation de la concentration en nutriments (en particulier le fer).

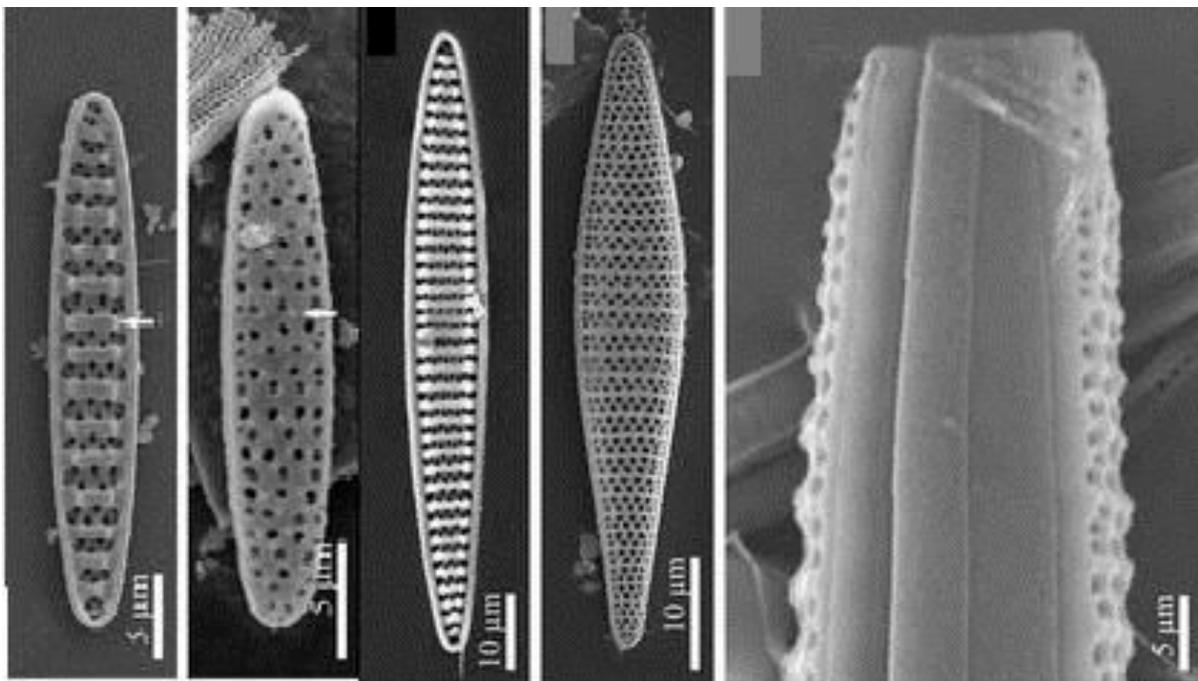


Figure 5.6. Différentes tailles et morphologies d'une espèce de diatomée pennée dominante dans les eaux de l'Océan Austral : *Fragilariopsis kerguelensis* (figure extraite de l'étude faite par Cefarelli et al. (2010) dans la mer d'Argentine et l'Océan Austral)

De nombreuses autres études ont souligné la dominance de *Fragilariopsis kerguelensis* dans les eaux de l'Océan Austral (Armand et al., 2005, 2008 ; Crosta et al., 2005 ; Romero et al., 2005 ; Zielinski et Gersonde, 1997 ; Cefarelli et al., 2010 ; Wright et al., 2010). D'autres études ont mis en évidence les mécanismes de la variabilité de la taille du squelette siliceux de ces cellules en rapport avec la température de surface (elle-même en rapport avec l'étendue de la glace polaire), la concentration en fer, ainsi que l'apport de poussières atmosphériques par le transport éolien (Cortese et Gersonde, 2007 ; Cortese et al.,

2012; Shukla et al., 2013). Cette diversité de taille intra-spécifique permet de poser l'hypothèse que les différentes signatures spectrales de Ra observées pour les diatomées ne sont nécessairement pas associées à différents genres, familles ou espèces mais pourraient également être associées à la même espèce, pouvant se trouver à différentes tailles ou stades de développement selon les conditions environnementales existantes. En effet, tel que précisé précédemment, la taille des particules impacte directement les valeurs de rétrodiffusion, elles-mêmes déterminant l'amplitude des spectres de Ra (Alvain et al., 2012).

Ainsi, d'après les analyses précédentes, il apparaît que les différentes signatures spectrales de Ra associées aux diatomées et leurs localisations géographiques observées sur la figure 5.2 pourraient correspondre à :

- La présence de différents genres ou espèces (dont la distribution est à mettre en relation avec la variabilité des conditions environnementales et la disponibilité des nutriments, en particulier le fer).
- La présence de mêmes espèces se trouvant à différents stades de développement, tel que l'illustre l'exemple de *Fragilariopsis kerguelensis* dans l'Océan Austral.
- La variabilité des propriétés optiques inhérentes qui explique directement celle des anomalies spectrales (Alvain et al., 2012). La variabilité de ces paramètres peut être mise en relation avec la composition phytoplanctonique, de même que l'influence d'autres composés non phytoplanctoniques (e.g. matière organiques dissoutes).

Toutefois, ces hypothèses restent à vérifier à l'avenir avec d'avantages de mesures *in situ* dans les régions considérées.

5.2.2. Variabilité temporelle des anomalies spectrales dans l'Océan Austral

L'objectif de cette analyse est de voir si, dans une région donnée, différentes signatures spectrales de Ra des diatomées se caractérisent par des cycles saisonniers spécifiques. Si c'est le cas, cela voudrait dire que ce signal pourrait être associé, par exemple, à l'existence de différentes diatomées à des moments différents. Il ne s'agit pas nécessairement de différents genres ou espèces mais le signal pourrait être également associé à une même espèce aux tailles différentes (exemple de *Fragilariopsis kerguelensis* cité

précédemment) ou se trouvant à différents stades de blooms. Cette analyse n'a toutefois pas comme objectif d'associer les différents signaux observés à une diatomée particulière car peu d'informations *in situ* existent actuellement. Jusqu'à présent, la plupart des études concernant la saisonnalité des différentes diatomées ont dressé un inventaire des espèces dominantes et les mécanismes régissant leurs apparitions dans des zones géographiques précises (Patagonie et Kerguelen par exemple), en rapport avec diverses conditions environnementales, tels que limitation en fer (e.g. HLNC) ou limitation en lumière (Hoffmann et al., 2008). Généralement, la distribution des diatomées est connue pour être majoritairement contrôlée par l'épaisseur de la couche de mélange. Ce groupe phytoplanctonique préfère les eaux stratifiées et riches en nutriments, qui vont permettre l'initiation de la floraison (Weber & El-Sayed, 1987; Jochem et al., 1995).

Concernant l'Océan Austral, même si certaines espèces ont été recensées comme étant dominantes dans cette zone (e.g. *Fragilariopsis kerguelensis*), on ne trouve toutefois pas d'information concernant la succession des différentes espèces de diatomées dominantes sur une échelle pluriannuelle. Toutefois, nous nous sommes intéressés au cycle saisonnier des différents neurones labellisés « diatomées » dans l'Océan Austral entre 40° et 70° Sud. La figure 5.7 représente les fréquences relatives d'apparition des cinq neurones labellisés « diatomées » entre 1998 et 2005 (calculées sur des moyennes mensuelles uniquement par rapport aux occurrences des neurones « diatomées », tel que décrit dans le paragraphe 5.2.1). Notons que durant l'hiver austral (Juin, Juillet, Aout), la fréquence est calculée à partir d'un nombre réduit de pixels valides étant donné que le capteur n'enregistre pas de données à partir d'une certaine latitude (environ 50° Sud) du fait de l'angle d'inclinaison de la Terre durant cette période.

La figure 5.7 montre que les différentes signatures spectrales (Ra) des diatomées ne sont pas caractérisées par le même cycle saisonnier. Ceci va dans le sens de l'hypothèse, stipulant que les différents spectres de Ra pourraient être associés à des diatomées différentes (taxonomiquement et/ou en termes de tailles) présentes à des moments différents de l'année ou à différents stades de développement. Le neurone 9 présente ainsi un pic de fréquence en début de période estivale (dans l'hémisphère sud) vers le mois de Décembre avec un pic moins élevé autour des mois de Février et Mars. Le neurone 10 est caractérisé par deux pics distincts, le premier à la même période que le neurone 9 et un autre pic de fréquence élevée autour du mois de Février. Notons que durant ces deux périodes de forte occurrence, le signal

correspondant au neurone 10 est le plus fréquemment détecté parmi les autres signaux, avec des fréquences d'occurrence entre 0,7 et 0,75 du total des pixels affectés aux diatomées. Ce signal pourrait donc correspondre à une diatomée dominante dans l'Océan Austral à cette période-là.

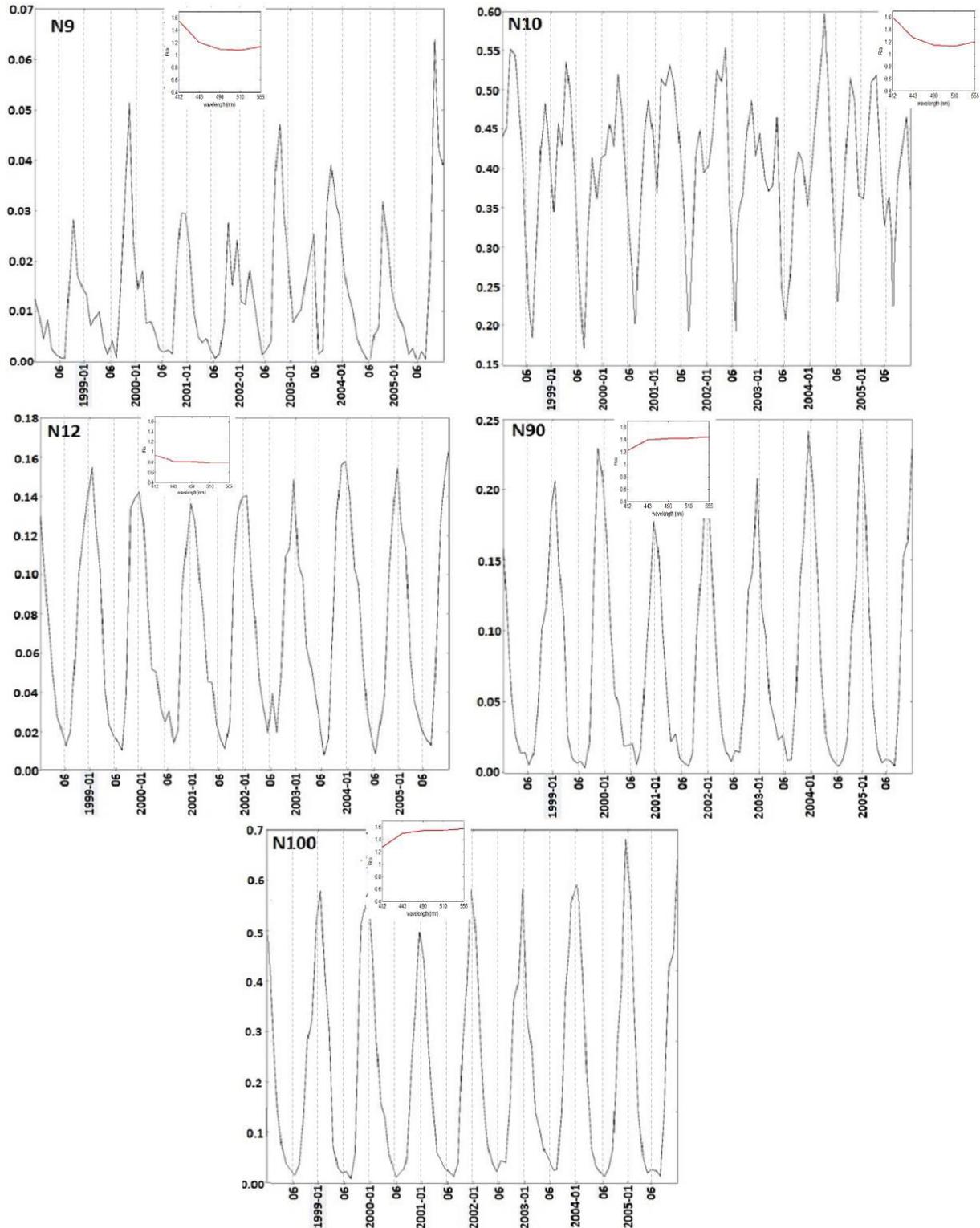


Figure 5.7. Fréquences relatives d'apparition des 5 neurones associés aux diatomées dans l'océan austral (entre 40° et 70° Sud) entre 1998 et 2005 (le référent de chaque neurone est représenté à titre indicatif).

Dans la littérature, on trouve par exemple référence à la dominance de certaines espèces telles que *Chaetoceros atlanticus* ou *Chaetoceros dichchaeta* en début d'été austral (Novembre-Décembre) puis à leur remplacement progressif par d'autres espèces telles que *Fragilariopsis kerguelensis* et *Corethron inerme* durant les mois de Janvier et Février. Ces successions se produisent essentiellement dans les zones où le fer ne représente pas un facteur limitant (Assmy et al., 2005) et peuvent être le résultat de l'interaction entre plusieurs mécanismes, parmi lesquels :

- Le broutage par le méso-zooplancton (500 μm à 5 mm), tels que les copépodes (Assmy et al., 2005, Kruse et al., 2009).
- Le contrôle par les nutriments qui peuvent progressivement représenter un facteur limitant pour certaines espèces ainsi que l'établissement progressif d'une limitation par l'acide silicique qui peut affecter différemment la croissance de certaines espèces (Armand et al., 2008).
- La senescence des cellules. Celle-ci peut se produire d'une manière naturelle (mort programmée ou apoptose) ou reliée à une épidémie (infection bactérienne ou virale par exemple) (Assmy et al., 2005). L'apoptose et le broutage par de petits copépodes ont été reconnus comme étant les principaux facteurs pouvant expliquer la senescence des populations de *Pseudo-nitzschia* dans l'Atlantique Sud en fin de bloom (Smetacek et al., 2002).
- Le ratio croissance / rythme des mortalités (Assmy et al., 2005, Kruse et al., 2009) et la capacité de l'espèce à s'adapter à son environnement écologique (limitations en nutriments, prédation, etc.).

Le neurone 12 présente un cycle régulier, caractérisé par un seul pic de fréquence (environ 0,12), atteint entre les mois de Janvier et Février. Il en est de même pour les neurones 90 et 100 (neurones voisins au niveau de la carte auto-organisatrice), qui présentent un cycle similaire, caractérisé par un seul pic de fréquence vers le mois de Janvier. Il est donc probable que ces signaux (plus réguliers) correspondent à des diatomées, dont le cycle de vie est moins influencé par l'interaction entre plusieurs facteurs complexes (Smetacek et al., 2002). Ce ne sont bien sûr ici que des hypothèses et cette analyse montre la complexité des facteurs pouvant entrer en jeu dans la détermination des formes et amplitudes des anomalies spectrales, facteurs qui nécessitent des études bio-optiques et des expertises de terrain plus approfondies afin de pouvoir comprendre leurs mécanismes et de mieux les appréhender.

5.3. Amélioration de la détection des blooms de diatomées dans des conditions spécifiques avec PHYSAT-SOM

Dans le précédent chapitre, les cartes climatologiques globales ont montré des zones où les blooms de diatomées étaient différemment détectés par PHYSAT-SOM et PHYSAT. Nous illustrons et analysons ici ces principales différences, en relation avec la littérature existante. Nous nous focaliserons ainsi sur certaines régions caractérisées par des blooms intensifs de diatomées, tels que les upwellings côtiers du Pérou (Pacifique Equatorial) et du Benguela (Atlantique Sud) de même que la zone du Pacifique Equatorial après l'épisode d'El Niño de 1997 et le phénomène La Niña de 1998 qui lui a succédé.

5.3.1. Amélioration de la détection des blooms de diatomées dans les upwellings côtiers

Dans les zones riches en nutriments, telles que les upwellings côtiers, des blooms de diatomées sont nettement visibles sur les cartes générées par PHYSAT-SOM par rapport à celles générées par la version 2008 de PHYSAT, où d'autres groupes semblent être plus dominants au niveau de ces zones (figure 5.8). L'intensité des floraisons de diatomées est particulièrement importante au niveau de l'upwelling du Benguela (e.g. Barlow, 1984 ; Kiørboe et al., 1998; Romero et al., 2002), dans l'Atlantique Sud (près des côtes angolaises, namibiennes et sud-africaines) ainsi qu'autour des côtes équatoriennes et péruviennes dans le Pacifique Equatorial, particulièrement durant l'été austral, tel que montré dans la figure 5.8.

Ces observations sont en accord avec de précédentes études et campagnes océanographiques. Par exemple, au niveau de l'upwelling du Pérou, Jackson et al. (2011) ont constaté d'intenses floraisons de diatomées dans une région située à environ 15° Sud, au niveau du large plateau continental péruvien. L'intensité de ces blooms a même été constatée durant le phénomène d'El Niño. Par ailleurs, DiTullio et al. (2005) ont recensé des floraisons massives de diatomées, atteignant 80 à 90% de la population phytoplanctonique, avec des concentrations en chlorophylle *a* dépassant les 45 mg/m³. Ces blooms de diatomées ont surtout été relevés au niveau de l'upwelling du plateau continental péruvien, où la remontée d'eaux riches en fer et autres nutriments engendre le développement de larges cellules de diatomées, tel que ça a été décrit dans la précédente partie de ce chapitre.

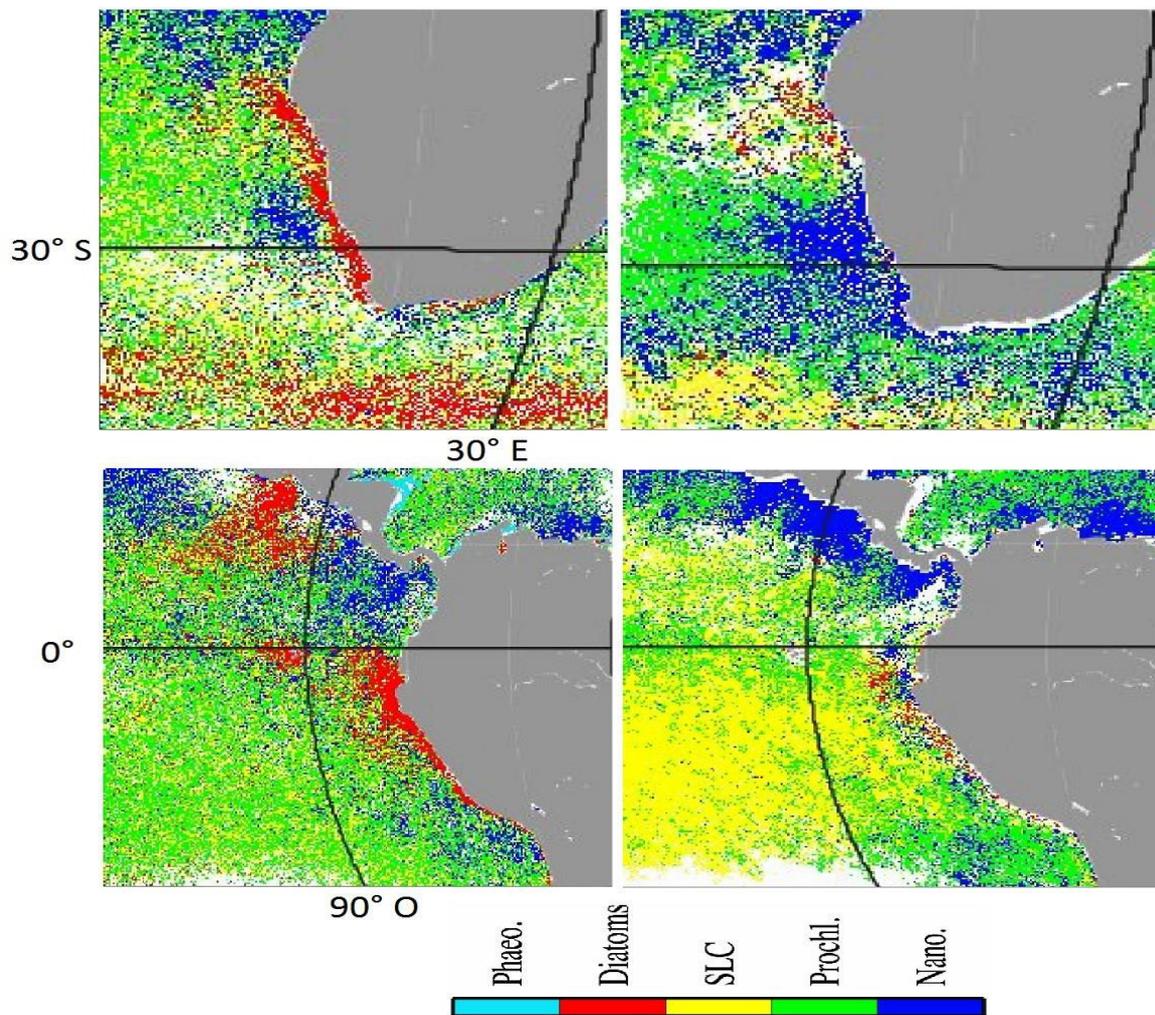


Figure 5.8. Cartes générées par PHYSAT-SOM (à gauche) et par la version 2008 de PHYSAT (à droite), représentant la distribution des groupes de phytoplancton dans l'upwelling du Benguela (en haut) et l'upwelling du Pérou (en bas) à partir d'une synthèse du mois de Janvier (climatologie sur la période d'observation satellite s'étendant de 1998 à 2010). A noter que la dominance des diatomées est nettement plus marquée sur les cartes de PHYSAT-SOM par rapport à celles de PHYSAT version 2008, où d'autres groupes semblent être plus dominants.

Par ailleurs, à travers une campagne océanographique menée dans l'upwelling du Pérou durant les mois d'Aout et Septembre 2000, Bruland et al. (2005) ont trouvé que les eaux extrêmement riches en chlorophylle, appelée « eaux brunâtres du Pérou » (avec des concentrations en chlorophylle situées entre 20 et 45 mg/m³), résultent de floraisons massives de diatomées au large du plateau continental. D'autres part, des floraisons de diatomées ont été recensées le long du courant de Humboldt, tout au long des côtes ouest d'Amérique du Sud, de l'extrémité sud du Chili jusqu'au nord du Pérou. Enfin, des expériences d'incubation établies par Hutchins et al. (2002) ont montré que, dans des conditions non limitées en fer, ces eaux contiennent des populations denses de diatomées pennées.

Il apparaît donc que les observations des blooms de diatomées obtenues avec PHYSAT-SOM sont en accord avec la littérature existante dans les régions étudiées. Tel qu'expliqué dans le précédent chapitre, l'amélioration de la détection de ces blooms est associée, d'une part, à la meilleure couverture de la diversité des anomalies spectrales avec SOM (en forme et amplitude) et à l'utilisation d'une base de données élargie, permettant de caractériser d'avantage de situations de dominance par rapport à l'ancienne approche de PHYSAT.

5.3.2. Blooms de diatomées succédant à un évènement climatique

La région du Pacifique Equatorial (de 5° Nord à 5° Sud et de 80° Ouest à 170° Est) est particulièrement intéressante du point de vue de la succession des groupes de phytoplancton en raison de sa sensibilité aux variations climatiques telles que « El Niño » et « La Niña » qui lui succède (Aceituno, 1988 ; Allan et al., 1996; Enfield, 1989). Le terme El Niño fait référence à un courant marin anormalement chaud qui remonte le long des côtes du Pérou. Le phénomène se distingue particulièrement bien sur les cartes satellite de température de surface des eaux (SST ou Sea Surface Temperature), qui affiche une élévation anormale durant cette période. Ce courant apparaît périodiquement, en réaction à un changement d'orientation des vents dans l'Ouest du bassin Pacifique. C'est un évènement majeur à l'échelle de la région directement touchée, mais également à l'échelle globale. Les années El Niño sont suivies du phénomène de La Niña, qui correspond notamment à un retour à une situation classique, associé à un upwelling équatorial plus intense que la normale accompagné de fortes remontées de sels nutritifs.

La variabilité des groupes est assez forte dans cette zone avec des évènements particuliers qui font, très certainement, suite à l'année El Niño de 1997. De façon générale, cette région est dominée par les *Prochlorococcus* et les SLC (zone oligotrophe) avec de temps à autre et localement, des floraisons de diatomées ou de nanoeucaryotes, tel que ça a précédemment été montré dans le chapitre 4 sur les cartes de climatologie, de même que dans d'autres études (Alvain et al., 2008 ; Gorgues et al., 2010 ; Masotti et al., 2011). Durant le phénomène El Niño de 1997, la concentration en chlorophylle *a* a diminué et des changements importants et à grande échelle ont affecté la composition de la communauté phytoplanctonique dans cette région. Ce changement était caractérisé par la diminution de la dominance des cyanobactéries picoplanctoniques (*Prochlorococcus* et *Synechococcus*), en

faveur des nanoeucaryotes (Masotti, 2011 ; Gorgues, 2010). La dominance relative des diatomées a, par la suite, fortement augmenté durant la période de La Niña en 1998. Il est généralement admis que les diatomées ne contribuent pas à plus de 20% de la biomasse phytoplanctonique totale dans cette région pauvre en nutriments (Everitt et al., 1990 ; Blanchot et al., 2001; Kobayachi & Takahashi, 2002; Mackey et al. 2002 ; Dandonneau et al., 2004). Elles sont donc généralement absentes de cette région, sauf exceptions et en dehors de quelques apparitions en bord de l'upwelling, le long des côtes du Pérou. Cependant, une floraison inhabituelle de diatomées a été observée de Juin à Décembre 1998 au niveau de l'équateur. La localisation de cette floraison de 1998 est entre 4° Nord et 4° Sud et de 80° Ouest à 140° Ouest. Les années 2003 et 2004 sont également associées à de légères floraisons, moins intenses mais de plus longue durée, localisés plus près des côtes en 2004 et entre 90° et 120° Ouest en 2003. Le bloom de 1998 a atteint son pic durant le mois d'Aout, tel que le montre la figure 5.9, dont la partie de droite affiche le gradient latitudinal de la fréquence relative des diatomées calculées à partir de moyennes mensuelles dans le Pacifique Equatorial entre 1998 et 2008 (occurrences de tous les neurones « diatomées » divisées par les occurrences de neurones caractérisant les autres groupes)

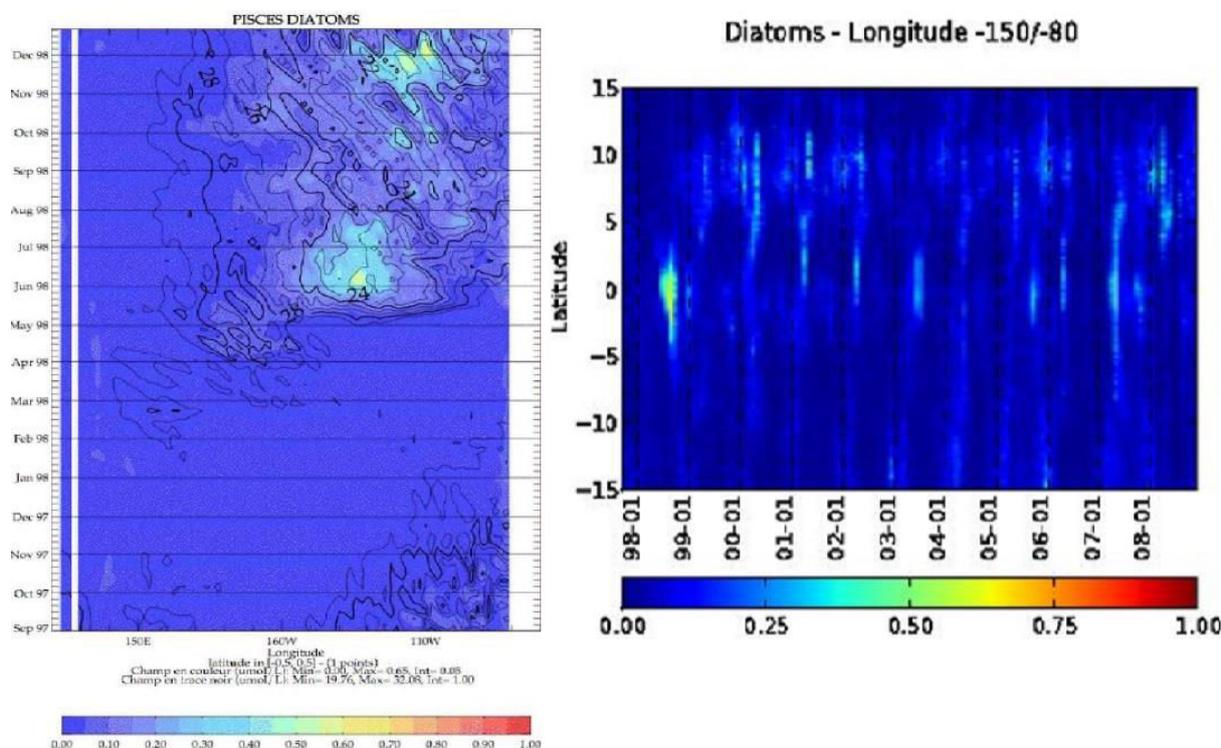


Figure 5.9. Gradient latitudinal de la fréquence relative des diatomées (tous neurones « diatomées » confondus) dans le Pacifique Equatorial entre 1998 et 2008 calculée par PHYSAT-SOM, montrant le bloom de diatomées inhabituel associé à l'évènement intense de La Niña d'Aout 1998 (à droite) et gradient longitudinal de la fraction relative de chlorophylle a associée aux diatomées dans la même zone d'après le modèle PISCES (à gauche, figure extraite du manuscrit de thèse de Séverine Alvain)

Des études *in situ* datant du précédent El Niño ou directement de 1997-1998, ont également observé cette floraison inhabituelle de diatomées faisant suite à l'intensification de l'upwelling équatorial, non limité aux régions côtières (Chavez et al., 1999 ; Strutton & Chavez, 2000; Ryan et al., 2002 ; Kobayashi et Takahashi., 2002). Ces observations sont également en accord avec les résultats de modèles biogéochimiques, dont le modèle PISCES (Aumont et al., 2003 ; Buitenhuis et al., 2006), tel que le montre la figure 5.9, dont la partie de gauche présente la fraction relative de la chlorophylle *a* associée aux diatomées selon le modèle PISCES dans le Pacifique Equatorial (Gorgues et al., 2010).

Les fréquences relatives des pixels associés aux diatomées dans le Pacifique Equatorial, calculées pour le mois d'Aout 1998 (synthèse mensuelle) avec PHYSAT-SOM, ont été comparées à celles obtenues avec la version 2008 de PHYSAT. Tel que montré par la figure 5.10, cette floraison est plus clairement visible sur la carte générée par PHYSAT-SOM avec une densité de pixels plus élevée, associée à la meilleure couverture spatiale obtenue avec SOM (chapitre 4). Des structures spatiales caractérisées par de larges méandres sont clairement visibles sur la carte de fréquences de PHYSAT-SOM. Ces structures pourraient être dues à une interaction entre les courants marins Nord et Sud sous l'effet des vents dominants dans cette zone. L'influence des ondes de Rossby (Rossby, 1939 ; Platzman, 1968 ; Dickinson, 1978) représente aussi une hypothèse probable expliquant ce phénomène de méandres. Les ondes océaniques de Rossby représentent des ondes de très faible fréquence au niveau de la surface des océans et de la thermocline. Elles sont causées par la variation du climat à travers l'interaction des vents et de la flottabilité et vice-versa. Les ondes barotropes et baroclines causent une variation de la hauteur de la surface de la mer. Les ondes baroclines donnent également un déplacement vertical significatif de la thermocline, souvent de l'ordre de plusieurs dizaines de mètres. Les satellites ont montré la progression des ondes de Rossby à travers les bassins océaniques, particulièrement aux latitudes moyennes (Beucher, 2010). Ces ondes ont une périodicité très longue pouvant prendre des mois ou même des années à traverser le Pacifique d'Est en Ouest en faisant déplacer les masses d'eaux sous la forme de larges méandres (Dickinson, 1978, Dandonneau et al., 2003). Elles jouent aussi un rôle important durant le phénomène El Niño en contribuant à la dispersion des masses d'eaux chaudes du Pacifique Est vers le Pacifique Ouest et participant ainsi à l'équilibre thermique des océans (Dickinson, 1978). Par ailleurs, une étude a montré que ces ondes pourraient accumuler des particules flottantes dans les zones de convergence au niveau des gyres subtropicaux causant des anomalies observables par satellite (Dandonneau et al., 2003). De

tels processus peuvent être déterminants dans la distribution des organismes vivants dans les eaux oligotrophes.

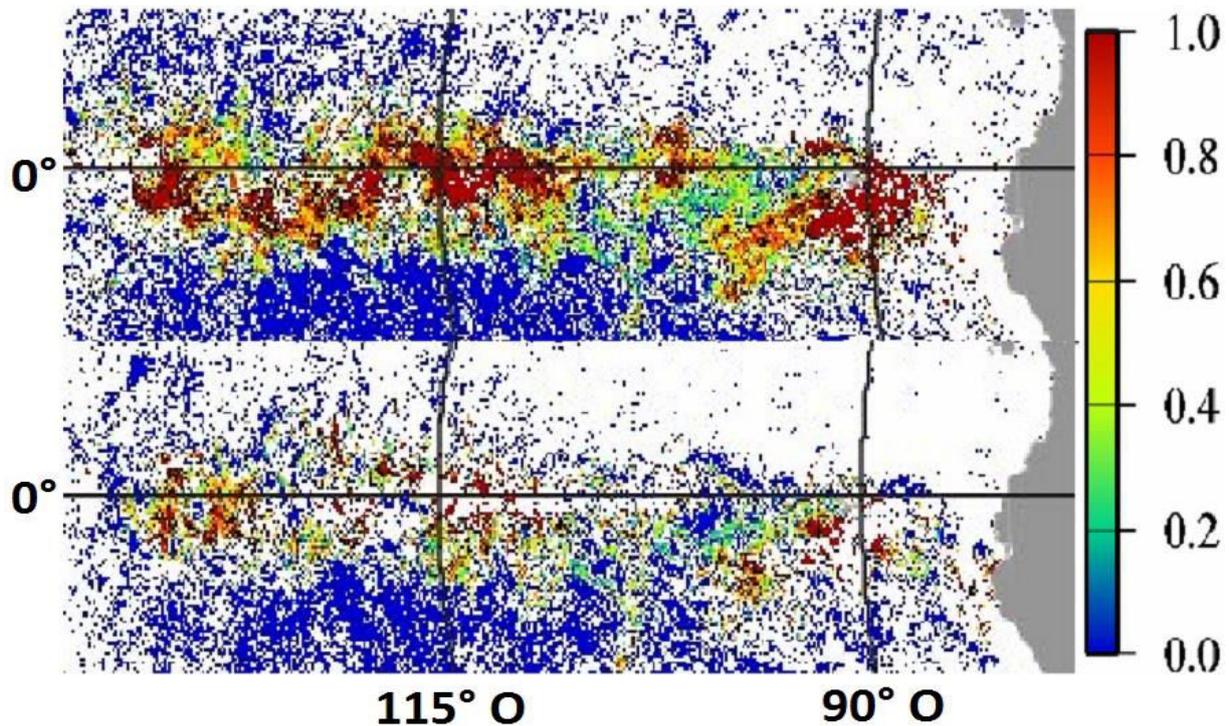


Figure 5.10. Fréquences relatives des pixels associés aux diatomées (tous neurones « diatomées » confondus) dans le Pacifique Equatorial (Au large des côtes du Pérou) durant l'évènement de La Niña d'Aout 1998 calculées par PHYSAT-SOM (en haut) et la version 2008 de PHYSAT (en bas) (moyenne mensuelle sur le mois d'Aout 1998)

Au-delà de l'amélioration de la distinction des structures spatiales (figure 5.10), la caractérisation de différents signaux associés aux diatomées permet également d'évaluer la contribution relative des différentes signatures spectrales de Ra durant ce bloom inhabituel. La figure 5.11 montre les fréquences relatives des différents neurones caractérisant les diatomées (calculées relativement aux neurones « diatomées » uniquement, de la même façon que pour les figures 5.2 et 5.7) pendant la floraison de La Niña en Aout 1998 (moyenne mensuelle sur le mois d'Aout 1998). On remarque que les différentes signatures spectrales ne sont pas présentes dans les mêmes proportions. Le spectre de Ra correspondant au neurone 10 est le plus fréquemment présent au moment de ce bloom. Le spectre du neurone 9, bien que très proche de celui du 10 (voir figure 5.2), est moins fréquemment observé. Il en est de même pour les neurones 90 et 100 qui présentent des spectres similaires en termes de forme et amplitude (voir figure 5.2) mais pas en fréquence d'apparition. La dissociation avec SOM du signal relatif à un bloom d'un même groupe de phytoplancton donne ainsi l'avantage d'observer en détails les différents spectres de Ra le caractérisant et de prospector les

mécanismes complexes, aussi bien bio-optiques que spatio-temporels qui sont susceptibles d'agir sur ces signatures spectrales.

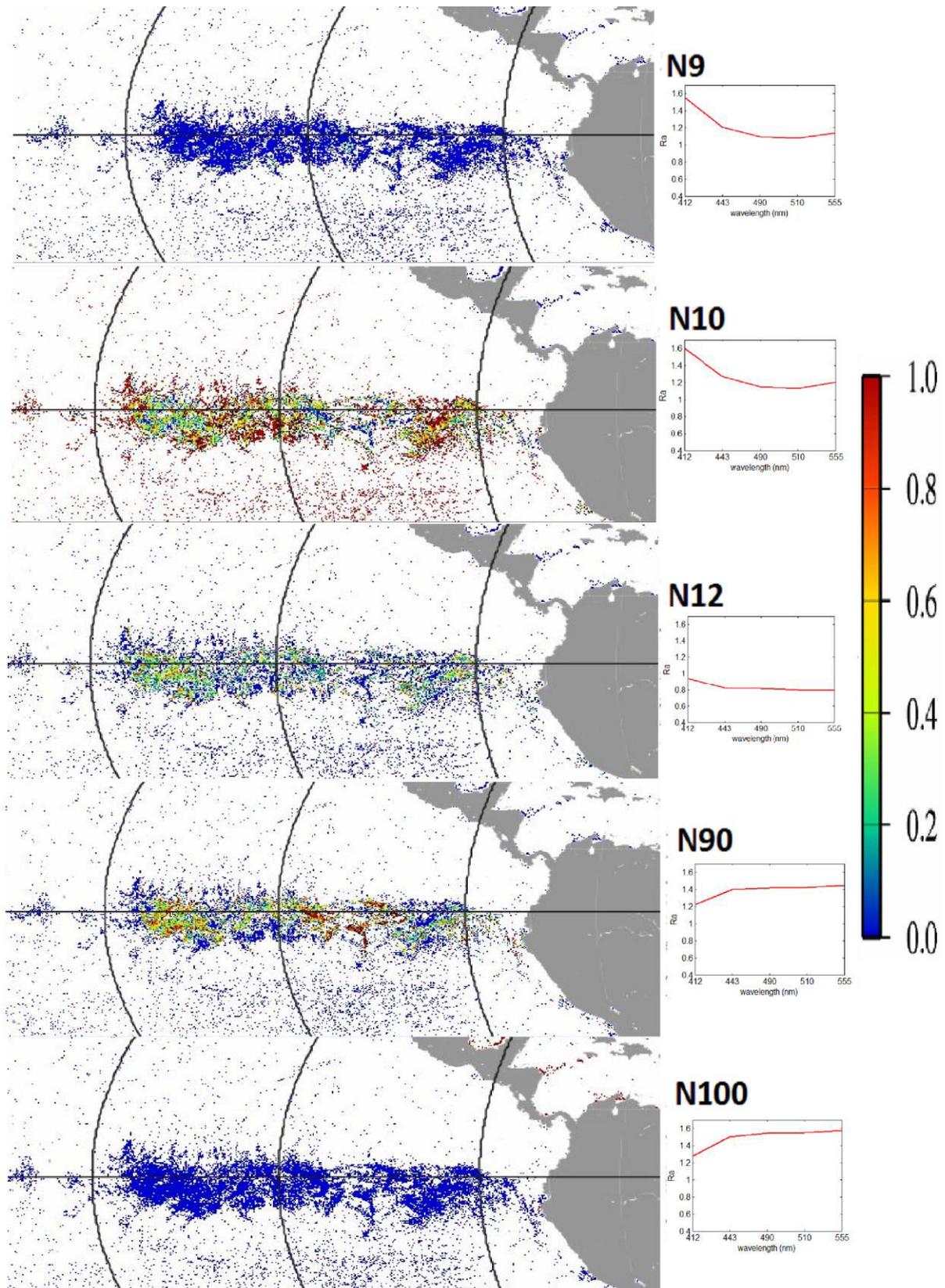


Figure 5.11. *Fréquence relative d'apparition des différentes signatures spectrales caractérisant les diatomées au moment du bloom inhabituel de La Niña en Août 1998 dans le Pacifique Equatorial*

5.4. Conclusion

La base de données *in situ* utilisée a permis d'associer les diatomées à des spectres référents significativement différents en termes de forme et d'amplitude. L'étude de la variabilité spatio-temporelle de ces spectres au niveau de certaines régions telles que le Pacifique Equatorial et l'Océan Austral a montré qu'ils n'occupaient pas la même aire géographique et qu'ils n'étaient pas caractérisés par la même saisonnalité. L'analyse des valeurs des propriétés optiques inhérentes correspondant aux localisations de chaque spectre a montré que les valeurs de ces paramètres étaient significativement différentes entre les spectres sur toutes les longueurs d'onde et que ceci peut être utilisé pour soutenir des hypothèses liées à la présence de différents genres ou stades de floraisons.

La littérature montre l'existence d'une grande variété de types de diatomées dans les régions étudiées. Leur taille est proportionnelle à la concentration en nutriment (fer en particulier) et la température de surface de l'eau, entre autres. Des études ont même montré une forte hétérogénéité de tailles intra-spécifique (exemple de l'espèce dominante dans l'Océan Austral, *Fragilariopsis kerguelensis*). Le fait que les spectres de Ra soient associés à différentes tailles d'une même espèce représente donc une des hypothèses susceptibles d'expliquer les différences des signatures spectrales. Cependant, la multitude de facteurs intervenant dans la détermination des formes et amplitudes des anomalies spectrales nécessitent plus de données *in situ* coïncidentes aux mesures spatiales ainsi des études optiques plus poussées pour déterminer l'origine précise de la différence entre les Ra observés pour les diatomées. Néanmoins, les nouveautés apportées par PHYSAT-SOM ouvrent plusieurs perspectives quant à la détection des groupes de phytoplancton dans des conditions autres que la dominance (stades de blooms, diverses conditions bio-optiques, espèces).

L'observation des blooms de diatomées dans les upwellings côtiers, tels que ceux du Pérou et du Benguela montre que ces efflorescences sont nettement plus visibles sur les cartes de PHYSAT-SOM par rapport à la version 2008 de PHYSAT. La détection de ces blooms est en accord avec de précédentes études ayant recensé les diatomées comme étant le principal groupe dominant dans ces régions riches en nutriments, quand la concentration en fer est suffisante pour leur croissance. L'augmentation de la couverture spatiale avec SOM (paragraphe 4.6 du chapitre 4) engendre une observation plus nette et plus détaillée des

structures spatiales des blooms par rapport à la version 2008 de PHYSAT. Ceci a été montré à travers l'exemple du bloom de diatomées succédant à l'évènement climatique de La Niña en Aout 1998. Ces améliorations permettent ainsi d'envisager plusieurs études visant, par exemple, l'interprétation des structures spatiales observées en relation avec les conditions environnementales et les forçages physiques, tels que l'hydrodynamisme (e.g. ondes de Rossby, tourbillons).

CHAPITRE 6

Application de PHYSAT-SOM dans les eaux du Cas 2 de la Manche et Mer du Nord : première approche et perspectives

6.1. Introduction

Dans les précédents chapitres, la classification automatique de spectres d'anomalies de PHYSAT avec les cartes auto-organisatrices, pour la détection des groupes de phytoplancton, a été développée pour les eaux dites du « Cas 1 ». Toutefois, l'application de la méthode PHYSAT-SOM ne peut être effectuée directement dans les eaux dites du « Cas 2 » car les spécificités de ces eaux ne permettent pas d'envisager une normalisation du signal des luminances marines en se basant uniquement sur celui de la chlorophylle *a*. En effet, ces zones sont caractérisées par des propriétés optiques influencées non seulement par le phytoplancton, mais aussi par la présence de matières terrigènes optiquement actives, telles que les matières organiques dissoutes (CDOM), les sédiments ou les particules minérales (Morel et Prieur, 1977 ; Smith et Baker, 1978 ; Gordon et Morel, 1983 ; Morel, 1988 ; Mobley et al., 2004), comme expliqué dans l'introduction. Dans ces régions, il est donc nécessaire de prendre en considération ces éléments qui influent sur le signal mesuré par les capteurs de la couleur de l'océan lorsqu'on souhaite l'exploiter pour la détection des groupes de phytoplancton.

Il existe quelques approches applicables dans les zones côtières visant à détecter des cas spécifiques de groupes de phytoplancton (Lubac et al., 2008 ; Astoreca et al., 2009 ; Pan et al., 2010, 2011). L'apport d'une méthode telle que PHYSAT-SOM permettrait de compléter fortement les possibilités d'observation des groupes phytoplanctoniques dans ces régions qui représentent souvent une importance particulière, de par les activités économiques qui s'y développent (e.g. tourisme, pêche) et leur sensibilité face aux pressions anthropiques et climatiques. Dans ce contexte, l'identification et le suivi des floraisons des différents groupes de phytoplancton constitue un enjeu important en océanographie, tant pour des raisons économiques, sanitaires, qu'écologiques (Anderson, 1989 ; Anderson et al., 2000, 2005 ; Sellner et al., 2003 ; Tomlison et al., 2004 ; Craig et al., 2006). Par exemple, les

floraisons d'algues nuisibles ou HAB (Harmful Algal Blooms) peuvent significativement affecter les activités touristiques. Par ailleurs, le compartiment phytoplanctonique, se trouvant à la base de la chaîne alimentaire marine, conditionne la présence d'espèces comestibles et leur pêche (Landsberg, 2002 ; Sellner et al., 2003).

Dans ce chapitre, on abordera une des premières étapes à effectuer pour envisager, à terme, l'application de PHYSAT-SOM dans les eaux du Cas 2. On se concentre donc sur l'analyse des Ra et la prise en compte de différentes classes d'eau, de façon à séparer les signatures optiques potentiellement liées à la présence de matière d'origine non phytoplanctonique du reste du signal. Ce travail a été effectué dans une région atelier du laboratoire d'Océanologie et de Géosciences, dont une partie est représentative des eaux du Cas 2 : la Manche et la Mer du Nord.

6.2. La Manche et la Mer du Nord

6.2.1. Contexte hydrographique

La Manche et la Mer du Nord (figure 6.1) représentent une région soumise à un important hydrodynamisme lié à de forts courants de marée. De plus, les pressions anthropiques (e.g. pêche, transport maritime) sont croissantes (Kostecki et Le Pape, 2011), induisant la mise sous surveillance de cette zone, notamment dans le cadre de la convention OSPAR (Oslo-Paris), les objectifs de la DCE (Directive Cadre sur l'Eau), de même que la DCSMM (Directive Cadre Stratégie pour le Milieu Marin).



Figure 6.1. Courants marins de la Manche et de la Mer du Nord et influence de l'océan Atlantique
(<http://commons.wikimedia.org/wiki/File:NorthSeaCurrents.jpg>)

La forme d'entonnoir du détroit du Pas-de-Calais crée un goulot d'étranglement qui génère une accélération des masses d'eau provenant de l'Océan Atlantique vers la Mer du Nord. La proximité des continents et la faible bathymétrie (au maximum 200m) renforce l'hydrodynamisme *via* les frottements (Salomon et Breton, 1993a, 1993b; Bailly Du Bois et Dumas, 2005). L'entrée d'eau au Nord est un courant chaud, provenant du gyre océanique Nord-Atlantique. Au sud, ces entrées sont moindres et diluées par les apports terrigènes importants, se chargeant progressivement en éléments dissous et matières organiques (Lacroix et al., 2007).

En Manche orientale, ce contexte marégraphique favorise la création d'une masse d'eau côtière individualisée, appelée « le fleuve côtier » (Brylinski et al., 1991). Plusieurs fleuves et cours d'eau du bassin hydrographique allant de la Baie de Seine jusqu'au détroit du Pas-de-Calais alimentent et entretiennent par leurs apports la masse d'eau côtière. Ces apports sont maintenus en zone côtière par un front qui freine leur dispersion vers les eaux du large (Brylinski et al., 1991). La circulation de ce fleuve côtier est régie principalement par la dérive résiduelle vers la Mer du Nord. Cependant, elle est soumise également au forçage du vent préférentiellement de secteur sud-ouest (Salomon et Breton, 1991).

6.2.2. L'étude du phytoplancton en Manche orientale et sud de la Mer du Nord

Cet écosystème, soumis à un fort régime marégraphique, est un lieu important d'interface avec le continent par le biais d'effluents. Breton (2000) indique que les assemblages phytoplanctoniques observés en Manche orientale sont essentiellement constitués de diatomées de grande taille représentant plus de 85% de la biomasse totale. Par ailleurs, Seuront et Vincent (2008) ont identifié la succession de trois assemblages de diatomées au cours du printemps 2004, correspondant à ceux observés en zone côtière de la Baie Sud de la Mer du Nord. Cependant de fortes concentrations correspondant à l'intense développement d'une Prymnésiophycée, *Phaeocystis*, sont recensées dans ces eaux (Hedin-Bougard, 1980 ; Brunet et al., 1996 ; Breton et al., 2000 ; Rousseau et al., 1994, 2002 ; Schapira et al., 2008). Gentilhomme et Lizon (1998) ont souligné le rôle important des sels nutritifs, plus particulièrement azotés, dans la prolifération printanière de cette micro-algue. Les blooms massifs de *Phaeocystis* sont des phénomènes récurrents dans les eaux côtières eutrophisées d'Europe (Lancelot et al., 1998 ; Schoemann et al., 2005). Certaines espèces de ce genre cosmopolite peuvent produire d'importantes quantités de colonies mucilagineuses au cours du printemps (figure 6.2). La prolifération massive de ces colonies est reconnue pour

avoir des effets négatifs sur les activités aquacoles et de pêche mais elle a également un impact très important sur le fonctionnement de l'écosystème (Weisse et al., 1994 ; Schoemann et al., 2005 ; Nejstgaard et al., 2007). Par conséquent, les blooms de colonies de *Phaeocystis* ont été classés parmi les blooms d'algues nuisibles (Anderson et al., 1998 ; Veldhuis et Wassmann, 2005).

Les contextes biologiques et hydrographiques, décrits ci-haut, impactent conjointement les propriétés optiques des eaux de cette région qui sont, par conséquent, influencées par le phytoplancton et par la présence de matières terrigènes organiques, minérales et détritiques. Ceci souligne la nécessité de développer une méthodologie adéquate afin d'être en mesure d'étudier les anomalies spectrales dans cette région.



Figure 6.2. Bloom de Phaeocystis produisant une mousse blanche (mucus algal) qui s'accumule sur les plages du littoral par la marée descendante (<http://www.seos-project.eu>)

6.3. Méthodologie suivie pour la classification des anomalies spectrales en Manche et Mer du Nord

6.3.1. Données satellites

Les données satellites utilisées dans l'exemple que nous présentons correspondent aux luminances marines normalisées (nLw) du capteur SeaWiFS, de niveau 2 et pour 5 longueurs d'onde du visible (412, 443, 490, 510 et 555 nm). Les nLw sont disponibles au laboratoire après utilisation de l'algorithme standard de correction atmosphérique (Gordon et Wang,

1994 ; Stumpf et al., 2003 ; Bailey et al., 2010). Afin de tenir compte des phénomènes potentiellement observables à une échelle régionale, telle que celle de la zone étudiée, la résolution spatiale utilisée est de 1 km (données MLAC à pleine résolution).

Un masque de nuages adapté aux eaux turbides a été appliqué (Nordkvist et al. (2009) à la place de celui utilisé par la NASA. En effet, ce dernier induit généralement une perte significative de données car il classe les pixels turbides en tant que nuages. Ceci est expliqué par le fait que cet algorithme standard utilise un seuil bas et fixe sur les réflectances dans le proche infra-rouge, au-dessus duquel le pixel est classé en tant que nuage. Ce seuil devient cependant invalide dans les eaux côtières turbides, où de fortes valeurs de réflectances peuvent être enregistrées dans le proche infra-rouge à cause des matières en suspension et les aérosols d'origine terrestre (Vantrepotte et al., 2012). Le masque de Nordkvist et al. (2009) est, quant à lui, basé sur la faible variabilité spectrale des nuages par rapport à celle de l'eau.

6.3.2. Classification préliminaire des eaux

Etant donné l'impact d'autres substances optiquement actives que le phytoplancton dans les eaux du Cas 2, il était nécessaire de passer par une étape permettant de prendre en compte cette spécificité lors de la normalisation des nLw. Pour cela, il a été choisi d'utiliser une classification des eaux côtières optiquement contrastées. Cette méthode, développée par Vantrepotte et al. (2012), permet de séparer les eaux dont les propriétés optiques sont majoritairement déterminées par le phytoplancton (typiquement les eaux du large), de celles qui sont impactées par la présence d'autres facteurs.

Cette méthode est basée sur une classification des spectres de réflectance (Rrs). Préliminairement à la phase de classification, ces spectres sont normalisés par leur intégrale (aire sous la courbe) dans le but de réduire la variabilité de 1^{er} ordre, représentée par l'amplitude du spectre, en faveur de sa forme. Ce critère qualitatif vise à mettre en évidence l'impact de la composition de l'eau en éliminant l'effet de certains facteurs qui influent sur les valeurs absolues des réflectances (principalement la transparence des eaux qui est dépendante de la concentration des constituants optiquement actifs (e.g. matières en suspension, CDOM). Les spectres de réflectances normalisés sont ensuite labellisés à l'aide de données *in situ*, comprenant divers paramètres biogéochimiques (e.g. concentration en chlorophylle *a*, matières en suspension, carbone organique particulaire, absorption par le phytoplancton et par

le CDOM, rétrodiffusion). La labellisation est faite en reliant les spectres de réflectances provenant des mesures satellites avec des classes de spectres de Rrs déterminées à partir des données *in situ*. Une méthode, appelée « Novelty Detection Technique » (N.D.T, Bishop, 1994), est utilisée pour trouver le spectre *in situ* le plus proche du spectre mesuré par le satellite. Cette technique se base sur la distance de Mahalanobis (D'Alimonte et al., 2003 ; Mélin et al., 2011 ; Vantrepotte et al., 2012). Elle permet de relier les spectres selon la distance qui les sépare et calcule au même temps une fonction de densité de probabilité qui sert à déterminer la probabilité d'appartenance d'un spectre à une classe d'eau donnée.

La méthode de Vantrepotte et al. (2012) permet, à ce jour, de distinguer les 5 classes d'eau suivantes :

- La classe 1 est caractérisée par la prédominance de matériel particulaire (principalement des sédiments) ainsi qu'une fraction relativement élevée de matériel détritique.
- La classe 2 est composée de matériel particulaire d'origine biologique (vivant et détritique) ainsi qu'une fraction importante de matières dissoutes (e.g. CDOM).
- La classe 3 est représentée par un mélange entre particules minérales et matières dissoutes.
- La classe 4 est caractérisée par un mélange de phytoplancton et de matières dissoutes.
- La classe 5 représente les eaux du large turbides, mais moins impactées par les apports terrigènes.

Les classes 1 et 3 ont été écartées lors du test présenté dans ce chapitre, n'étant pas, d'après cette classification, caractérisées par une présence significative de phytoplancton. Les différentes classes d'eau étant maintenant déterminées, l'étape de normalisation peut être réalisée pour chacune d'elles. De cette façon, les spécificités des signaux dus aux substances non phytoplanctoniques seront intégrées au signal moyen utilisé lors de la normalisation.

6.3.3. Calcul des anomalies spectrales relatives à chaque classe d'eau

Pour chaque classe d'eau, une LUT (Look-Up Table) de signaux moyens spécifiques a été utilisée afin de tenir compte des valeurs de chlorophylle *a* et de luminances marines normalisées relatives à chaque classe (tables disponibles en annexe 4). Ainsi, une référence ($nL_{wref}(\lambda, Chla)$) est calculée pour chaque type d'eau et est par la suite utilisée pour calculer les Ra suivant l'équation 2.1 (chapitre 2). Les spectres moyens de $nL_{wref}(\lambda, Chla)$ utilisés pour le calcul des Ra de chacune des 3 classes d'eau sont montrés sur la figure 6.3.

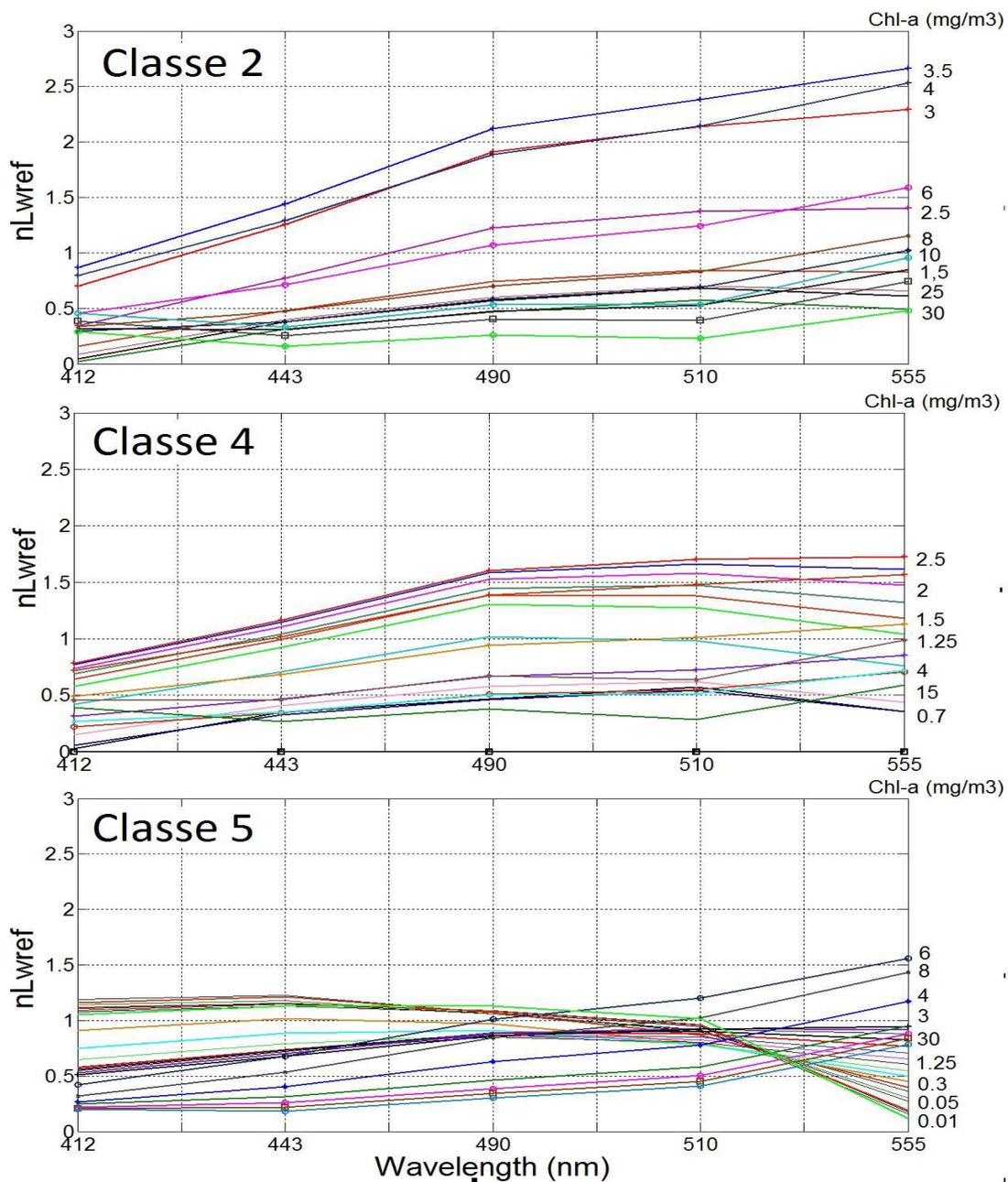


Figure 6.3. Spectres de luminances marines normalisées $nL_{wref}(\lambda, Chla)$ utilisés pour calculer les anomalies spectrales de chaque classe d'eau

Des différences significatives sont observables entre les 3 groupes de spectres de $nLw_{ref}(\lambda, Chla)$ en termes d'amplitude et de forme. Ceci est particulièrement le cas entre la classe 5 et les deux autres classes. De même, de plus faibles valeurs de $nLw_{ref}(\lambda, Chla)$ sont observables dans les grandes longueurs d'onde pour les classes 4 et 5 (entre 0,25 et 1,5 à 555 nm) comparativement à la classe 2 (entre 0,75 et 2,25 à 555 nm). Par ailleurs, les nLw_{ref} des 3 types d'eau se distinguent par des intervalles de concentration en chlorophylle *a* différents, tenant compte des propriétés bio-optiques spécifiques à chaque cas. Ainsi, la classe 2 se caractérise par des valeurs de *Chla* entre 1,5 et 30 mg/m³, la classe 4 par des valeurs entre 0,7 et 15 mg/m³ et enfin, la classe 5, représentant les eaux du large turbides, se caractérise par le plus large intervalle de *Chla*, situés entre 0,01 et 30 mg/m³. Par ailleurs, le nombre de spectres à partir desquels les LUTs ont été générées diffère significativement d'un type d'eau à un autre pour une concentration en chlorophylle *a* donnée (pour une *Chla* donnée, le nombre de spectres figure sur la dernière colonne de chaque LUT en annexe 4).

Les valeurs de $nLw_{ref}(\lambda, Chla)$ sont montrés sur la figure 6.4 pour les 3 classes d'eau et chacune des 5 longueurs d'onde, prises en compte séparément. On constate des différences significatives entre les 3 classes, particulièrement entre la classe 5 et les deux autres. Ainsi, pour la classe 2, le maximum de nLw_{ref} se trouve à 3,5 mg/m³ de *Chla* alors que pour la classe 4, le maximum de nLw_{ref} se situe à 2,5 mg/m³ de *Chla* (pour toutes les longueurs d'onde). La classe 5 se distingue par une gamme de *Chla* plus large (*Chla* varie entre 0,01 et 30 mg/m³) que les deux autres classes, de même que par une différence relativement plus marquée de $nLw_{ref}(\lambda, Chla)$ en fonction de la longueur d'onde. Par exemple, les valeurs maximales de nLw_{ref} se situent à 0,18 mg/m³ de *Chla* à 412 nm ($nLw_{ref}(412) = 1,19$) alors qu'à 555 nm, les valeurs maximales de nLw_{ref} se trouvent à 6 mg/m³ de *Chla* ($nLw_{ref}(555) = 1,55$). En revanche pour les classes 2 et 4, les valeurs maximales de nLw_{ref} se trouvent respectivement à 3,5 et 2,5 mg/m³ de *Chla*, indépendamment de la longueur d'onde considérée.

Les différences en termes de forme et d'amplitude entre les spectres de $nLw_{ref}(\lambda, Chla)$ (figure 6.3), de même que les différences des relations entre nLw_{ref} et la concentration en chlorophylle *a* pour une longueur d'onde donnée dans chacune des 3 classes d'eau (figure 6.4) soulignent l'importance de les considérer séparément pour le calcul des anomalies spectrales, via la classification de Vantrepotte et al. (2012). En appliquant pour chaque classe l'équation 2.1 du chapitre 2, un jeu d'anomalies spectrales spécifique à chaque type d'eau a ainsi été calculé. La prochaine étape consiste à appliquer séparément

l'algorithme des cartes auto-organisatrices pour classer automatiquement les spectres de Ra de chaque type d'eau.

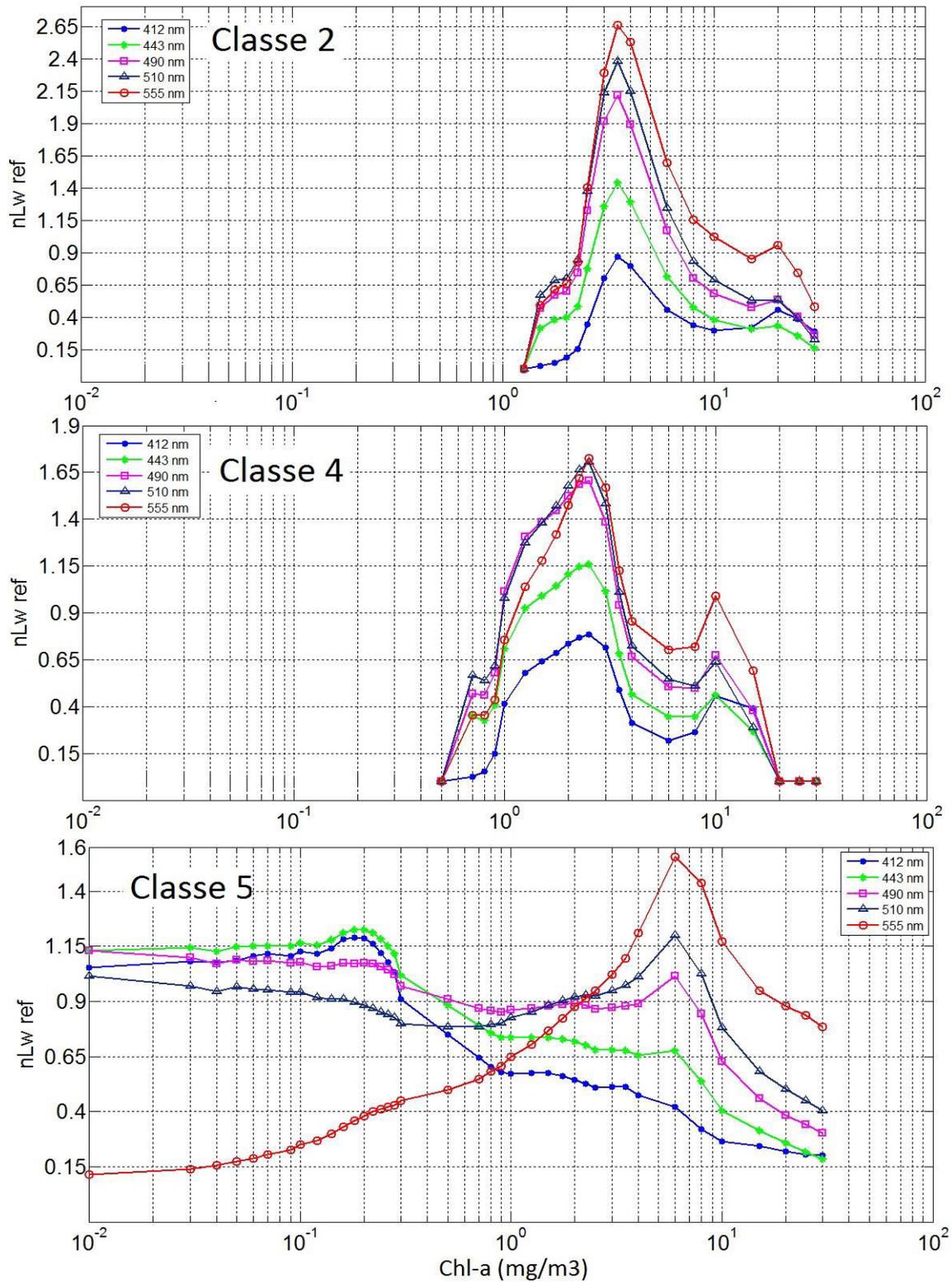


Figure 6.4. Valeurs de $nLW_{ref}(\lambda, Chla)$ pour les 3 classes d'eau (chaque longueur d'onde prise séparément)

6.3.4. Application des cartes auto-organisatrices séparément dans chaque classe d'eau

Pour chaque type d'eau, un apprentissage distinct a été fait avec une carte auto-organisatrice bidimensionnelle de 20x5 neurones. On a donc 3 cartes SOM distinctes de 100 neurones pour chaque classe d'eau. La forme et la taille de la carte ont été déterminées en se basant sur les paramètres d'erreurs (erreurs de quantification vectorielle, erreurs topologiques et distorsion moyenne de tous les neurones), décrits dans le chapitre 3. Dans l'exemple présenté dans ce chapitre, on vise à avoir une classification fine des diverses formes et amplitudes des spectres de Ra pouvant exister au sein de chaque type d'eau, indépendamment du nombre de groupes ou d'assemblage de groupes de phytoplancton susceptibles d'être détectés à l'avenir. Lorsque cette classification automatique sera, en perspective, appliquée à la détection des groupes phytoplanctoniques, on pourrait envisager une simplification de la classification en réduisant, par exemple, le nombre de neurones ou en regroupant les différents spectres référents.

Pour l'apprentissage de SOM l'ensemble de Ra utilisé comprend 477 460, 2 885 913 et 3 535 204 spectres respectivement pour les classes 2, 4 et 5. Cet ensemble a été obtenu à partir des données d'une année complète d'images SeaWiFS (2000). A noter que, pour la classe 5, un échantillonnage a été fait en sélectionnant 1 pixel sur 5 compte tenu du grand nombre de spectres associés. A la fin de l'apprentissage, les spectres de Ra de chaque type d'eau sont ainsi séparés en 100 classes. Chaque neurone est associé à un spectre référent (figure 6.5) qui résume l'information classée par ce neurone (ensemble d'anomalies spectrales). L'ordre topologique induit par l'algorithme des SOM est visible sur la figure 6.5. En effet, les spectres référents similaires se retrouvent dans une même zone de la carte auto-organisatrice grâce à la fonction de voisinage incluse dans l'algorithme.

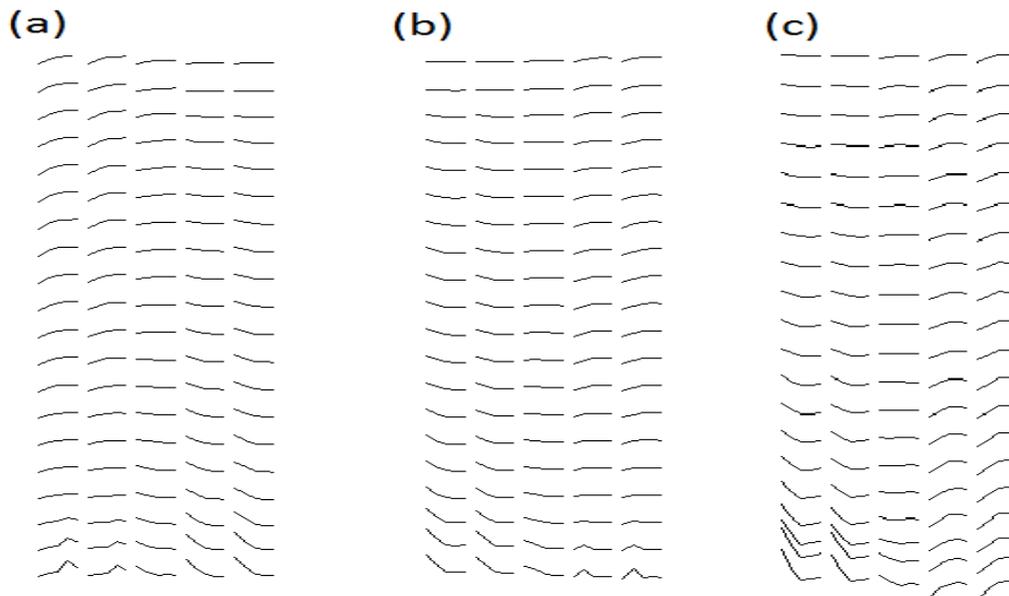


Figure 6.5. Les 100 spectres référents associés aux classes d'eau n°=2 (a), 4 (b) et 5 (c). Les référents représentent une synthèse de la variabilité des spectres de Ra relatifs à chacune des classes d'eau sur les 5 longueurs d'onde considérées

La figure 6.6 représente la projection des composantes des spectres référents correspondant à la classe d'eau numéro 2 (matériel particulaire d'origine biologique mélangé à des matières dissoutes) comparées entre chacune des 5 longueurs d'onde prises en compte deux à deux. Puisque ces spectres référents (figure 6.5) représentent une synthèse (ou une compression) de l'information contenue dans le jeu de données d'apprentissage, les relations entre les valeurs des spectres référents sont représentatives de la variabilité des anomalies spectrales dans le type d'eau considéré. Cette figure montre des relations semblables à celles trouvées à l'échelle globale entre les différentes longueurs d'onde considérées (figure 3.7 du chapitre 3). En effet, les plus grandes variations dans les valeurs des anomalies spectrales se produisent entre 412 nm et les plus grandes longueurs d'onde, où le signal devient relativement constant (généralement à partir de 490 nm). Le plus faible coefficient de corrélation est observé entre Ra (412nm) et Ra (510 nm) ($R^2 = 0,72$). On retrouve également le même type de relations pour les 2 autres classes d'eau (voir annexe 4). On en déduit que les valeurs d'anomalies spectrales présentent leur maximum de différences entre les courtes longueurs d'onde (412, 443 nm) et les plus grandes longueurs d'onde quel que soit le type d'eau et sa composition. Les formes de certains spectres (figure 6.5) et les valeurs d'anomalies spectrales vont toutefois différer d'une classe d'eau à une autre.

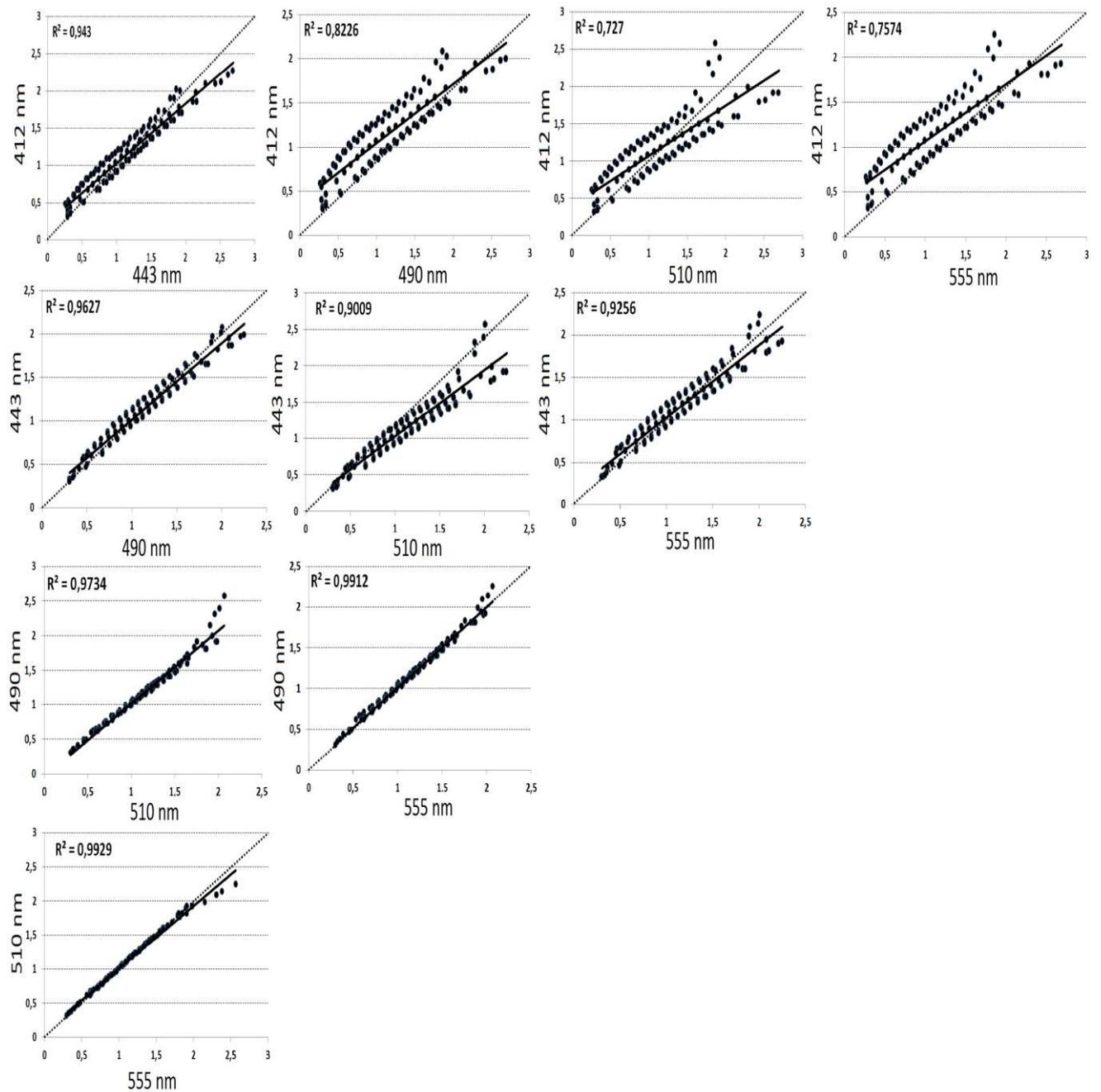


Figure 6.6. Projection des composantes des spectres référents correspondant à la classe d'eau numéro 2 (matériel particulaire d'origine biologique mélangé à des matières dissoutes) comparées entre chacune des 5 longueurs d'onde prises en compte deux à deux

Comme pour l'application des SOM à l'échelle globale, les spectres référents ainsi calculés pour chaque type d'eau vont être utilisés pour associer chaque pixel des images SeaWiFS de niveau 2 à 1 km de résolution à son neurone le plus proche (spectre référent le plus proche en termes de distance euclidienne) à l'aide de la fonction d'affectation de SOM.

Cette projection permet la production de cartes de distribution de neurones et donc, des Ra associés. La variabilité spatio-temporelle de ces spectres (100 spectres par classe d'eau) a ainsi pu être déduite à partir des occurrences et fréquences d'apparition de chaque neurone. Le diagramme montré dans la figure 6.7 résume la méthodologie suivie pour avoir la variabilité spatio-temporelle des Ra dans la région atelier, en partant des données satellites de luminances marines normalisées. Toutes ces étapes pourraient être reproduites dans d'autres régions en vue d'une labellisation des signaux trouvés en termes de groupes ou d'assemblages de groupes de phytoplancton, selon les informations *in situ* disponibles.

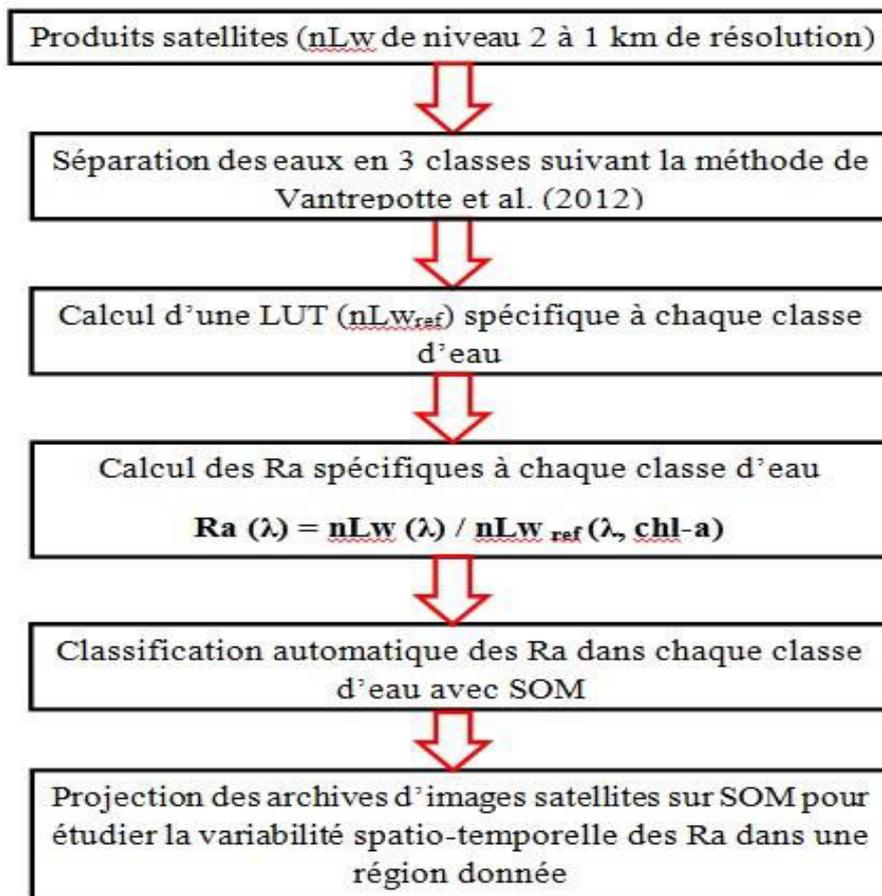


Figure 6.7. Méthodologie suivie pour avoir la variabilité spatio-temporelle des Ra en partant des données satellites de luminances marines normalisées.

6.4. Variabilité spatio-temporelle des anomalies spectrales en Manche et Mer du Nord : résultats préliminaires

Nous illustrons dans ce qui suit quelques cas de variabilités spatio-temporelles des Ra pour les 3 classes d'eau considérées. Ces résultats préliminaires pourront être utiles à l'avenir afin de trouver des relations potentielles entre les différentes structures spatiales et

temporelles des anomalies spectrales fréquemment observées et les données *in situ* relatives à la dominance ou la présence de groupes phytoplanctoniques dans la région étudiée. Il sera également utile de les mettre en relation avec des forçages physiques (e.g. température, courants, vents, lumière disponible). Un exemple de distribution spatiale des anomalies spectrales est illustré dans la figure 6.8. Chaque carte représente une synthèse pluriannuelle (climatologie) de la localisation des 100 neurones relatifs à la classe 5 (eaux turbides du large). Chaque pixel est représenté par le numéro du neurone ayant la plus forte occurrence à travers le mois entier pour la période d'observation de SeaWiFS allant de 1997 à 2005 (les données à haute résolution au-delà de 2005 n'étant plus librement mises à disposition par la NASA). Les mêmes nuances de couleur représentent des neurones proches ou voisins au niveau de la carte auto-organisatrice.

Des structures spatiales bien définies (groupements de même couleur) apparaissent sur les cartes, ce qui veut dire que les neurones adjacents se retrouvent, le plus souvent, spatialement regroupés. Puisque les neurones adjacents sont représentés par des Ra similaires en termes de forme et d'amplitude, on peut supposer que les similarités des propriétés biophysiques du milieu résultent en des similarités dans ses propriétés optiques. Cette hypothèse est cohérente avec les résultats trouvés théoriquement par Alvain et al. (2012), montrant que les Ra dépendent essentiellement des propriétés optiques inhérentes du milieu (absorption par le phytoplancton, rétrodiffusion par les particules et absorption par le CDOM), mais devra être validée à l'avenir en procédant à une étude comparative entre la localisation des Ra et les différents paramètres optiques correspondant. Toutefois, les structures spatiales observées représentent un résultat préliminaire encourageant, en perspective d'une labellisation de ces différents signaux en termes de groupes phytoplanctoniques.

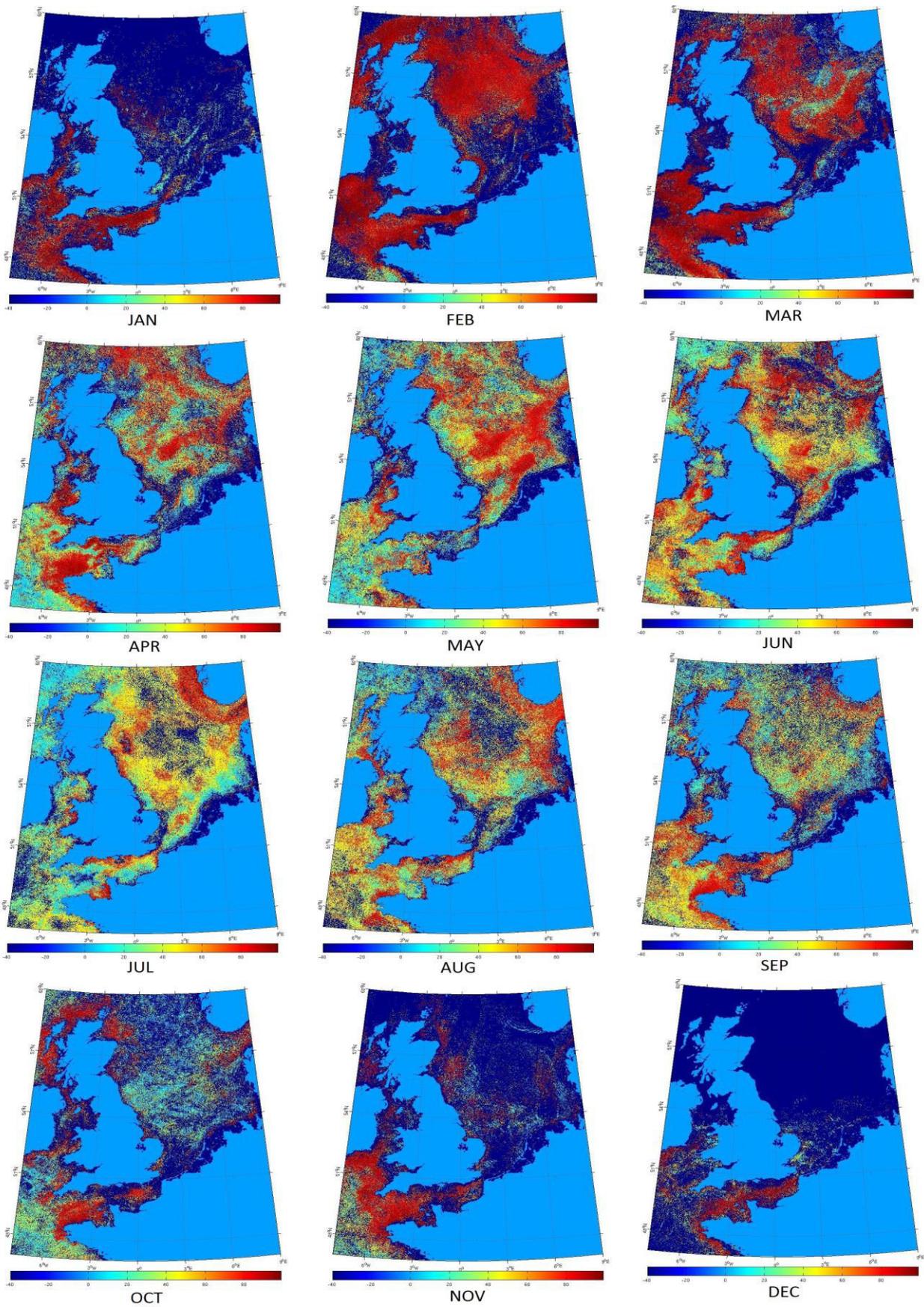


Figure 6.8. Synthèses mensuelles (climatologies sur la période allant de 1997 à 2005) de la distribution des 100 neurones en Manche et Mer du Nord pour la classe 5 (eaux turbides du large)

L'étude de la variabilité temporelle des anomalies spectrales automatiquement classées avec SOM constitue une étape importante afin d'avoir toutes les informations nécessaires pour une future étape de labellisation. Pour chacune des classes d'eau, les fractions relatives journalières des pixels associés à chaque neurone de SOM entre 1997 et 2005 ont été calculées. Un lissage par la méthode des moyennes mobiles a permis de mettre en évidence des cycles interannuels d'apparition de signatures spectrales spécifiques de Ra, plus ou moins marqués en fonction des neurones et du type d'eau. La figure 6.9 montre un exemple de variabilité temporelle associée à 2 neurones (44 et 86) appartenant à la classe 2 (prédominance de matériel particulaire d'origine biologique et de matières dissoutes).

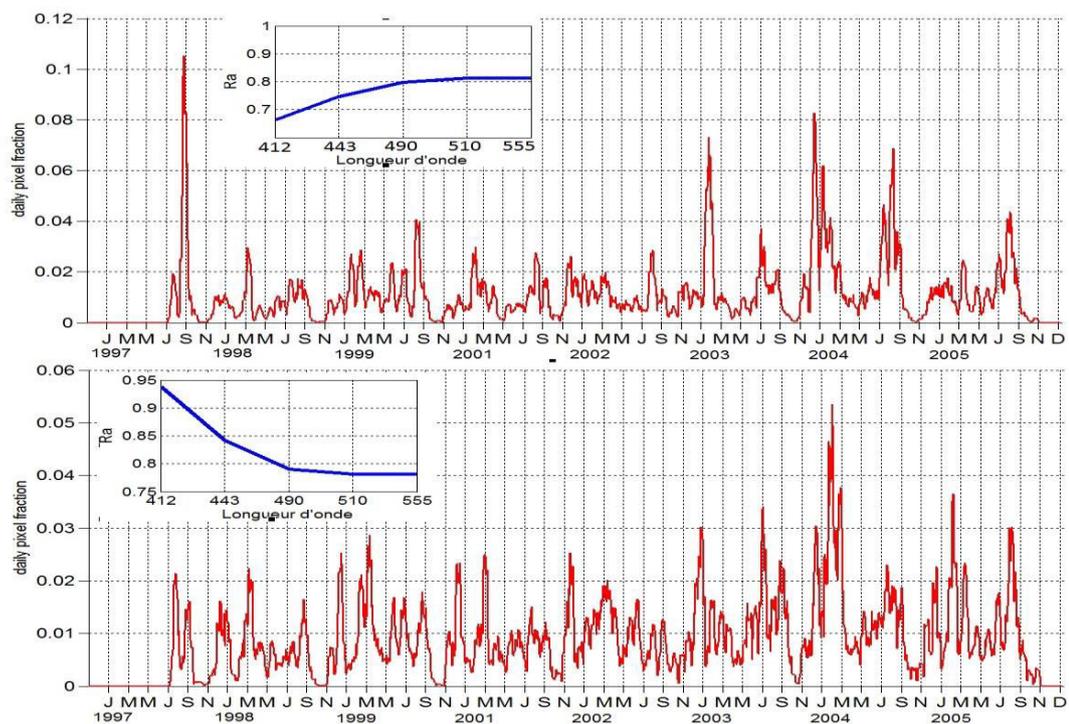


Figure 6.9. Fraction relative des pixels associés aux neurones 44 (en haut) et 86 (en bas) appartenant à la classe d'eau $n=2$ (mélange de matériel particulaire d'origine biologique et de matières dissoutes). Les spectres référents de chaque neurone sont aussi représentés.

On remarque des différences entre les fréquences d'apparitions de ces neurones (représentant des signatures spectrales de Ra différentes en forme et amplitude) tout au long de la période considérée. Par exemple, en Septembre 1997, le neurone 44 (en haut) est caractérisé par une fraction relative de pixels associés de 0,1 alors qu'à la même période, l'autre neurone affiche une fréquence relative de 0,02. On voit aussi que les pixels associés à ces neurones ne se caractérisent pas par des cycles interannuels d'apparitions bien définis. On note également que les spectres référents de ces deux neurones sont significativement différents en termes de formes bien qu'ils appartiennent à la même classe optique.

Un autre exemple est présenté dans la figure 6.10, montrant les fractions relatives des pixels associés à deux paires de neurones adjacents (neurones 68 et 69 ; 92 et 93) de la classe d'eau n=5 (eaux turbides du large).

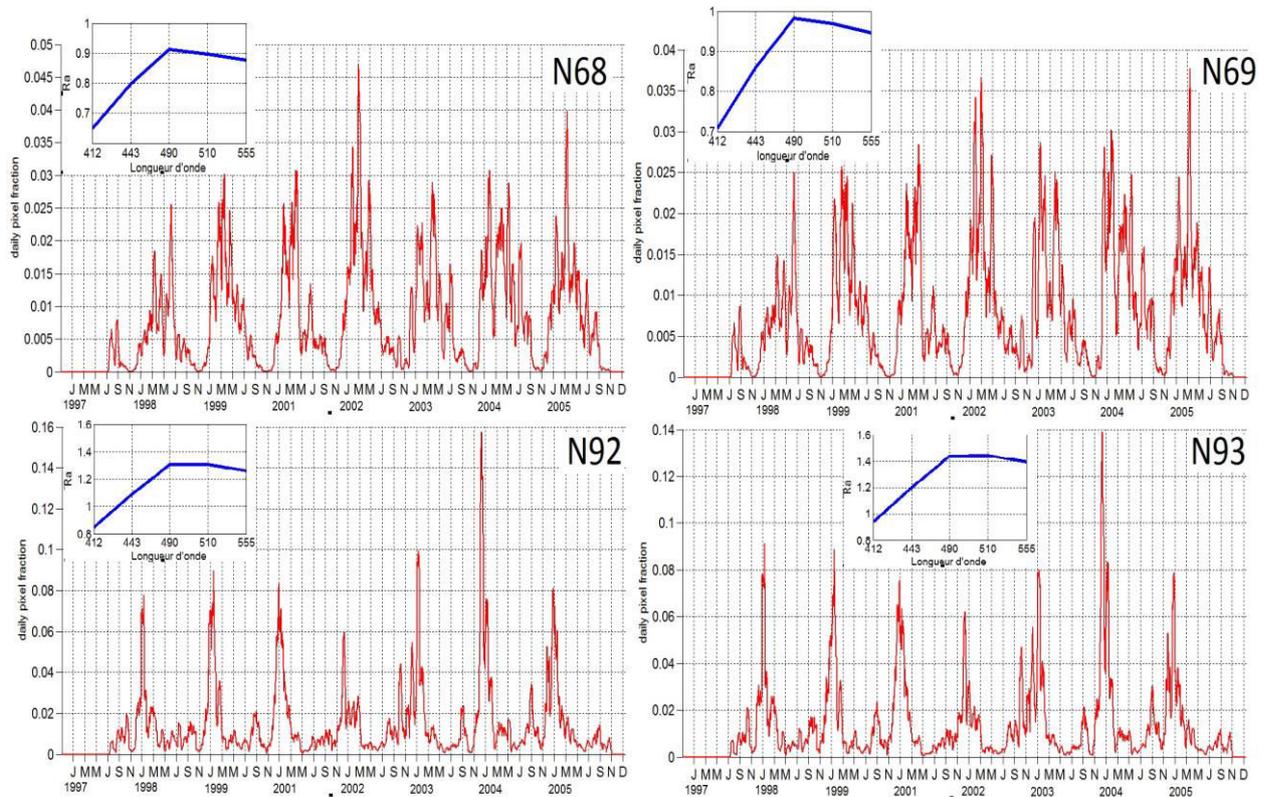


Figure 6.10. Fraction relative des pixels associés à deux paires de neurones adjacents de la classe 5 (eaux turbides du large): les neurones 68 et 69 (en haut) et les neurones 92 et 93 (en bas).

Contrairement à l'exemple présenté dans la figure 6.9, un cycle interannuel est clairement observable pour ces neurones. Par ailleurs, on remarque que les neurones adjacents au niveau de la carte SOM présentent la même périodicité, ce qui confirme que les anomalies spectrales similaires en termes de forme et amplitude sont caractérisées par le même cycle saisonnier. En effet, les neurones 68 et 69 présentent leur maximum d'occurrence entre les mois de Mars et Mai alors que les neurones 92 et 93 sont fréquemment détectés entre les mois de Décembre et Janvier. Ces observations pourraient être expliquées par le fait que les eaux de la classe 5 (eaux du large moins impactées par les apports terrigènes) soient influencées par une variabilité saisonnière d'origine biologique (e.g. phytoplancton et matériel détritique associé). En revanche, la classe 2, caractérisée par une prédominance de matériel particulaire ainsi qu'une fraction importante de matières dissoutes (e.g. CDOM) pourrait être, au vu de sa composition, moins sujette à des cycles interannuels clairs. Toutefois, ceci demeure une hypothèse devant être vérifiée en perspective, en relation avec des mesures *in situ* (e.g.

données relatives aux propriétés optiques inhérentes, matières en suspension, carbone organique particulaire). La labellisation de ces différents signaux en termes de groupes phytoplanctoniques dominants ou d'assemblages de groupes représente une perspective à cette étude préliminaire.

6.5. Perspectives

Les sorties de la méthode PHYSAT-SOM en région Manche et Mer du Nord pourront être utilisées à l'avenir pour une phase de labellisation des signaux en termes de groupes ou d'assemblages de groupes de phytoplancton. Ils ont par exemple été exploités lors d'un stage de master 1 effectué au LOG par Anne-Hélène Rêve (encadrée par N. Guiselin et S. Alvain). Ce travail a permis une première labellisation en se basant sur la base de données conséquente disponible au laboratoire. L'originalité est que cette base se compose de données d'abondance issues essentiellement du C.P.R. (Continuous Plankton Recorder, Warner et Hays, 1994 ; Reid et al., 2003), de comptages et de données pigmentaires. Grâce à l'utilisation d'une partie de cette base et à la définition d'une dominance, via des seuils multiples, différents spectres référents de la classe 5 ont pu être associés à la présence spécifique d'assemblages de groupes microphytoplanctoniques (20-200 μm).

Les spectres référents des neurones 7 et 40 sont par exemple associés à de fortes abondances en microphytoplancton par rapport aux neurones 25 et 52, qui sont associés à de faibles abondances. De même, l'assemblage des classes de phytoplancton varie d'un neurone à un autre. Ainsi, certaines signatures spectrales vont être associées à une large dominance par les diatomées (e.g. neurones 45 et 57) et d'autres vont être associées à une dominance des dinoflagellés (e.g. neurones 19 et 22). Quarante-deux spectres (figure 6.11) ont ainsi été associés à une classe dominante ou une co-dominance de plusieurs classes lors de ce premier test encourageant :

- Dominance de diatomées : 11 spectres.
- Dominance de prymnésiofycées : 2 spectres.
- Dominance de dinoflagellés : 16 spectres.
- Dominance de dictyofycées : 2 spectres.
- Composition mixte : 11 spectres.

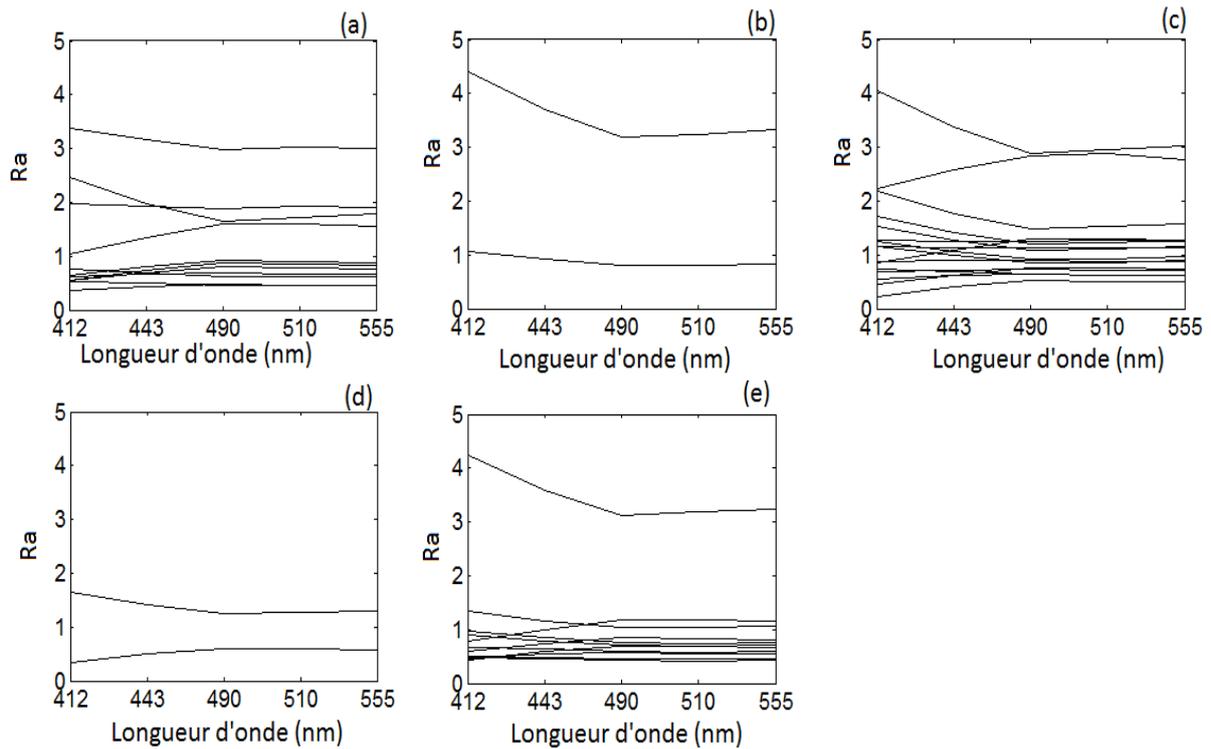


Figure 6.11. Spectres référents des neurones associés aux différentes compositions microphytoplanctoniques lors du stage d’A.H. Rêve: diatomées (a), prymnésiofycées (b), dinoflagellés (c), dictyophycées (d) et composition mixte (e)

On observe une forte variabilité des formes et amplitudes des spectres associés à une classe de microphytoplancton donnée. Cette hétérogénéité est en relation avec la complexité des interactions entre les différentes propriétés optiques inhérentes du milieu (Alvain et al., 2012). La diversité des tailles et formes des espèces existantes au sein de chaque classe (se référer à l’exemple des diatomées dans le chapitre 5) ou les stades de développement sont des explications possibles à cette forte variabilité spectrale.

A partir de ces premières informations et des études à venir, il sera possible d’élaborer des cartes de distribution des classes phytoplanctoniques, dont deux exemples sont montrés dans la figure 6.12 (répartition géographique des assemblages de microphytoplancton pour la journée du 13 Mai 2000 à gauche et fréquence relative des pixels caractérisés par une dominance des diatomées pour le mois de Mai 2001 à droite). Les premiers résultats de cartographie des spectres labellisés en Manche-Mer du Nord sont prometteurs puisque l’on retrouve des structures spatiales bien définies qui apparaissent sur les cartes et semblant suivre le courant Atlantique venant du Nord (pixels dominées par les diatomées et les dinoflagellés apparaissant sur la figure de gauche). Le travail de master a également permis de montrer

qu'il est possible de distinguer différentes espèces de diatomées, indiquant par là qu'une caractérisation précise des différents Ra était une étape indispensable.

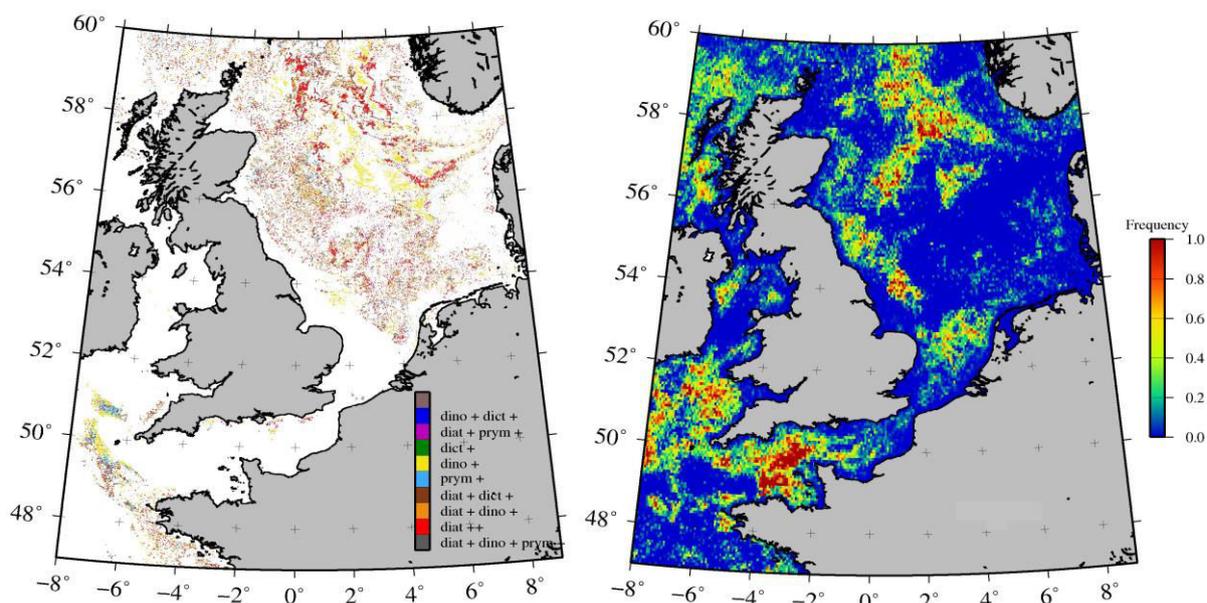


Figure 6.12. Premier test de labellisation des anomalies spectrales en Manche et Mer du Nord. A gauche, la répartition géographique des assemblages de microphytoplancton pour la journée du 13 Mai 2000 et à droite, la fréquence relative des pixels caractérisés par une dominance des diatomées pour le mois de Mai 2001 (Figure issue du rapport de Master d'Anne Hélène Rêve)

Ce chapitre a eu pour but principal de montrer la possibilité d'utiliser les anomalies spectrales de PHYSAT (Alvain et al., 2005, 2008) et de les classer avec les cartes auto-organisatrices dans les eaux du Cas 2. L'approche s'est basée sur une séparation préliminaire de différents types d'eau, selon leurs propriétés optiques (Vantrepotte et al., 2012). Cette classification préliminaire a permis une caractérisation de la variété de spectres de Ra spécifiques à chaque type d'eau, dans un environnement aux propriétés bio-optiques complexes. La distribution géographique des spectres montre des structures spatiales bien définies et l'existence de cycles interannuels (pour plus de détails se référer aux autres exemples de distributions spatio-temporelles, illustrés en annexe 4). Ces cycles sont surtout mis en évidence pour les classes d'eau moins sujettes aux interférences dues à du matériel d'origine non phytoplanctonique (en particulier la classe 5 représentant les eaux du large turbides, mais relativement moins impactées par les apports terrigènes).

L'utilisation future de l'intégralité de la base de données *in situ* disponible au laboratoire (sans se limiter aux données C.P.R.) permettra d'affiner la labellisation. On peut citer l'exemple de *Phaeocystis* qui n'est pas comptabilisé dans les données des campagnes

C.P.R. alors que sa présence en Manche et Mer du Nord est connue (Brunet et al., 1996 ; Lancelot et al., 1998 ; Breton et al., 2000 ; Rousseau et al., 2002; Schoemann et al., 2005; Schapira et al., 2008). Des travaux complémentaires seront donc nécessaire pour aller plus loin dans cette voie afin d'ajouter d'autres classes, assemblages ou même espèces, lorsque leurs caractéristiques particulières pourront être associées à un spectre de Ra spécifique. L'association de ces informations avec des mesures d'IOPs *in situ* ou dans le cadre d'expériences en microcosme (en projet dans l'équipe) devrait également permettre d'améliorer la détection des groupes à l'avenir.

CONCLUSIONS ET PERSPECTIVES

Rappel du Contexte et des objectifs

Dans le cadre de ma thèse, j'ai développé une méthode de classification automatique des anomalies de luminances marines (Ra, Alvain et al., 2005) dans le but de détecter les groupes de phytoplancton à l'échelle globale, à partir des mesures de la couleur de l'océan. L'identification des différentes populations de phytoplancton et la connaissance de leurs mécanismes de successions représentent des enjeux majeurs en océanographie. En effet, le phytoplancton est le premier maillon de la chaîne alimentaire marine. De par cette position clé au sein de l'écosystème marin, il conditionne la présence et la pérennité de nombreuses espèces (Rapport technique de la FAO, 2001). Par ailleurs, son action sur les cycles biogéochimiques des éléments, dont le carbone qu'il absorbe de l'atmosphère sous forme de CO₂ pour réaliser la photosynthèse, lui attribue un rôle important dans la régulation du climat global, de même que dans l'équilibre acido-basique des océans (Longhurst et al., 1995 ; Claustre et al., 1997 ; Behrenfeld et al., 2002 ; Hashimoto et Shiomoto, 2002). Le rôle du phytoplancton dans les cycles biogéochimiques des éléments et le climat est, cependant, étroitement lié aux groupes considérés (taxinomiquement) et à leur taille (Michaels et Silver, 1988 ; Claustre *et al.*, 1997 ; Hashimoto et Shiomoto, 2002).

Dans ce contexte, des algorithmes visant à quantifier ou distinguer différents groupes de phytoplancton depuis l'espace, en allant au-delà de la simple concentration en chlorophylle *a*, ont été développés ces dernières années. Ce travail de thèse s'est basé sur l'amélioration de l'une de ces méthodes, appelée PHYSAT (Alvain et al., 2005, 2008). Cette approche permet de distinguer différents groupes de phytoplancton à l'échelle globale, en se basant sur l'utilisation conjointe des pigments biomarqueurs et de la variabilité de second ordre des luminances marines normalisées (nLw, mesurées par les capteurs de la couleur de l'océan). Cette variabilité de second ordre des nLw, appelée « Ra » (Radiance Anomaly), est ainsi obtenue après l'élimination de l'effet de la concentration en chlorophylle *a* (variabilité de premier ordre) sur les spectres de nLw à différentes longueurs d'onde du visible. Grâce aux données *in situ* de pigments utilisées, des spectres d'anomalies de luminances (Ra), spécifiques en termes d'amplitude et de forme, ont été mis en évidence et ont permis la détection de groupes de phytoplancton dominants à l'échelle globale. Cependant, la méthode

PHYSAT restait limitée par l'approche suivie pour définir les signatures spectrales de chaque groupe de phytoplancton, nécessitant la définition d'enveloppes spectrales (tubes). En effet, il était nécessaire d'introduire des seuils minimaux et maximaux ainsi que des critères sur la forme des spectres afin d'éviter les chevauchements entre les différentes signatures spectrales des groupes de phytoplancton. Cette approche rendait difficile la prise en compte de certaines formes de spectres si celles-ci se retrouvaient «à cheval» entre différents tubes. Les possibilités de développement de PHYSAT étaient donc restreintes et l'entière variabilité spectrale des Ra ne pouvait être caractérisée avec précision. Les possibilités d'utilisation de grandes bases de données *in situ* de pigments, incluant une large variabilité de situations de dominances, étaient également limitées.

Il était donc nécessaire de développer une approche statistique objective, permettant une meilleure exploitation des anomalies spectrales pour l'étude de la répartition des groupes de phytoplancton. La méthode développée durant ma thèse, basée sur les anomalies spectrales de PHYSAT et leur classification automatique avec un type de réseaux de neurones (SOM ou Self-Organizing Maps, Kohonen, 1984), vise une caractérisation plus fine et fiable des spectres de Ra en termes de forme et d'amplitude et une intégration efficace de plus grandes bases de données *in situ* afin d'élargir les perspectives en termes de détection des groupes de phytoplancton.

Données et méthode

Afin d'obtenir une caractérisation précise des différentes formes et amplitudes des spectres de Ra, pouvant être rencontrées à l'échelle globale, une approche de classification automatique non supervisée, basée sur l'utilisation des cartes auto-organisatrices ou SOM (Kohonen, 1984), a été choisie. C'est une méthode de classification très adaptée aux grandes bases de données bruitées ou pouvant présenter des informations manquantes, telles que les données satellite. L'algorithme SOM a donc été appliqué à la classification d'un grand jeu de spectres de Ra : c'est la phase d'apprentissage. Plusieurs tests d'apprentissage ont été fait afin d'ajuster au mieux les paramètres du modèle à la base de données de spectres devant être classée.

Une fois ces spectres séparés en différents groupes (100 classes de spectres référents dans cette étude) et la qualité de la classification évaluée, une base de données *in situ* de pigments biomarqueurs, coïncidentes aux mesures du capteur SeaWiFS, a été utilisée dans le but de labelliser (étiqueter) les différentes signatures spectrales (i.e. neurones) en termes de

groupes de phytoplancton. L'identification des groupes de phytoplancton à partir des inventaires pigmentaires s'est basée sur l'utilisation de ratios sur les concentrations relatives (par rapport à celles de la chlorophylle *a*) des pigments biomarqueurs (Mackey et al., 1996 ; Alvain et al., 2005). Une approche de labellisation des neurones, visant à garder uniquement les cas de dominance (afin d'être en mesure de comparer les résultats de PHYSAT-SOM avec ceux de la précédente version de PHYSAT) et les signatures spectrales associées à une information *in situ* représentative, a été appliquée. L'étape suivante a consisté à utiliser ces différents spectres labellisés pour déduire la répartition globale des groupes de phytoplancton à partir de l'archive d'images de SeaWiFS. Pour cela, chaque pixel (i.e. spectre de R_a) est « projeté » sur les spectres référents de la carte SOM entraînée afin d'être affecté à son neurone le plus proche (i.e. spectre référent le plus proche en termes de distance euclidienne). Ainsi, des cartes globales de distribution des principaux groupes de phytoplancton ayant été identifiés à partir des inventaires de pigments ont été générées.

Résultats principaux de la méthode PHYSAT-SOM

La classification automatique des spectres de R_a avec SOM a permis de générer un ensemble de spectres référents. Il s'agit d'une compression de l'information contenue dans le jeu de données d'apprentissage. Ces vecteurs référents donnent ainsi une idée générale sur les formes et les amplitudes des spectres d'anomalies qu'on peut fréquemment retrouver à l'échelle globale. Leur analyse sur les différentes longueurs d'onde prises en compte a montré que les plus grandes variations des valeurs d'anomalies spectrales se produisent entre 412 nm et les plus grandes longueurs d'onde, où les valeurs de R_a deviennent relativement constantes (e.g. $R^2 = 0,46$ entre les valeurs de R_a à 412 nm et 555 nm). Ceci avait été mis en relation avec la sensibilité des R_a dans les courtes longueurs d'onde face aux variations des propriétés optiques inhérentes (Alvain et al., 2012).

L'avantage d'utiliser les cartes auto-organisatrices (PHYSAT-SOM) par rapport à l'ancienne approche de PHYSAT (en termes de caractérisation potentielle de la variabilité des anomalies spectrales) a été démontré dans le chapitre 3 en se basant sur l'ensemble des 41 spectres de R_a ayant servi à développer la précédente approche (PHYSAT version 2008). Cette comparaison a montré qu'avec SOM, il était possible d'exploiter toute l'information spectrale en classant l'intégralité des spectres contenus dans l'échantillon. En revanche, l'ancienne approche ne permettant pas de gérer d'une manière optimale les cas des recouvrements (spectres à cheval entre différents tubes d'anomalies), l'intégralité de

l'information ne pouvait être exploitée, 21 spectres sur 41 ne remplissant pas les critères de formes spécifiés au départ (Alvain et al., 2005, 2008). Cette étude a permis de démontrer le potentiel de SOM pour classifier et déterminer de façon fine et précise les spectres de Ra, répondant aux limitations et seuils imposés à PHYSAT.

Les signatures spectrales moyennes (spectres référents) des 4 groupes de phytoplancton identifiés à partir des inventaires de pigments (Nanoeucaryotes, *Prochlorococcus*, *Synechococcus* Like Cyanobacteria ou SLC et diatomées) sont relativement proches de celles obtenues par la version 2008 de PHYSAT. Ainsi, les nanoeucaryotes restent associés à des valeurs de Ra inférieures à 1. Les cyanobactéries (*Prochlorococcus*, *Synechococcus*) se caractérisent par des valeurs de Ra relativement constantes (spectres plats), situées entre 0,9 et 1,2. En revanche, l'utilisation d'une base de données *in situ* différente (par rapport à celle utilisée pour développer l'ancienne approche) a permis d'associer les diatomées à deux types de signatures spectrales, significativement différentes en termes de forme et d'amplitude.

La répartition globale des groupes est en accord avec de précédentes études ayant analysé la distribution du phytoplancton dans diverses régions du globe (Dandonneau et al., 2004; Zubkov et al., 2000 ; Marty et Chiavérini, 2002; DuRand et al., 2001; Steinberg et al., 2001; Longhurst, 2007). Comparativement aux résultats de la version originale de PHYSAT, les cartes de distribution des groupes de phytoplancton montrent des différences concernant, notamment, les blooms de diatomées dans certaines régions qu'on distingue mieux avec PHYSAT-SOM (e.g. upwellings, Océan Austral) ainsi que dans la répartition relative des *Prochlorococcus* et SLC. Ces derniers restent toutefois principalement dominants dans les eaux oligotrophes d'après les deux méthodes. L'étude du taux de couverture spatiale en pixels labellisés a montré que PHYSAT-SOM permettait d'avoir une moyenne de 2,75 fois plus de pixels labellisés pour les 4 groupes de phytoplancton. Cette amélioration du taux de couverture est toutefois associée à un changement de méthode (apport de SOM) et de base de données *in situ*. Le seul apport de cette dernière en termes de pourcentage de couverture a été estimé à 35 % de pixels labellisés en plus à l'échelle globale et en moyenne pour les 4 groupes de phytoplancton identifiés à partir des données pigmentaires. L'apport de la seule base de données *in situ* a été montré dans le chapitre 4 en faisant deux tests de labellisation de SOM (avec les 41 inventaires de pigments ayant servis à développer la précédente version de PHYSAT et avec toute la base de données *in situ* de pigments).

L'analyse de la variabilité spatio-temporelle des différentes signatures spectrales associées aux diatomées dans le Pacifique Equatorial et l'Océan Austral a montré que ces signatures occupaient des zones géographiques spécifiques et se caractérisaient pas des cycles saisonniers distincts. En outre, les diverses études réalisées dans ces régions montrent l'existence d'une grande diversité taxonomique de diatomées, mais aussi, la présence d'une grande hétérogénéité intra-spécifique en termes de formes et tailles de cellules (e.g. *Fragilariopsis Kerguelensis* dans l'Océan Austral). Ces différentes informations, de même que l'étude des valeurs d'IOPs, montrant que chaque signature spectrale se trouve dans un environnement optique spécifique, peuvent être utilisés pour soutenir des hypothèses liées à la présence de différents genres ou stades de floraisons afin d'expliquer la différence des signatures spectrales caractérisant les diatomées.

Les résultats exposés dans le chapitre 6 ont montré la possibilité d'utiliser les anomalies spectrales de PHYSAT dans les eaux du Cas 2, moyennant une approche adéquate de normalisation des luminances marines selon les différents types d'eau considérés (Vantrepotte et al., 2012). Le premier test de labellisation établi à partir des spectres de Ra classés avec SOM et d'une partie des informations *in situ* disponibles en Manche-Mer du Nord (dans le cadre du stage de master 1 d'A.H Rêve, encadrée par S.Alvain et N.Guiselin) a abouti à des résultats préliminaires encourageants. En effet, des structures spatiales bien définies, relatives à la distribution de certaines classes microphytoplanctoniques, ont été mises en évidence dans la région atelier.

Perspectives

Les perspectives émergent de ce travail de thèse concernent, en premier lieu, l'exploitation de l'outil SOM, conjointement à d'autres données *in situ* (e.g. pigments, IOPs, paramètres physiques tels que température de surface et vents), pour approfondir nos connaissances actuelles des mécanismes bio-optiques et physiques intervenant dans la détermination des formes et amplitudes des spectres de Ra. En effet, différentes signatures spectrales existantes à l'échelle globale n'ont été affectées à aucun groupe de phytoplancton dominant. Il sera donc intéressant de comprendre à quelle (s) information (s) pourraient correspondre ces spectres en les liant à des bases de données *in situ* dans différentes régions et en ne considérant, par exemple, pas uniquement les cas de dominance de groupes. Pour cela, un travail préliminaire sur les bases de données *in situ* de pigments disponibles actuellement à l'échelle globale reste à faire afin de séparer les mesures selon les cas de compositions mixtes

des groupes pouvant être observés. Ce travail sera essentiel pour simplifier la tâche d'affectation des différentes signatures spectrales actuellement non expliquées par une situation de dominance. Par ailleurs, il serait intéressant de tester d'autres approches de détermination de la dominance des groupes de phytoplancton, en établissant des critères basés, par exemple, sur la biomasse ou l'abondance spécifique au lieu des ratios sur la concentration relative des pigments biomarqueurs, utilisés actuellement.

Par ailleurs, une étude récente (Alvain et al., 2013) a montré l'évolution des populations de diatomées dans les hautes latitudes en fonction de paramètres climatiques (e.g. NAO ou North Atlantic Oscillation et SAM ou Southern Annular Mode) ou de l'évolution de l'épaisseur de la couche de mélange qui représente un des principaux facteurs déterminant la dominance des diatomées (Weber & El-Sayed, 1987; Jochem et al., 1995). Cette étude a cependant été faite avec l'ancienne version de PHYSAT (Alvain et al., 2005, 2008). Compte-tenu des améliorations apportées par PHYSAT-SOM, notamment en ce qui concerne la détection des blooms de diatomées dans certaines conditions, et l'existence d'archives satellites sur des échelles décennales, il sera essentiel à l'avenir d'appliquer la nouvelle approche proposée dans le cadre de ma thèse dans des études similaires. En outre, les différentes hypothèses formulées dans le chapitre 5, concernant l'origine des différentes signatures spectrales associées aux diatomées, devront être étayées au fur et à mesure de la disponibilité de données *in situ* dans les régions où elles sont observées. Des études optiques poussées (e.g. expériences en microcosme actuellement en projet dans l'équipe) devraient également aider à étudier séparément les différentes hypothèses formulées dans le cadre de ma thèse pour connaître l'origine précise de ces différentes signatures spectrales.

L'adaptation de PHYSAT-SOM à d'autres capteurs de la couleur de l'Océan, actuellement en service (e.g. VIIRS, MODIS, GOCI), devra également être élaborée. L'essentiel du travail d'adaptation de la méthode d'un capteur à un autre concernera, en premier lieu, le calcul d'une LUT de valeurs de nLw_{ref} spécifiques à chaque capteur afin de tenir compte de la variation des estimations de la concentration en chlorophylle *a* et de nLw , pouvant significativement varier d'un capteur et d'une longueur d'onde à une autre (Zibordi et al., 2006 ; Lee et al., 2010). Ensuite, il faudra identifier les différences (biais) engendrées par le changement de capteur sur les anomalies spectrales et les spectres référents (après un nouvel apprentissage). Cette étape essentielle devrait permettre par la suite d'évaluer la nécessité de refaire un apprentissage (ou non) quand on change de capteur. D'autres tests seront également à prévoir, notamment en ce qui concerne la nécessité de refaire l'étape de

labellisation. Un essai préliminaire de la méthode PHYSAT-SOM sur les données du capteur MODIS à l'échelle globale a cependant été fait en utilisant la même approche suivie pour le capteur SeaWIFS (apprentissage et labellisation), montrant que les structures de groupes sont relativement bien conservées entre les deux capteurs, malgré la différence des bases de données *in situ* utilisées pour labelliser SOM pour chaque capteur (Annexe 3).

Les données de variations spatio-temporelles des anomalies spectrales fréquemment observées en Manche et Mer du Nord ont été utilisées conjointement à une base de données *in situ* disponible dans la région atelier, pour établir un premier test de labellisation des spectres d'anomalies. Ce test préliminaire a montré des résultats encourageants et constitue une première étape dans l'exploitation des cartes auto-organisatrices et des informations de terrain dans un environnement optiquement complexe. Il devra être consolidé avec d'autres études (e.g. des études optiques en relation avec la composition phytoplanctonique) et complété en utilisant d'autres approches de labellisation que celle basée sur l'abondance. En outre, les données *in situ* considérées dans le cadre ce test concernent uniquement le microphytoplancton (jeu de données C.P.R.). L'utilisation future de l'intégralité de la base de données *in situ* disponible au laboratoire (sans se limiter aux données C.P.R.) permettra d'affiner la labellisation. Ce premier travail montre cependant qu'il est possible de détecter par satellite la présence de classes phytoplanctoniques dans la région atelier étudiée. Certaines d'entre elles ont ainsi pu être associées à des signatures spectrales. Bien que ces dernières soient très variables en termes de formes et d'amplitudes (complexité des interactions entre les différentes propriétés optiques inhérentes du milieu, variabilité intra-classe, etc.), les premières cartes montrent des répartitions géographiques caractérisées par l'existence de structures spatiales bien définies. Des travaux complémentaires peuvent donc être envisagés pour aller plus loin dans cette voie afin d'ajouter d'autres classes, assemblages ou espèces, lorsque leurs caractéristiques particulières pourront être associées à un spectre de Ra spécifique.

Ces premiers résultats laissent également envisager la détection d'assemblages particuliers de groupes de phytoplancton. Ceci pourrait permettre, par exemple, l'établissement de modèles de niches écologiques en Manche-Mer du Nord, en montrant le comportement des classes et espèces étudiées au sein d'un environnement aux caractéristiques hydrodynamiques et optiques complexes.

Bibliographie

Aceituno, P. (1988). On the functioning of the Southern Oscillation in the South American Sector. Part I : Surface Climate. *Monthly Weather Review*, 116 : 505-524.

Aiken, J., Fishwick, J. R., Lavender, S., Barlow, R., Moore, G. F., Sessions, H., et al. (2007). Validation of MERIS reflectance and chlorophyll during the BENCAL cruise October 2002: Preliminary validation of new demonstration products for phytoplankton functional types and photosynthetic parameters. *International Journal of Remote Sensing*, 28 : 497-516.

Ainsworth, E. J. (1999). Visualization of ocean colour and temperature from multispectral imagery captured by the Japanese ADEOS satellite. *Journal of Visualization*, 2 : 195-204.

Ainsworth, E. J., et Jones, S. F. (1999). Radiance spectra classification from the ocean color and temperature scanner on ADEOS. *IEEE Transactions on Geoscience and Remote Sensing*, 37 : 1645-1656.

Allan, R.J., Lindesay, J.A., et Parker, D.E., 1996. El Niño Southern Oscillation and Climatic Variability, CSIRO, Collingwood (Australia), 405 p.

Alvain, S., Moulin, C., Dandonneau, Y., et Breon, F. M. (2005). Remote sensing of phytoplankton groups in case 1 waters from global SeaWiFS imagery. *Deep Sea Research I*, 52 : 1989-2004.

Alvain, S., Moulin, C., Dandonneau, Y., et Loisel, H. (2008). Seasonal distribution and succession of dominant phytoplankton groups in the global ocean: A satellite view. *Global Biogeochemical Cycles*, 22, GB3001, doi: 10.1029/2007GB003154.

Alvain, S., Loisel H., et Dessailly D. (2012). Theoretical analysis of ocean color radiances anomalies and implications for phytoplankton groups detection in case 1 waters. *Optics Express*, 20, 1070-1083.

Alvain, S., Le Quéré, C., Bopp, L., Racault, M.F., Beaugrand, G., Dessailly, D. et Buitenhuis, E. (2013). Rapid climatic driven shifts of diatoms at high latitudes. *Remote Sensing of Environment*, 132 : 195-201.

Anderson, D.M. (1989). Toxic algal blooms and red tides: A global perspective, pp. 11-16. In: T. Okaichi, D. M. Anderson, and T. Nemoto (ed.) Red Tides: Biology, Environmental Science and Toxicology. Elsevier, New York.

Anderson, D.M., Cembella, A.D., et Hallegraeff, G.M. (1998). *Physiological ecology of harmful algal blooms*, NATO ASI series, vol G41. Springer-Verlag, Berlin: 662 pp.

- Anderson**, D.M., Hoagland, P., Kaoru, Y., et White, A.W. (2000). Estimated Annual Economic Impacts from Harmful Algal Bloom (HABs) in the United States, *Technical Report WHOI-2000-11 Woods Hole Oceanographic Institute*, Woods Hole, Mass.
- Armand**, L.K., Crosta, X., Romero, O., et Pichon, J.-J. (2005). The biogeography of major diatom taxa in Southern Ocean sediments: 1. Sea ice related species. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 223 : 93-126.
- Armand**, L.K., Cornet-Barthaux, V., Mosseri, J., et Quéguiner, B. (2008). Late summer diatom biomass and community structure on and around the naturally iron-fertilised Kerguelen Plateau in the Southern Ocean. *Deep-Sea Research II*, 55 : 653-676.
- Assmy**, P., Henjes, J., Schmidt, K., Smetacek, V., et Montresor, M. (2005). The wax and wane of an iron-induced diatom bloom in the Southern Ocean, *Berichte zur Polar und Meeresforschung*, 500 : 89-100.
- Astoreca**, R., Rousseau, V., Ruddick, K., Knechciak, C., Van Mol, B., Parent, J-Y., et Lancelot, C. (2009). Development and application of an algorithm for detecting *Phaeocystis globosa* blooms in the Case 2 Southern North Sea waters. *Journal of Plankton Research*, 31 : 287-300.
- Aumont**, O., Maier-Reimer, E., Blain, S., et Pondaven, P. (2003). An ecosystem model of the global ocean including Fe, Si, P co-limitations. *Global Biogeochemical Cycles*, 17, doi:10.1029/2001GB001745.
- Bailey**, S.W., et Werdell, P.J. (2006). A multi-sensor approach for the on-orbit validation of ocean color satellite data products. *Remote Sensing of Environment*, 102 : 12-23.
- Bailey**, S.W., Hooker, S.B., Antoine, D., Franz, B.A., et Werdell, P.J. (2008). Sources and assumptions for the vicarious calibration of ocean color satellite observations. *Applied Optics*, 47: 2035—2045.
- Bailey**, S.W., Franz, B.A., et Werdell, P.J. (2010). Estimation of near-infrared water-leaving reflectance for satellite ocean color data processing. *Optics Express*, 18 : 7521–7527.
- Bailly Du Bois**, P., et Dumas, F. (2005). Fast hydrodynamic model for medium and long-term dispersion in seawater in the English Channel and southern North Sea, qualitative and quantitative validation by radionuclide tracers. *Ocean Modelling*, 9 : 169-210.
- Banerjee**, A.K., Arora, N., et Murty, U.S. (2009). Clustering and Classification of Anopheline Spacer Sequences using Self Organizing Maps. *The Internet Journal of Genomics and Proteomics*, 4, doi: 10.5580/15b.

Barlow, R.G. (1984). Dynamics of the decline of a phytoplankton bloom after an upwelling event. *Marine Ecology Progress Series*, 16: 121-126.

Barlow, R.G., Aiken, J., Holligan, P.M., Cummings, D.G., Maritorea, S., et Hooker, S. (2002). Phytoplankton pigment and absorption characteristics along meridional transects in the Atlantic Ocean. *Deep-Sea Research I*, 49 : 637-660.

Behrenfeld, M. J., Esaias, W.E., et Turpie, K.R. (2002). Assessment of Primary Production at the Global Scale. *In: Phytoplankton Productivity: Carbon assimilation in marine and freshwater ecosystems*, P. J. I. B. Williams, D. N. Thomas et R. A. Reynolds (eds), Blackwell Publishing Company: 156-186.

Belviso, S., Moulin, C., Bopp, L., et Stefels, J. (2004). Assessment of a global climatology of 686 oceanic dimethylsulfide (DMS) concentrations based on SeaWiFS imagery (1998 – 2001), *Canadian Journal of Fisheries and Aquatic Sciences*, 61 : 804-816.

Beucher, F. (2010). *Manuel de météorologie tropicale : des alizés au cyclone (2 tomes)*, Météo-France.

Bidigare, R. R., Prézelin, B.B., et Smith, R.C. (1992). Bio-optical models and the problems of scaling. *In: Primary Productivity and Biogeochemical Cycles in the Sea*, P. G. Falkowski et A. D. Woodhead (eds), Plenum Press: 175-212.

Bishop, C.M. (2006). *Pattern Recognition And Machine Learning*, Springer.

Blanchot, J., André, J.M., Navarette, C., Neveux, J., et Radenac, M.H. (2001). Picophytoplankton in the equatorial Pacific: Vertical distributions in the Warm Pool and in the High nutrient low chlorophyll conditions. *Deep-Sea Research I*, 48 : 297-314.

Borrione, I., et Schlitzer, R. (2012). Distribution and recurrence of phytoplankton blooms around South Georgia, Southern Ocean. *Biogeosciences Discussions*, 9 : 10087-10120, doi: 10.5194/bgd-9-10087-2012.

Bouman, H. A., Platt, T., Sathyendranath, S., Li, W.K.W., Stuart, V., Fuentes-Yaco, C., Maass, H., Horne, E.P.W., Ulloa, O., Lutz, V.A., et Kyewalyanga, M. (2003). Temperature as indicator of optical properties and community structure of marine phytoplankton: implications for remote sensing. *Marine Ecology Progress Series*, 258: 19-30.

Bracher, A., Vountas, M., Dinter, T., Burrows, J. P., Röttgers, R., et Peeken, I. (2009). Quantitative observation of cyanobacteria and diatoms from space using PhytoDOAS on SCIAMACHY data. *Biogeosciences*, 6, 751-764.

Brajard, J., Jamet, C., Moulin, C., et Thiria, S. (2006). Use of a neuro-variational inversion for retrieving oceanic and atmospheric constituents from satellite ocean colour sensor: Application to absorbing aerosols. *Neural Networks*, 19 : 178-185.

- Brajard, J., Niang, A., Sawadogo, S., Fell, F.R., Santer, R., et Thiria, S. (2007).** Estimating aerosol parameters above ocean from MERIS observation using topological maps, *International Journal of Remote Sensing*, 28 : 781-795.
- Breton, E., Brunet, C., Sautour, B., et Brylinski, J-M. (2000).** Annual variations of phytoplanktonic biomass in the Eastern English Channel: comparison by pigment signatures and microscopic counts. *Journal of Plankton Research*, 22 : 1423-1440.
- Brewin, R. J. W., Sathyendranath, S., Hirata, T., Lavender, S., Barciela, R. M., et Hardman-Mountford, N. J. (2010).** A three-component model of phytoplankton size class for the Atlantic Ocean. *Ecological Modelling*, 221, 1472-1483.
- Brewin, R.J.W., Hardman-Mountford, N.J., Lavender, S., Raitsos, D.E., Hirata, T., Uitz, J., Devred, E., Bricaud, A., Ciotti, A.M., et Gentili, B. (2011).** An intercomparison of bio-optical techniques for detecting dominant phytoplankton size class from satellite remote sensing. *Remote Sensing of Environment*, 115, 325–339.
- Brzezinski, M.A., Villareal, T.A., et Lipschultz, F. (1998).** Silica production and the contribution of diatoms to new and primary production in the central North Pacific. *Marine Ecology Progress Series*, 167 : 89-104.
- Bricaud, A., Morel, A., et Prieur, L. (1983).** Optical efficiency factors of some phytoplankters. *Limnology and Oceanography*, 28 : 816-832.
- Bricaud, A., Bédhomme, A.L., et Morel, A. (1988).** Optical properties of diverse phytoplanktonic species : experimental results and theoretical interpretation, *Journal of Plankton Research*, 10 : 851-873.
- Bricaud, A., Morel, A., Babin, M., Allali, K., et Claustre, H. (1998).** Variations of light absorption by suspended particles with chlorophyll a concentration in oceanic (case 1) waters: Analysis and implications for bio-optical models. *Journal of Geophysical Research*, 103 : 31033-31044.
- Bricaud, A., Claustre H., Ras J., et Oubdelkhier K. (2004).** Natural variability of phytoplanktonic absorption in oceanic waters : influence of the size structure of algal populations, *Journal of Geophysical Research*, 109, doi:10.1029/2004JC002419.
- Bricaud, A., Ciotti, A.M., et Gentili, B. (2012).** Spatial-temporal variations in phytoplankton size and colored detrital matter absorption at global and regional scales, as derived from twelve years of SeaWiFS data (1998-2009). *Global Biogeochemical Cycles*, 26, doi:10.1029/2010GB003952.
- Brown, C. W., et Yoder, J. A. (1994).** Coccolithophorid blooms in the global ocean. *Journal 712 of Geophysical Research*, 99 : 7467-7482.

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 : 81-103.

Brunet, C., Brylinski, J-M., Bodineau, L., Thoumelin, G., Bentley, D., et Hilde, D. (1996). Phytoplankton dynamics during the spring bloom in the South-Eastern English Channel. *Estuarine, Coastal and Shelf Science*. 43 : 469-483.

Brylinski, J.-M., Lagadeuc, Y., Gentilhomme, V., Dupont, J.-P., Lafite, R., Dupeuble, P.-A., Huault, M.-F., Auger, Y., Puskaric, E., Wartel, M., Cabioch, L. (1991). Le "fleuve Côtier": un phénomène hydrologique important en Manche Orientale. Exemple du Pas de Calais. *Oceanologica Acta*, 11: 197-203.

Buitenhuis, E.T., Le Quéré, C., Aumont, O., Bunker, A., Hirst, A., Ikeda, T., O'Brien, T., et Piontkovski, S. (2006). Biogeochemical fluxes through mesozooplankton. *Global Biogeochemical Cycles*, 20, GB2003, doi: 10.1029/2005GB002511.

Burd, A. B., et Jackson, G. A. (2009). Particle aggregation. *Annual Review of Marine Science*, 1 : 65-90.

Burges, C.J.C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, 2 : 121–167.

Campbell, J., Antoine, D., Armstrong, R., Arrigo, K.R., Balch, W.M., Barber, R.T., Behrenfeld, M.J., Bidigare, R.R., Bishop, J., Carr, M.E., Esaias, W., Falkowski, P.G., Hoepffner, N., Iverson, R., Kiefer, D.A., Lohrenz, S.E., Marra, J., Morel, A., Ryan, J., Vedernikov, V., Waters, K., Yentsch, C., et Yoder, J.A. (2002). Comparison of algorithms for estimating ocean primary production from surface chlorophyll, temperature, and irradiance. *Global Biogeochemical Cycles*, 16, doi : 10.1029/2001GB001444.

Cefarelli, A.O., Ferrario, M.E., Almandoz, G.O., Atencio, A.G., Akselman, R., et Vernet, M. (2010). Diversity of the diatom genus *Fragilariopsis* in the Argentine Sea and Antarctic waters: morphology, distribution and abundance. *Polar Biology*, 33 : 1463-1484.

Chazottes, A., Annick, B., Crépon, M., et Thiria, S. (2006). Statistical analysis of a database of absorption spectra of phytoplankton and pigment concentrations using self-organizing maps. *Applied Optics*, 45 : 8102-8115.

Chazottes, A., Crépon, M., Bricaud, A., Ras, J., et Thiria, S. (2007). Statistical analysis of absorption spectra of phytoplankton and of pigment concentrations observed during three POMME cruises using a neural network clustering method. *Applied Optics*, 46 : 3790-3799.

Chen, D. G., et Ware, D. M. (1999). A neural network model for forecasting fish stock recruitment. *Canadian Journal of Fisheries and Aquatic Sciences*, 56 : 2385-2396.

- Ciotti**, A. M., et Bricaud, A. (2006). Retrievals of a size parameter for phytoplankton and spectral light absorption by coloured detrital matter from water-leaving radiances at SeaWiFS channels in a continental shelf off Brazil. *Limnology and Oceanography: Methods*, 4 : 237–253.
- Claustre**, H. (1994). Phytoplankton pigment signatures of the trophic status in various oceanic regimes. *Limnology and Oceanography*, 39 : 1206-1211.
- Claustre**, H., Hooker, S.B., Van Heukelem, 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 methods using in situ samples: Application to remote sensing and database activities. *Marine Chemistry*, 85 : 41-61.
- Claustre**, H., Moline, M.A., et Prezelin, B.B. (1997). Sources of variability in the column photosynthetic cross section for Antarctic coastal waters. *Journal of Geophysical Research*, 102: 047-25.
- Chavez**, F.P., Strutton, P.G., Friederich, G.E., Feely, R.A., Feldman, G.C., Foley, D.G., et McPhaden, M.J. (1999). Biological and chemical response of the equatorial Pacific ocean to the 1997–98 El Niño. *Science*, 286, 2126-2131.
- Chisholm**, S. W. (1992). Phytoplankton size. *In: Primary Productivity and Biogeochemical Cycles in the Sea*, P. G. Falkowski et A. D. Woodhead (eds), Plenum Press: 213-237.
- Cortese**, G., et Gersonde, R. (2007). Morphometric variability in the diatom *Fragilariopsis kerguelensis*: Implications for Southern Ocean paleoceanography. *Earth and Planetary Science Letters*, 257 : 526-544.
- Coreste**, G., Gersonde, R., Maschner, K., et Medley, P. (2012). Glacial-interglacial size variability in the diatom *Fragilariopsis kerguelensis*: Possible iron/dust controls ? *Paleoceanography*, 27 : doi : 10.1029/2011PA002187.
- Cottrell** M., Fort J.C., et Pagès G. (1997). *Theoretical aspects of the SOM algorithm, WSOM'97*, Helsinki, pp : 246-267.
- Craig**, S.E., Lohrenz, S.E., Lee, Z., Mahoney, K.L., Kirkpatrick, G.J., Schofield, O.M., et Steward, R.G. (2006). Use of hyperspectral remote sensing reflectance for detection and assessment of the harmful alga, *Karenia brevis*. *Applied Optics*, 45 : 5414–5425.
- Crosta**, X., Romero, O., Armand, L.K., Pichon, J.-J. (2005). The biogeography of major diatom taxa in Southern Ocean sediments: 2. Open ocean related species. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 223 : 66-92.

- D'Ortenzio**, F., Antoine, D., Martinez, E., Ribera d'Alcalà, M. (2012). Phenological changes of oceanic phytoplankton in the 1980s and 2000s as revealed by ocean-color remote-sensing observations. *Global Biogeochemical Cycles*, 26, doi:10.1029/2011GB004269.
- Darecki**, M., and Stramski, D. (2004). An evaluation of MODIS and SeaWiFS bio-optical algorithms in the Baltic Sea. *Remote Sensing of Environment*, 89 :326-350.
- Dandonneau**, Y., Vega, A., Loisel, H., Du Penhoat, Y., et Menkes, C. (2003). Oceanic Rossby waves acting as a “hay rake” for ecosystem floating by-products. *Science*, 302 : 1548-1551.
- Dandonneau**, Y., Deschamps, P.Y., Nicolas, J.M., Loisel, H., Blanchot, J., Montel, Y., Thieuleux, F., et Bécu, G. (2004). Seasonal and interannual variability of ocean color and composition of phytoplankton communities in the North Atlantic, Equatorial Pacific and South Pacific. *Deep-Sea Research II*, 51 : 303-318.
- Dayhoff**, J. E. (1990). Neural network architectures : An introduction. New York: Van Nostrand Reinhold.
- Deboeck** G., et Kohonen T. (1998). *Visual Explorations in Finance with Self-Organizing Maps*, Springer.
- Devred**, E., Sathyendranath, S., Stuart, V., Maas, H., Ulloa, O., et Platt, T. (2006). A two-component model of phytoplankton absorption in the open ocean: Theory and applications. *Journal of Geophysical Research*, 111, doi:10.1029/2005JC002880.
- Djavidnia**, S., Mélin, F., et Hoepffner, N. (2010). Comparison of Global Ocean Colour Data Records. *Ocean Science*, 6 : 61-76.
- Dickinson**, R.E. (1978). Rossby waves - long-period oscillations of oceans and atmospheres. *Annual Review of Fluid Mechanics*, 10 : 10-195.
- DiTullio**, G.R., Geesey, M.E., Maucher, J.M., Alm, M.B., Riseman, S.F et Bruland, K.W. (2005). Influence of iron on algal community composition and physiological status in the Peru upwelling system. *Limnology and Oceanography*, 50 : 1887-1907.
- Diouf**, D., Niang, A., Brajard, J., Crépon, M., et Thiria, S. (2013). Retrieving aerosol characteristics and sea-surface chlorophyll from satellite ocean color multi-spectral sensors using a neural-variational method. *Remote Sensing of Environment*, 130, 74-86.
- Dreyfus**, G., Martinez, J. M., Samuelides, M., Gordon, M. B., Badran, F., Thiria, S. et Hérault, L. *Réseaux de neurones : méthodologie et applications*. Éditions Eyrolles, Paris, 386p.

- DuRand**, M. D., Olson, R.J., et Chisholm, S.W. (2001). Phytoplankton population dynamics at the Bermuda Atlantic time-series station in the Sargasso Sea. *Deep-Sea Research II*, 48 : 1983-2003.
- Enfield**, D.B., 1989. El Niño, past and present. *Reviews of Geophysics*, 27 : 159-187.
- Eplee**, R.E., Robinson, W.D., Schieber, B.D., McClain, C.R., et Darzi, M. (1998). Vicarious Calibration of SeaWiFS. *Proceedings IGARSS*, 5 : 2707-2709.
- Eplee**, R.E., Robinson, W.D., Bailey, S.W., et Clark, D.K. (2001). Calibration of SeaWiFS. II. Vicarious techniques. *Applied Optics*, 40: 6701-6718.
- Eplee**, R.E., Bailey, S.W., Barnes, R.A., Kieffer, H.H., et McClain, C.R. (2006). Comparison of SeaWiFS on-orbit lunar and vicarious calibrations. *Proceedings of SPIE*, 6296, 629610.
- Eplee**, R.E., Xiong, X., Sun, J-Q., Meister, G., et McClain, C.R. (2009). The Cross Calibration of SeaWiFS and MODIS using On-Orbit Observations of the Moon. *Proceedings of SPIE*, 7452, 74520.
- Everitt**, D.A., Wright, S.W., Volkman, J.K., Thomas, D.P., et Lindstrom, E. (1990) Phytoplankton community compositions in the western equatorial Pacific determined from chlorophyll and carotenoid pigment distribution. *Deep Sea Research I*, 37 : 975–997.
- Falkowski**, P. G., Katz, M., Knoll, A.H., Quigg, A., Raven, J.A., Schofield, O., et Taylor, F.J.R. (2004). The Evolution of Modern Eukaryotic Phytoplankton. *Science*, 305: 354-360.
- FAO**. *Rapport sur les pêches*. No. 634. Rome, FAO. 2001. 44p. (<http://www.fao.org/docrep/meeting/003/x8623f.htm>)
- Flombaum**, P., Gallegos, J.L., Gordillo, R.A., Rincon, J., Zabala, L.L., Jiao, N., Karl, D.M., Li, W.K.W. et al. (2013). Present and future global distributions of the marine Cyanobacteria Prochlorococcus and Synechococcus. *Proceedings of the National Academy of Sciences*, 110: 9824-9829.
- Fourtanier**, E., et Kociolek, J.P. (2003). Catalogue of the diatom genera (vol 14, pg 190, 1999). *Diatom Res* 18: 245-258.
- Franz**, B.A., Bailey, S.W., Werdell, P.J., et McClain, C.R. (2007). Sensor-independent Approach to the Vicarious Calibration of Satellite Ocean Color Radiometry. *Applied Optics*, 46: 5068-5082.
- Frederiksen**, M., Edwards, M., Richardson, A.J., Halliday, N., et Wanless, S. (2006). From plankton to top predators: bottom-up control of a marine food web across four trophic levels. *Journal of Animal Ecology*, 75, doi : 10.1111/j.1365-2656.2006.01148.x.

Garver, S., et Siegel, D. A. (1997). Inherent optical property inversion of ocean color spectra and its biogeochemical interpretation: 1 time series from the Sargasso Sea. *Journal of Geophysical Research*, 102 : 18607-18625.

Gentilhomme, V., et Lizon, F. (1998). Seasonal cycle of nitrogen and phytoplankton biomass in a well-mixed coastal system (Eastern English Channel). *Hydrobiologia*, 361: 191-199.

Gibbs, C.F. (1996). Chlorophyll b interference in the fluorometric determination of chlorophyll a and « phaeo-pigments ». *Australian journal of marine freshwater research.*, 30 : 597-606.

Gieskes, W.W.G., Kraay, G.W., Nontji, A., Setiapermana, D., et Sutomo (1988). Monsoonal alteration of a mixed and a layered structure of a euphotic zone of the Banda Sea (Indonesia): a mathematical analysis of algal pigment fingerprints. *Netherlands Journal of Sea Research*, 22 : 123-13.

Gonçalves, L.F., Bosa, J.L., Balen, T.R., Lubaszewski, M.S., Schneider, E.L., Henriques, R.V. (2011). Fault Detection, Diagnosis and Prediction in Electrical Valves Using Self-Organizing Maps. *Journal of Electron Test*, doi : 10.1007/s10836-011-5220-0.

Gordon, H.R. (1997). Atmospheric correction of ocean color imagery in the Earth Observing System era. *Journal of Geophysical Research*, 102, doi: 10.1029/96JD02443.

Gordon, H. R., Du, T., et Zhang, T. (1997). Remote sensing of ocean color and aerosol properties : resolving the issue of aerosol absorption. *Applied Optics*, 36 : 8670-8684.

Gordon, H., et Morel, A. (1983). Remote Assessment of Ocean Color for Interpretation of Satellite Visible Imagery: A Review. *Lecture Notes on Coastal and Estuarine Studies*, 4, Springer Verlag, New York, 114 pp.

Gordon, H.R., et Wang, M. (1994). Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with SeaWiFS: a preliminary algorithm. *Applied Optics*, 33 : 443-452.

Gorgues, T., Menkes, C., Slemmons, L., Aumont, O., Dandonneau, Y., Radenac, M.H., Alvain, S., et Moulin, C. (2010). Revisiting the La Niña 1998 phytoplankton blooms in the equatorial Pacific. *Deep-Sea Research I*, 57 : 567-576.

Goyens, C., Jamet, C., et Schroeder, T. (2013). Evaluation of four atmospheric correction algorithms for MODIS-Aqua images over contrasted coastal waters. *Remote Sensing of Environment*, 131 : 63-75.

Gregg, W.W., et Casey, N.W. (2007). Sampling biases in MODIS and SeaWiFS ocean chlorophyll data. *Remote Sensing of Environment*, 111 : 25-35.

Hasan, S., et Shamsuddin, S.M. (2011). Multistrategy Self-Organizing Map Learning for Classification Problems. *Computational Intelligence and Neuroscience*, doi : dx.doi.org/10.1155/2011/121787.

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: 83-87.

Hedin-Bougart, M. (1980). *Hydrologie littorale: phytoplancton (région de Dunkerque – Mer du Nord)*. Thèse de doctorat, Université des Sciences et Technologies de Lille I: 204 pp.

HENSEN V. (1887). Über die Bestimmung des Planktons oder des im Meere treibenden Materialen an Pflanzen und Tieren. *Ber. Komm. Wiss. Untersuch. Deutsch. Meere* 5: 1-109.

Hirata, T., Aiken, J., Hardman-Mountford, N. J., Smyth, T. J., et Barlow, R. G. (2008). An absorption model to derive phytoplankton size classes from satellite ocean colour. *Remote Sensing of Environment*, 112 : 3153-3159.

Hu, C., Wang, Y., Yang, Q., He, S., Hu, L., et He, M. (2003). Comparison of ocean color data products from MERIS, MODIS, and SeaWiFS: Preliminary Results for the East China seas.

Hoffmann, L.J., Peeken, I., et Lochte, K. (2008). Iron, silicate, and light co-limitation of three Southern Ocean diatom species. *Polar Biology*, 31 : 1067–1080.

Houliez E., Lizon F., Thyssen M., Artigas L.F., et Schmitt F.G. (2012). Spectral fluorometric characterization of Haptophyte dynamics using the FluoroProbe : an application in the Eastern English Channel for monitoring *Phaeocystis globosa*. *Journal of Plankton Research*, 34 : 136-151.

Huot, Y., Babin, M., Bruyant, F., Grob, C., Twardowski, M.S., et Claustre, H. (2007). Does chlorophyll a provide the best index of phytoplankton biomass for primary productivity studies? *Biogeosciences*, 4 : 853-868.

Hutchins, D.A., Hare, C.E., Weaver, R.S., Zhang, Y., Firme, G.F., DiTullio, G.R., Alm, M.B., Riseman, S.F., Maucher, J.M., Geesey, M.E., Trick, C.G., Smith, G.J., Rue, E.L., Conn, J., et Bruland, K.W. (2002). Phytoplankton iron limitation in the Humboldt Current and Peru Upwelling, *Limnology and Oceanography*, 47 : 997-1011.

IOCCG. (2007). Ocean-Colour Data Merging. Reports of the international ocean-colour coordinating group, no. 6 (pp. 68). Darmouth, Canada: IOCCG.

IOCCG. (2010). Atmospheric correction for remotely-sensed ocean-colour products. In M. H. Wang (Ed.), Reports of the international ocean-colour coordinating group, no. 10 (pp. 78). Darmouth, Canada: IOCCG.

Jackson, T., Bouman, H.A., Sathyendranath, S., et Devred, E. (2011). Regional-scale changes in diatom distribution in the Humboldt upwelling system as revealed by remote sensing : implications for fisheries. *ICES Journal of Marine Science*, 68 : 729-736.

Jamet, C., Loisel, H., Kuchinke, C.P., Ruddick, K., Zibordi, G., et Feng, H. (2011). Comparison of three atmospheric correction algorithms for turbid waters using AERONET-OC measurements, *Remote Sensing of Environment*, 115 : 1955-1965.

Jassby, A. D. et Platt, R. (1976). Mathematical formulation of the relationship between photosynthesis and light for phytoplankton. *Limnology and Oceanography*, 21 : 540-541.

Jeffrey, S. W. et Vesk, M. (1997). Introduction to marine phytoplankton and their pigment signatures. *In: Phytoplankton pigment in oceanography: Guidelines to modern methods*, S. W. Jeffrey, R. F. C. Mantoura et S. W. Wright (eds), UNESCO: 33-84.

Jochem, F. J., Mathot, S., et Queguiner, B. (1995). Size fractionated primary production in the open Southern Ocean in austral spring. *Polar Biology*, 15 : 381-392.

Jouini, M., Lévy, M., Crépon, M., et Thiria, S. (2013). Reconstruction of satellite chlorophyll images under heavy cloud coverage using a neural classification method. *Remote Sensing of Environment*, 131 : 232-246.

Jeffrey, S. W., Mantoura, R.F.C et Wright, S.W. (1997). Phytoplankton pigments in oceanography, UNESCO, 661 pp.

Kjørboe, T., Tiselius, P., Mitchell-Innes, B., Hansen, J.L.S., Visser, A.W., et Mari, X. (1998). Intensive aggregate formation with low vertical flux during an upwelling-induced diatom bloom. *Limnology and Oceanography*, 43 : 104-116.

Kobayashi, F., et Takahashi, K. (2002). Distribution of diatoms along the equatorial transect in the western and central Pacific during the 1999 La Niña conditions. *Deep-Sea Research II*, 49 : 2801-2821.

Kohonen, T. (1984). *Self-organization and associative memory*. (2nd ed.). Berlin, Heidelberg: Springer-Verlag.

Kohonen, T. (1988). *Self organizing and associative memory* : Springer-Verlag, Berlin.

Kohonen, T. (1995). *Self-organizing maps*. Berlin, Heidelberg: Springer.

Kohonen, T. (1997). *Self-organizing maps*. Berlin: Springer.

Kohonen, T. (2001). *Self-Organizing Maps*. Third, extended edition : Springer.

Kohonen, T. (2013). Essentials of the self-organizing map. *Neural Networks*, 37 : 52-65.

Korb, R.E., Whitehouse, M.J., et Ward, P. (2004). SeaWiFS in the southern ocean: spatial and temporal variability in phytoplankton biomass around South Georgia. *Deep-Sea Research II*, 51 : 99-116.

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, doi:10.1029/2009jc005303.

Kostecki, C., et Le Pape, O. (2011). *Analyse de l'effet de différentes pressions de nature anthropique sur les populations de soles dans les eaux sous juridiction française de la sous région marine Manche - Mer du Nord*. Rapport scientifique. Les publications du Pôle halieutique AGROCAMPUS OUEST n°5, 9 p.

Kruse, S., Jansen, S., Krägersky, S. et Bathman U. (2009). Gut content analysis of three dominant Antarctic copepod species during an induced phytoplankton bloom EIFEX (European iron fertilization experiment). *Marine Ecology*, 30 : 301-312.

Kuchinke, C. P., Gordon, H. R., et Franz, B. A. (2009). Spectral optimization for constituent retrieval in Case II waters I: Implementation and performance. *Remote Sensing of Environment*, 13 : 571-587.

Lacroix, G., Ruddick, K., Gypens, N., Lancelot, C. (2007). Modelling the relative impact of rivers (Scheldt/Rhine/Seine) and Western Channel waters on the nutrient and diatoms/Phaeocystis distributions in Belgian waters (Southern North Sea). *Continental Shelf Research*, 27: 1422-1446.

Lampert, L. (2001). Dynamique saisonnière et variabilité pigmentaire des populations phytoplanctoniques dans l'Atlantique Nord (golfe de Gascogne), *Ph.D. Thesis, Univ. Bretagne Occidentale*, 294 pp.

Lancelot, C., et Mathot, S. (1987). Dynamics of a Phaeocystis-dominated spring bloom in Belgian coastal waters. 1. Phytoplankton activities and related parameters. *Marine Ecology Progress series*, 37 : 239-248.

Lancelot, C., Keller, M.D., Rousseau, V., Smith, W.O., et Mathot, S. (1998). Autecology of the marine haptophyte Phaeocystis sp. In: Anderson DM, Cembella AD, Hallegraeff GM (Eds.), *Physiological Ecology of Harmful Algal Blooms*, NATO ASI series, vol G41. Springer-Verlag, Berlin: 209-224.

Landsberg, J.H. (2002) The effects of harmful algal blooms on aquatic organisms. *Reviews in Fisheries Science*, 10: 113-390.

Le Quéré, C., Harrison, S.P., Prentice, I.C., Buitenhuis, E.T., Aumont, O., Bopp, L., Claustre, H., Cotrim da Cunha, L., Geider, R.J., Giraud, X., Klaas, C., Kohfeld, K.E., Legendre, L., Manizza, M., Platt, T., Rivkin, R.B., Sathyendranath, S., Uitz, J., Watson, A.J., et Wolf-Gladrow, D. (2005). Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. *Global Change Biology*, 11: 2016-2040.

Leblanc, K., Arístegui, R., Armand, L., Assmy, P., Beker, B., Bode, A., Breton, E., Cornet, V., Gibson, J., Gosselin, M.-P., Kopczynska, E.E., Marshall, H.G., Peloquin, J.M., Piontkovski, S., Poulton, A.J., Quéguiner, B., Schiebel, R., Shipe, R., Stefels, J., Van Leeuwe, M.A., Varela, M., Widdicombe, C.E., Yallop, M. (2012): A global diatom database - abundance, biovolume and biomass in the world ocean. *Earth System Science Data*, 4 : 149-165.

Lee, Z., Arnone, R., Hu, C., Werdell, P.J., et Lubac, B. (2010). Uncertainties of optical parameters and their propagations in an analytical ocean color inversion algorithm. *Applied Optics*, 49 : 369-381.

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

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: 739-752.

Longhurst, A. R., Sathyendranath, S., Platt, T., et Caverhill, C.M. (1995). An estimate of global primary production in the ocean from satellite radiometer data. *Journal of Plankton Research*, 17: 1245-1271.

Longhurst, A. (2007). *Ecological Geography of the Sea*. (2nd ed). San Diego, California: Academic press.

Lubac, B., Loisel, H., Guiselin, N., Astoreca, R., Artigas, L.F et Mériaux, X. (2008). Hyperspectral and multispectral ocean color inversions to detect *Phaeocystis globosa* blooms in coastal waters. *Journal of Geophysical Research*, 113, C06026, doi: 10.1029/2007JC004451.

Mackey, M.D., Higgins, H.W., et Wright, S.W. (1996). CHEMTAX - A program for estimating class abundances from chemical markers: application to HPLC measurements of phytoplankton. *Ecology Progress Series*, 144 : 265-283.

Mackey D.J., Blanchot, J., Higgins, H.W., et Neveux, J. (2002). Phytoplankton abundances and community structure in the equatorial Pacific. *Deep Sea Research II*, 49 :2561-2582.

Marty, J. C., et Chiavérini, J. (2002). Seasonal and interannual variations in phytoplankton production at DYFAMED time-series station, northwestern Mediterranean Sea. *Deep-Sea Research II*, 49 : 2017-2030.

Masotti, I., Moulin, C., Alvain, S., Bopp, L., Tagliabue, A., et Antoine, D., (2011). Large-scale shifts in phytoplankton groups in the Equatorial Pacific during ENSO cycles. *Biogeosciences*, 8 : 539-550.

Loyola, D.G.R. (2006). Applications of neural network methods to the processing of earth observation satellite data. *Neural Networks*, 19 : 168-177.

Malin, G., Turner, S.M., et Liss, P.S. (1997). Sulfur: The plankton/climate connection. *Journal of Phycology*, 28 : 590–597.

Mantoura, R.F.C., Wright, S.W., Jeffrey, S.W., Barlow, R.G., et Cummings, D.E., (1997): Filtration and storage of pigments from microalgae : 283–305. *In* : *Phytoplankton Pigments in Oceanography*, ed. by S. W. Jeffrey, R. F. C. Mantoura and S.W. Wright, Monographs on oceanographic methodology, UNESCO, Paris.

Maritonera, S., Siegel, D. A., et Peterson, A. R. (2002). Optimization of semianalytical ocean color model for global scale applications. *Applied Optics*, 41 : 2705–2714.

Martinez E., Antoine, D., D'Ortenzio, F., et Gentili, B. (2009). Climate-driven basin-scale decadal oscillations of oceanic phytoplankton, *Science*, 36 : 1253-1256.

Martinez, E., Antoine, D., D'Ortenzio, F., et De Boyer Montégut, C. (2011). Phytoplankton spring and fall blooms in the North Atlantic in the 1980s and 2000s, *Journal of Geophysical research*, 116, C11029, doi:10.1029/2010JC006836.

McCulloch, W.S., et Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5 : 115-133.

Meister, G., Franz, B.A., Kwiatkowska, E.J., et McClain, C.R. (2012). Corrections to the Calibration of MODIS Aqua Ocean Color Bands Derived From SeaWiFS Data. *IEEE Transactions on Geoscience and Remote Sensing*, 50 : 310-318.

Mélin, F. (2011). Comparison of SeaWiFS and MODIS time series of inherent optical properties for the Adriatic Sea. *Ocean Science*, 6 : 61-76.

Mélin, F., et Zibordi, G. (2007). Optically Based Technique for Producing Merged Spectra of Water-leaving Radiances from Ocean Color Remote Sensing. *Applied Optics*, 46 : 3856-3869.

- Mélin**, F., Zibordi, G., et Djavidnia, S. (2007). Development and Validation of a Technique for Merging Satellite Derived Aerosol Optical Depth from SeaWiFS and MODIS. *Remote Sensing of Environment*, 108 : 436-450.
- Mélin**, F., Vantrepotte, V., Clerici, M., D'Alimonte, D., Zibordi, G., Berthon, J.-F., et Canuti, E. (2011). Multi-sensor satellite time series of optical properties and chlorophyll-*a* concentration in the Adriatic Sea. *Progress in Oceanography*, 91 : 229-244.
- Mélin**, F., Zibordi, G., Carlund, T., Holben, B.N., Stefan, S. (2013). Validation of SeaWiFS and MODIS Aqua/Terra aerosol products in coastal regions of European marginal seas. *Oceanologia* : doi:10.5697/oc.55-1.027.
- Merkl**, D. (1999). *Document Classification with Self-Organizing Maps*, springer.
- Metzl**, N., Brunet, C., Jabaud-Jan, A., Poisson, A., et Schauer, B. (2006). Summer and winter air-sea CO₂ fluxes in the Southern Ocean. *Deep-Sea Research I*, 53, 1548-1563.
- Michaels**, A. F., et Silver, M.W. (1988). Primary Production, sinking fluxes and the microbial food web. *Deep Sea Research I: Oceanographic Research Papers*, 35: 473-490.
- Mobley**, C.D. (1994). *Light and Water: Radiative Transfer in Natural Waters*. Academic Press, New York, 592 pp.
- Mobley**, C.D., Stramski, D., Bissett, W.P., et Boss, E. (2004). Optical modeling of ocean waters: Is the Case 1 - Case 2 classification still useful? *Oceanography*, 17 : 60-67.
- Mohan**, R., Quarshi, A.A., Thamban, M., Sudhakar, M. (2011). Diatoms from the surface waters of the Southern Ocean during the austral summer of 2004. *Current Science*, 100 : 1323-1327.
- Moreira**, J., et Costa, L.F. (1996). *Neural-based color image segmentation and classification using self-organizing maps*, Universidade Federal de São Carlos.
- Morel**, A., and L. Prieur. (1977). Analysis of variations in ocean color. *Limnology and Oceanography*, 22(4) : 709-722.
- Morel**, A., et Maritorena, S. (2001). Bio-optical properties of oceanic waters: A reappraisal. *Journal of Geophysical Research*, 106 : 7163–7180.
- Morel**, A. (1988). Optical modeling of the upper ocean in relation to its biogenous matter content (case 1 water), *Journal of Geophysical Research*, 93, 10749-10768.

Morel, A. (2009). Are the empirical relationships describing the bio-optical properties of case 1 waters consistent and internally compatible ? *Journal of Geophysical Research*, 114: doi: 10.1029/2008JC004803.

Morel, A., et Gentili, B. (1996). Diffuse reflectance of oceanic waters. 3. Implication of bidirectionality for the remotesensing problem. *Applied Optics*, 35 : 4850-4862.

Nair, A., Sathyendranath, S., Platt, T., Morales, J., Stuart, V., Forget, M. -H., et al. (2008). Remote sensing of phytoplankton functional types. *Remote Sensing of Environment*, 112(8) : 3366-3375.

Nejstgaard, J.C., Tang, K.W., Steinke, M., Dutz, J., Koski, M., Antajan, E., et Long, J. (2007). Zooplankton grazing on Phaeocystis: a quantitative review and future challenges. *Biogeochemistry*, 83: 137-145.

Nelson, D.M., et Brzezinski, M.A. (1997). Diatom growth and productivity in an oligotrophic midocean gyre : A 3-yr record from the Sargasso Sea near Bermuda. *Limnology and Oceanography*, 42 : 473-486.

Neukermans, G., Loisel, H., Mériaux, X., Astoreca, R. et McKee, D. (2012). In situ variability of mass-specific beam attenuation and backscattering of marine particles with respect to particle size, density, and composition. *Limnology and Oceanography*, 57: 24-144.

Niang, A., Gross, L., Thiria, S., Badran, F., et Moulin, C. (2003). Automatic neural classification of ocean colour reflectance spectra at the top of the atmosphere with introduction of expert knowledge. *Remote Sensing of Environment*, 86 : 257-271.

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 : 82-94.

Nordkvist, K., Loisel, H., et Duforêt-Gaurier, L. (2009). Cloud masking of SeaWiFS images over coastal waters using spectral variability. *Optics Express*, 17 : 12246–12258.

O'Reilly, J. E., Maritonera, S., Michell, 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 : 24937-24953.

O'Reilly, J. E., et al. (2000), Ocean color chlorophyll-a algorithms for SeaWiFS, OC2 and OC4, *In* : SeaWiFS Postlaunch Calibration and Validation Analyses: Part 3, SeaWiFS Postlaunch Tech. Rep. Ser., vol. 11, version 4, edited by S. B. Hooker and E. R. Firestone : 9-23, NASA Goddard Space Flight Cent., Greenbelt, Md.

Pan, X., Mannino, A., Russ, M. E., Hooker, S. B., et Harding, L. W. (2010). Remote sensing of phytoplankton pigment distribution in the United States northeast coast. *Remote Sensing of Environment*, 114 : 2403-2416.

Pan, X., Mannino, A., Marshall, H.G., Filippino, K.C., et Mulholland, M.R. (2011). Remote sensing of phytoplankton community composition along the northeast coast of the United States. *Remote Sensing of Environment*, 115 : 3731-3747.

Pan, Y., Tang, D., et Weng, D. (2010). Evaluation of the SeaWiFS and MODIS Chlorophyll a Algorithms Used for the Northern South China Sea during the Summer Season. *Terrestrial Atmospheric and Oceanic Sciences*, 21 : 997-1005.

Partensky, F., Hess, W.R., et Vaultot, D. (1999). Prochlorococcus, a marine photosynthetic Prokaryote of global significance. *Microbiology and Molecular Biology Reviews*, 63 : 106-127.

Peeken, I. (1997). Photosynthetic pigment fingerprints as indicators of phytoplankton biomass and development in different water masses of the Southern Ocean during austral spring. *Deep Sea Research II*, 44, 261-282.

Platzman, G. (1968). The Rossby wave. *Quarterly Journal of The Royal Meteorological Society*, 94 : 94-248.

Raitsos, D. E., Lavender, S. J., Maravelias, C. D., Haralambous, J., Richardson, A. J., et Reid, P. C. (2008). Identifying four phytoplankton functional types from space: An ecological approach. *Limnology and Oceanography*, 53 : 605-613.

Raven, J. A. (1998). Small is beautiful. *Functional Ecology*, 12: 503-513.

Reid, P.C., Colebrook, J.M., Matthews, J.B.L., et Aiken, J. (2003). The Continuous Plankton Recorder: concepts and history, from plankton indicator to undulating recorders. *Progress in Oceanography*, 58: 117-175.

Richardson, A. J., Pfaff, M. C., Field, J. G., Silulwane, N. F., et Shillington, F. A. (2002). Identifying characteristic chlorophyll *a* profiles in the coastal domain using an artificial neural network. *Journal of Plankton Research*, 24 : 1289-1303.

Richardson, A.J., Risien, C., et Shillington, F.A. (2003). Using self-organizing maps to identify patterns in satellite imagery. *Progress in Oceanography*, 59 : 223, 239.

Rosenblatt, F. (1958). The Perceptron : probabilistic model for information storage and organization in the brain. *Psychological Review*, 65: 386-408.

- Romero, O.,** Boeckel, B., Donner, B., Lavik, G., Fisher, G., et Wefer, G. (2002). Seasonal productivity dynamics in the pelagic central Benguela System inferred from the flux of carbonate and silicate organisms. *Journal of Marine Systems*, 37 : 259-278.
- Romero, O.E.,** Armand, L.K., Crosta, X., et Pichon, J.-J. (2006). The biogeography of major diatom taxa in Southern Ocean surface sediments: 3. Tropical/Subtropical species. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 223 : 49-65.
- Rossby, C.-G.** (1939). Relation between variations in the intensity of the zonal circulation of the atmosphere and the displacements of the semi-permanent centers of action. *Journal of Marine Research*, 1939 : 38-55.
- Rousseau, V.,** Vaultot, D., Casotti, R., Carcou, V., Lenz, J., Gunkel, J., et Baumann, M. (1994). Phaeocystis (Prymnesiophyceae) life cycle: evidence and hypotheses. In: Lancelot C, Wassmann P (Eds.), Ecology of Phaeocystis-dominated ecosystems. *Journal of Marine Systems*, 55: 23-40.
- Rousseau, V.,** Leynaert, A., Daoud, N., et Lancelot, C. (2002). Diatom succession, silicification and availability in Belgian coastal waters (southern North Sea). *Marine Ecology Progress Series*, 236: 61-73.
- Rousseau, V.,** Lancelot, C., et Cox, D. (2006). *Current Status of Eutrophication in the Belgian Coastal Zone*. Brussels: Presses Universitaires de Bruxelles.
- Ruddick, K. G.,** Ovidio, F., et Rijkeboer, M. (2000). Atmospheric correction of SeaWiFS imagery for turbid coastal and inland waters. *Applied Optics*, 39 : 897-912.
- Ryan, J.P.,** Polito, P.S., Strutton, P.G., et Chavez, F.P. (2002). Unusual large-scale phytoplankton blooms in the equatorial Pacific. *Progress in Oceanography*, 55 : 263–285.
- Salomon, J.-C** et Breton, M. (1991). Courants de marée et courants résiduels dans la Manche. *Oceanologica Acta*, 11: 47-53.
- Salomon, J.-C.,** et Breton, M. (1993a). An atlas of long-term currents in the Channel. *Oceanologica Acta*, 16 :439-448.
- Salomon, J.-C.,** et Breton, M. (1993b). Computed residual flow through the Dover Strait. *Oceanologica Acta*, 16 : 449-455.
- Sathyendranath, S.,** Stuart, V., Cota, G., Maas, H., et Platt, T. (2001). Remote sensing of phytoplankton pigments: A comparison of empirical and theoretical approaches. *International Journal of Remote Sensing*, 22 : 249-273.

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 Ecological Progress Series*, 272 : 59-68.

Schapira, M., Vincent, D., Seuront, L., et Gentilhomme, V. (2008). Temporal patterns of phytoplankton assemblages, size spectra and diversity during the wane of a *Phaeocystis globosa* spring bloom in hydrologically contrasted coastal waters. *Journal of the Marine Biological Association of United Kingdom*, 88: 649-662.

Schölkopf, B., et Smola, A.J. (2002). *Learning With Kernels: Support Vector Machines, Regularization, Optimization and Beyond*, MIT Press.

Schoemann, V., Becquevort, S., Stefels, J., Rousseau, V., Lancelot, C. (2005). *Phaeocystis* blooms in the global ocean and their controlling mechanisms: a review. *Journal of Sea Research*, 53: 43-66.

Self, K.E., Landry, M.R., Allen, C.B., Calbet, A., Christensen, S., Bidigare, R.R. (2001). Microbial community composition and growth dynamics in the Antarctic Polar Front and seasonal ice zone during late spring 1997. *Deep Sea Research II*, 48 : 4059-4080.

Sellner, K.G., Doucette, G.J., et Kirkpatrick, G.J. (2003). Harmful algal blooms: Causes, impacts and detection. *Journal of Industrial Microbiology and Biotechnology*, 30: 383-406.

Seuront L., Vincent, D. (2008). Increased seawater viscosity, *Phaeocystis globosa* spring bloom and *Temora longicornis* feeding and swimming behaviours. *Marine Ecology Progress Series*. 363: 131-145.

Shawe-Taylor, J., et Cristianini, N. (2000). *Support Vector Machines and other kernel-based learning methods*, Cambridge University Press.

Shukla, S.K., Crosta, X., Cortese, G., et Nayak, G.N. (2013). Climate mediated size variability of diatom *Fragilariopsis kerguelensis* in the Southern Ocean. *Quaternary Science Reviews*, 69 : 49-58.

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: 1256-1263.

Siegel, D.A., Wang, M.H., et Maritorena, S., et al. (2000). Atmospheric correction of satellite ocean color imagery: the black pixel assumption. *Applied Optics*, 39: 3582-3591.

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., et Yoder, J.A. (2013). Regional to global assessments of phytoplankton dynamics from the SeaWiFS mission, *Remote Sensing of Environment*, 135 : 77-91.

Silulwane, N. F., Richardson, A. J., Shillington, F. A., et Mitchell-Innes, B. A. (2001). Identification and classification of vertical chlorophyll patterns in the Benguela upwelling system and Angola-Benguela Front using an artificial neural network, In A. I. L. Payne, S. C. Pillar, et R. J. M. Crawford (eds.), A Decade of Namibian Fisheries Science. *South African Journal of Marine Science*, 23 : 37-51.

Sinistro, R. (2010). Top-down and bottom-up regulation of planktonic communities in a warm temperate wetland. *Journal of Plankton Research*, 32 : 209-220.

Sladeczek, V. (1986). Diatoms as indicators of organic pollution. *Acta hydrochimica et hydrobiologica*, 14 : 555-566.

Smetacek, V. H., Scharek, R., et Nothig, E.M. (1990), Seasonal and regional variation in the pelagial and its relationships to the life cycle of Krill, in Antarctic Ecosystems. *Ecological Change and Conservation* : 103– 114.

Smetacek, V., Klaas, C., Menden-Deuer, S., et Rynearson, T.A. (2002). Mesoscale distribution of dominant diatom species relative to the hydrographical field along the Antarctic Polar Front. *Deep-Sea Research II*, 49 : 3835-3848.

Smith, R.C., et Baker, K.S. (1978). The bio-optical state of ocean waters and remote sensing. *Limnology and Oceanography*, 23 : 247-259.

Steinberg, D.K., Carlson, C.A., Bates, N.R., Johnson, R.J., Michaels, A.F., et Knap, A.H. (2001). Overview of the USJGOFS Bermuda Atlantic Time-series Study (BATS): A decade scale look at ocean biology and biogeochemistry. *Deep-Sea Research II*, 48 : 1405-1447.

Stramski D., Bricaud, A. et Morel, A. (2001). Modeling the inherent optical properties of the ocean based on the detailed composition of the planktonic community, *Applied Optics*, 40 : 2929-2945.

Stramski D., Boss, E., Bogucki, D., et Voss, K.J. (2004). The role of seawater constituents in light backscattering in the ocean, *Progress in Oceanography*, 61 : 27-56.

Strutton, P.G., et Chavez, F.P. (2000). Primary productivity in the equatorial Pacific during the 1997–1998 El Niño. *Journal of Geophysical Research*, 105 : 26089-26101.

Stumpf, R. P., Arnone, R. A., Gould, R. W., et Ransibrahmanakul, V. (2003). A partly coupled ocean–atmosphere model for retrieval of water-leaving radiance from SeaWiFS in coastal waters. NASA technical memorandum 2003-206892(22) (pp. 51–59). Greenbelt, MD: NASA Goddard Space Flight Center.

Subramaniam, A., Brown, C. W., Hood, R. R., Carpenter, E. J., et Capone, D. G. (2002). Detecting Trichodesmium blooms in SeaWiFS imagery. *Deep-Sea Research II*, 49 : 107-121.

Sudhakar, G., Jyothi, B., et Venkateswarlu, V. (1994). Rôle of diatoms as indicators of pollution gradients. *Environmental Monitoring and Assessment*, 33 : 85-99.

Thiria, S., Lechevalier, Y., Gascuel, O. et Canu, S. (1997). *Statistique et méthodes neuronales*. Éditions Dunod, Paris, 336 p.

Timmermans, K. R., Van der Woerd, H.J., Wernand, M.R., Sligting, M., Uitz, J., et De Baar, H.J.W. (2008), In situ and remote-sensed chlorophyll fluorescence as indicator of the physiological state of phytoplankton near the Isles Kerguelen (Southern Ocean), *Polar Biology*, 31(5) : 617-628.

Tomlinson, M.C., Stumpf, R.P., Ransibrahmanakul, V., Truby, E.W., Kirkpatrick, G.J., Pederson, B.A., Vargo, G.A., et Heil, C.A. (2004). Evaluation of the use of SeaWiFS imagery for detecting *Karenia brevis* harmful algal blooms in the eastern Gulf of Mexico, *Remote Sensing of Environment*, 91 : 293-303.

Tremblay, J. E., Lucas, M.I., Kattner, G., Pollard, R., Strass, V.H., Bathmann, U., et Bracher, A. (2002), Significance of the Polar Front zone for large-sized diatoms and new production during summer in the Atlantic sector of the Southern Ocean, *Deep Sea Research II*, 49 : 3793-3811.

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 (C08005) : 1-23.

Uitz, J., Stramski, D., Gentili, B., D'Ortenzio, F., et Claustre, H. (2012). Estimates of phytoplankton class-specific and total primary production in the Mediterranean Sea from satellite ocean color observations. *Global Biogeochemical Cycles*, 26, doi:10.1029/2011GB004055.

Vanormelingen, P., Verleyen, E. et Vyverman, W. (2008). The diversity and distribution of diatoms: from cosmopolitanism to narrow endemism. *Biodiversity and Conservation*, 17, 393-405.

Vantrepotte, V., Loisel, H., Dessailly, D et Mériaux, X. (2012). Optical classification of contrasted coastal waters. *Remote Sensing of Environment*, 123 : 306–323.

- Vapnik**, V. (1998). *Statistical Learning Theory*. Wiley-Interscience, New York.
- Veldhuis**, M.J.W., et Wassmann, P. (2005). Bloom dynamics and biological control of a high biomass HAB species in European coastal waters: a *Phaeocystis* case study. *Harmful Algae*, 4: 805-809.
- Vesanto**, J., Himberg, J., Alhoniemi, E., et Parhankangas, J. (1999). Self-organizing map in Matlab: the SOM Toolbox. *Proceedings of the Matlab DSP conference 1999* (pp. 35-40), Espoo, Finland.
- 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) : 19939-19956.
- Wang**, M.H., et Franz, B.A. (2000). Comparing the ocean color measurements between MOS and SeaWiFS: A vicarious intercalibration approach for MOS. *IEEE Transactions on Geoscience and Remote Sensing*, 38 :184-197.
- Warner**, A.J., et Hays, G.C., (1994). Sampling by the Continuous Plankton Recorder survey. *Progress in Oceanography*, 34: 237–256.
- Wassermann**, P.D. (1989). *Neural computing theory and practice*. New York: Von Nostrand Reinhold.
- Weisse**, T., Tande, K., Verity, P., Hansen, F., et Gieskes, W. (1994). Their trophic significance of *Phaeocystis* blooms. *Journal of Marine Systems*, 5: 67-79.
- Weber**, L.H., et El-Sayed, S. (1987). Contributions to the net, nano- and picoplankton standing crop and primary production in the Southern Ocean. *Journal of Plankton Research*, 9 : 973-994.
- Werdell**, P. J., et Bailey, S.W. (2005). An improved in-situ bio-optical data set for ocean color algorithm development and satellite data product validation. *Remote Sensing of Environment*, 98 : 122-140.
- Werdell**, P.J., Franz, B.A., Bailey, S.W., et McClain, C.R. (2006). Recent advances in the operational vicarious calibration of visible and near-infrared ocean color satellite radiometry. *Proc. Ocean Optics XVIII*, 14 pp.
- Werdell**, P.J., Bailey, S.W., Franz, B.A., Morel, A., et McClain, C.R. (2007). On-orbit vicarious calibration of ocean color sensors using an ocean surface reflectance model. *Applied Optics*, 46 : 5649-5666.

Whitmire, A.L., Scott Pegau, W., Karp-Boss, L., Boss, E., et Cowles, T.J. (2010). Spectral backscattering properties of marine phytoplankton cultures. *Optics Express*, 18 : 15073-15093.

Wright, S.W., Van den Enden, R.L., Pearce, I., Davidson, A.T., Scott, F.J., Westwood, K.J. (2010). Phytoplankton community structure and stocks in the Southern Ocean (30–80°E) determined by CHEMTAX analysis of HPLC pigment signatures. *Deep Sea Research II*, 57 : 758-778.

Zhang, C., Hu, C., Shang, S., Müller-Karger, F.E., Li, Y., Dai, M., Huang, B., Ning, X., et Hong, H. (2006). Bridging between SeaWiFS and MODIS for continuity of chlorophyll-a concentration assessments off Southeastern China. *Remote Sensing of Environment*, 102 : 250-263.

Zibordi, G., Mélin, F., Berthon, J.-F. (2006). Comparison of SeaWiFS, MODIS and MERIS Radiometric Products at a Coastal Site. *Geophysical Research Letters*, 33, doi : 10.1029/2006GL025778.

Zibordi, G., Mélin, F. (2007). Optically Based Technique for Producing Merged Spectra of Water-leaving Radiances from Ocean Color Remote Sensing. *Applied Optics*, 46 : 3856-3869.

Zielinski, U., et Gersonde, R. (1997). Diatom distribution in Southern Ocean surface sediments (Atlantic sector): Implications for paleoenvironmental reconstructions. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 129 : 213-250.

Zorn, S. R., Drewnick, F., Schott, M., Hoffmann, T., et Borrmann, S. (2008). Characterization of the 15 South Atlantic marine boundary layer aerosol using an aerodyne aerosol mass spectrometer. *Atmospheric Chemistry and Physics*, 8 : 4711-4728.

Zubkov, M. V., Sleigh, M.A., Burkill, P.H., et Leakey R.J.G. (2000). Picoplankton community structure on the Atlantic Meridional transect: A comparison between seasons. *Progress in Oceanography*, 45 : 369-386.

ANNEXES

Annexe 1 :

Automatic classification of water-leaving radiance anomalies from global SeaWiFS imagery: Application to the detection of phytoplankton groups in open ocean waters

Article initialement soumis le 29/09/2012 à Remote Sensing of Environment

Accepté le 30/09/2013



Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Automatic classification of water-leaving radiance anomalies from global SeaWiFS imagery: Application to the detection of phytoplankton groups in open ocean waters

Zied Ben Mustapha, Séverine Alvain^{*}, Cédric Jamet, Hubert Loisel, David Dessailly

LOG, Univ Lille Nord de France, ULCO, CNRS, F-62930 Wimereux, France

ARTICLE INFO

Article history:

Received 29 September 2012
 Received in revised form 25 August 2013
 Accepted 30 August 2013
 Available online xxx

Keywords:

Ocean color
 PHYSAT
 Phytoplankton
 Self-organizing maps

ABSTRACT

Remote sensing of ocean color is a powerful tool for monitoring phytoplankton in the ocean with a high spatial and temporal resolution. Several methods were developed in the past years for detecting phytoplankton functional types from satellite observations. In this paper, we present an automatic classification method based on a neural network clustering algorithm in order to classify the anomalies of water leaving radiance spectra (Ra), introduced in the PHYSAT method by Alvain et al. (2005), and analyze their variability at the global scale. The use of an unsupervised classification aims at improving the characterization of the spectral variability of Ra in terms of shape and amplitude as well as the expansion of its potential use to larger in situ datasets for global phytoplankton remote sensing. The Self-Organizing Map algorithm (SOM, Kohonen, 1984) aggregates similar spectra into a reduced set of pertinent groups, allowing the characterization of the Ra variability, which is known to be linked with phytoplankton community composition (Alvain et al., 2012).

Based on the same sample of Ra spectra, a comparison between the previous version of PHYSAT (Alvain et al., 2005, 2008) and the new one using SOM shows that it is now possible to take into consideration all the types of spectra. This was not possible with the previous approach, based on thresholds, defined in order to avoid overlaps between the spectral signatures of each phytoplankton group. The SOM-based method is relevant for characterizing a wide variety of Ra spectra through its ability to handle large amounts of data, in addition to its statistical reliability and processing speed compared to the previous PHYSAT. The former approach might have introduced potential biases and thus, its extension to larger databases was very restricted. This is not the case with the new statistical design presented in this work.

In the second step, some new Ra spectra have been related to phytoplankton groups using collocated field pigment inventories from a large in situ database. Phytoplankton groups were identified based on biomarker pigment ratio thresholds taken from the literature. SOM was then applied to the global daily SeaWiFS (Sea-viewing Wide Field-of-view Sensor) imagery archive between 1997 and 2010. Global distributions of major phytoplankton groups were analyzed and validated against in situ data. Thanks to its ability to capture a wide range of spectra and to manage a larger in situ pigment dataset, the neural network tool allows to classify a much higher number of pixels (2 times more) than the previous PHYSAT method for the five phytoplankton groups taken into account in this study (*Synechococcus*-like-cyanobacteria, diatoms, *Prochlorococcus*, Nanoecaryotes and *Phaeocystis*-like). In addition, different Ra spectral signatures have been associated to diatoms. These signatures are located in various environments where the inherent optical properties affecting the Ra spectra are likely to be significantly different. Local phenomena such as diatom blooms in the upwelling regions or during climatic events (i.e. *La Niña*) are more clearly visible with the new method.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

The main parameter derived from ocean color satellite observations is the chlorophyll-a concentration (chl-a). It is commonly used as a proxy of

the phytoplankton concentration in surface waters. However, chl-a provides a limited information about phytoplankton groups present in the surface waters since this photosynthetic pigment is present in all groups. Phytoplankton groups play different roles in the various biogeochemical cycles. For instance, diatoms are responsible for about 20% of global carbon fixation (Nelson, Treguer, Brzezinski, Leynaert, & Queguiner, 1995) and are major contributors to the biogeochemical cycling of silicon (Falcatore, d'Alcalá, Croot, & Bowler, 2000) whereas prymnesiophytes, such as *Phaeocystis*, produce atmospheric Dimethyl Sulfide (DMS), a

^{*} Corresponding author at: Laboratoire d'Océanologie et Géosciences, LOG UMR8187, Université du Littoral Côte d'Opale, Université de Lille 1, CNRS, 32 Avenue du Maréchal Foch, 62930 Wimereux, France. Tel.: +33 3 21 99 64 21.

E-mail address: severine.alvain@univ-littoral.fr (S. Alvain).

compound that impacts the climate as a precursor of cloud condensation nuclei (Belviso, Moulin, Bopp, & Stefels, 2004). An accurate and adaptive detection of phytoplankton groups from space is thus a major challenge in ocean color and for better understanding the biogeochemical cycles.

Based on different approaches, several algorithms are now able to detect phytoplankton groups from space, such as coccolithophorids (e.g. Ackleson, Balch, & Holligan, 1994; Brown & Podesta, 1997; Brown & Yoder, 1994), diatoms (e.g. Sathyendranath et al., 2004), the cyanobacteria *Synechococcus* (Morel, 1997), N₂-fixing cyanobacteria *Trichodesmium* (Dupouy et al., 2008; Subramaniam, Brown, Hood, Carpenter, & Capone, 2002) and *Phaeocystis globosa* (Astoreca et al., 2009; Lubac et al., 2008). These spectral-response approaches (among others) rely on the covariation between some spectral features of the ocean optical properties and the dominant phytoplankton group.

Among the spectral response approaches developed during the past years, a method called “PHYSAT” was developed for detecting multiple phytoplankton groups (Alvain, Moulin, Dandonneau, & Breon, 2005) at the global scale. This algorithm empirically relates the anomalies of the SeaWiFS normalized water leaving radiances (nLw) with the presence of dominant phytoplankton groups. This anomaly, also called “Radiance Anomaly”, simply mentioned by Ra further in the manuscript, represents the second order variation of the normalized water-leaving radiances after removal of the first order effect of the chl-a variation (Alvain, Moulin, Dandonneau, & Loisel, 2008). A set of radiance anomaly spectra characteristics in terms of shapes and amplitudes was associated to specific phytoplankton groups (Nanoeucaryotes, *Prochlorococcus*, *Synechococcus-like* cyanobacteria (SLC) and diatoms) by using coincident in situ pigment inventories performed in the framework of the GeP&CO program (Geochemistry, Phytoplankton, and Color of the Ocean, Dandonneau et al., 2004). The phytoplankton group spectral signatures have been first defined empirically by matching a set of radiance anomaly spectra with 41 coincident pigment inventories. A set of criteria (thresholds) was defined in order to determine the spectral signatures for each phytoplankton group. These criteria were then used to compute maps of dominant groups. The thresholds have been roughly defined by taking the minimum and maximum values of the spectra associated to each phytoplankton group. Moreover, additional criteria were introduced on the radiance anomaly spectra in order to avoid any overlapping between the spectral signatures of each phytoplankton group. These are the main limitations of PHYSAT version 2008 (PHYSAT-v2008), making the method difficult to adapt to new satellite or large in situ data sets (i.e. from other sensors, new match-up) and prone to not consider specific shapes of spectra. Consequently, the possibility of future identification of other phytoplankton groups, mixture of several groups or even different stages of blooms (Alvain, Loisel, & Dessailly, 2012) becomes restricted.

This paper focuses on the development of an automatic classification approach that allows the discrimination between different radiance anomaly spectra computed from a large satellite dataset and the extraction of pertinent objective information embedded in these spectra, without any a priori knowledge (unsupervised classification). This new approach aims to better characterize the variability of the most frequently observed radiance anomalies at the global scale by using a neural network approach, the “Self-Organizing Maps” (SOM, Kohonen, 1984) to automatically classify their spectral diversity in terms of amplitude and shape. Self-Organizing Maps were already successfully applied in remote sensing for pattern extraction (Ehsani & Quiel, 2008; Gorricha & Lobo, 2012; Hu & Weng, 2009; Iskandar, 2010; Richardson, Risien, & Shillington, 2003), and more specifically to ocean color for aerosol characterization (Brajard et al., 2007; Niang, Badran, Moulin, Crépon, & Thiria, 2006; Niang, Gross, Thiria, Badran, & Moulin, 2003), radiance spectra classification (Ainsworth & Jones, 1999) as well as for phytoplankton absorption spectra analysis (Chazottes, Bricaud, Crépon, & Thiria, 2006; Chazottes, Crépon, Bricaud, Ras, & Thiria, 2007), among others.

This study aims at providing a more accurate and robust method in order to go further in phytoplankton group remote sensing than

PHYSAT-v2008 and most importantly, to significantly enhance the development of the former approach, which was restricted by the introduction of thresholds and shape criteria on the radiance anomaly spectra. The use of SOM gives the ability to automatically determine different classes of radiance anomaly spectra based on their similarities in terms of shape and amplitude. In addition, the method gives the advantage to take into account a larger in situ dataset to relate these spectra to phytoplankton groups in a timely manner, with a higher flexibility and reliability, which would be much more difficult to achieve with PHYSAT-v2008. The different classes of spectra determined by SOM will be related to dominant phytoplankton groups using in situ pigment inventories (labeling procedure) as done previously for the first PHYSAT version. This new approach should give a significant improvement to the existing method (PHYSAT-v2008) by enabling the identification of phytoplankton groups through a detailed characterization of the radiance anomalies from a large satellite dataset (enhanced global representativeness of characteristic radiance anomalies). SOM gives also the possibility to perform a detailed study of the global spectral variability and diversity of the radiance anomalies regardless of the labeling possibilities, related to the in situ data availability.

After a description of the different datasets and the algorithm used to perform the automatic classification approach, named PHYSAT-SOM, the performances of the new method and PHYSAT-v2008 are compared in terms of spectral information potential characterization, based on the set of 41 radiance anomaly spectra that were initially used to develop PHYSAT-v2008. Then, the improvements brought by the use of SOM and its ability to handle a larger in situ dataset are shown through a detailed study of the automatically classified spectra, a global climatology of the phytoplankton groups distribution (using the available global SeaWiFS archive), as well as statistics on the spatial coverage in comparison with PHYSAT-v2008.

2. The data sets

2.1. The PHYSAT-v2008 radiance anomalies

The radiance anomalies (Ra), as defined by Alvain et al. (2005) were used as inputs to develop our classification method. These Ra were computed from global SeaWiFS level 3 daily normalized water-leaving radiance (nLw) binned products provided by the NASA/GSFC/DAAC at 9 km resolution. The Ra represents the second order variation of nLw after removal of the first order effects of chl-a variation. It is computed at 412, 443, 490, 510 and 555 nm as follows:

$$Ra(\lambda) = nLw(\lambda)/nLw_{ref}(\lambda, chl-a) \quad (1)$$

where $nLw_{ref}(\lambda, chl-a)$ represents the average nLw spectrum for a given value of chl-a. The nLw_{ref} is computed for chl-a values within the following interval: 0.02–3 mg·m⁻³ with an increment of 0.1 mg·m⁻³. This reference can then be used to remove the first order effect of chl-a on nLw (λ) measurements, as shown in Eq. (1). Ra is, by definition, independent of the chl-a level and represents therefore the second order variation of nLw (λ) (Alvain et al., 2005).

The same validity criteria as Alvain et al. (2005) were used for the calculation of the Ra spectra: the aerosol optical thickness at 865 nm has to be lower than 0.15 in order to minimize the impact of atmospheric correction errors and of sub-pixel cloud contamination. Moreover, all pixels where chl-a is not in the range 0.04–3 mg·m⁻³ were discarded. This latter criterion is used in order to avoid ultra-oligotrophic waters where it is unlikely to find a dominant phytoplankton group in ocean color data and to exclude water masses potentially impacted by coastal matter. In order to reduce the effects of shallow sandy-bottom on the marine reflectance, pixels characterized by a depth lower than 10 m were removed using a bathymetric mask. Indeed, shallow sandy-bottom has a reflective effect on the sunlight and therefore could

increase artificially the values of the upwelling radiances (Mobley & Sundman, 2003; Mobley, Zhang, & Voss, 2003).

2.2. The pigment database

Inventories of phytoplankton pigments are commonly used to discriminate between phytoplankton groups (Barlow et al., 2002; Claustre, 1994; Mackey, Higgins, & Wright, 1996; Uitz, Claustre, Morel, & Hooker, 2006; Vidussi, Claustre, Manca, Luchetta, & Marty, 2001). Five biomarker pigments were considered to identify phytoplankton groups: *divinyl-chlorophyll-a*, 19'HF (19'hexanoxyfucoxanthine), *zeaxanthin*, *fucoxanthin* and *peridinin*. The approach based on biomarker pigment ratio thresholds described in Alvain et al., 2005 was used.

A set of Ra spectra collocated with 41 in situ pigment inventories, collected during the GeP&CO program cruises (Dandonneau et al., 2004) was used to compare the performances of PHYSAT-SOM and PHYSAT-v2008 approaches in terms of Ra spectra retrieval. This is the dataset that has been used to develop the original method (Alvain et al., 2005).

A larger in situ dataset, including samples collected between 1999 and 2009 during several cruises, has been used in the second part of the study in order to show the improvements brought by the automatic classification method, which gives the advantage to quickly process larger datasets. The pigment inventories used here gather data collected from the ICOTA (Coastal Ichthyology in Adelle Land, A. Goffart and D. Vincent, personal communication; Zorn, Drewnick, Schott, Hoffmann, & Borrmann, 2008), OOMPH (Organics over the Ocean Modifying Particles in the both Hemispheres, E. Peeken, personal communication), HALIFAX (E. Devred, personal communication), and the NOMAD dataset (the NASA bio-Optical Marine Algorithm Data set, Werdell & Bailey, 2005) in addition to the whole GeP&CO dataset. The geographic distribution of the 1068 coincident in situ and satellite measurements is displayed in Fig. 1. Although these coincident measurements represent only 12.7% of the total in situ database, they cover all the seasons and various areas of the world's ocean especially in the North Atlantic and the tropical Pacific thanks to the GeP&CO dataset, allowing taking into account larger spatial and temporal scales

as well as various environmental conditions. However, some areas are poorly represented such as the Mediterranean, the North Pacific and the Indian Ocean.

3. The clustering method: Self-Organizing Map algorithm

Using an automatic classification approach gives the ability of considering a large satellite data set in order to have the largest possible coverage of characteristic spectral anomalies. It is often difficult to extract interpretable patterns from satellite images, as the datasets are large, often non-linear and usually need a strong preprocessing phase, related to the distribution of the data and outlier removal. SOM, among other unsupervised classification methods, has a number of advantages over traditional statistical methods (Wassermann, 1989). It can solve non-linear problems of almost infinite complexity and is more robust in handling noisy data (Kohonen, 2013). This is especially desirable for satellite data. SOM is particularly adapted to pattern recognition and classification tasks from large datasets (Richardson et al., 2003) and thus, is particularly suitable for our study.

A Self-Organizing Map is a type of artificial neural network that is trained using an unsupervised learning to produce a low-dimensional (typically two-dimensional), discretized representation of the input data (Kohonen, 2013). This neural network approach provides the advantages of a fully automated method without any a priori knowledge on the data. SOM is different from other artificial neural networks in the sense that it uses a neighborhood function to preserve the topological properties of the input space. The first step of the classification is to summarize the information contained in the training data set by producing a set of reference vectors (**rvs**), which constitutes a compression of the information embedded in the data set (vector quantization in formal statistic). Each neuron of the SOM is associated with a particular reference vector and thus corresponds to a group of Ra spectra (see Appendix A). The different neurons of the topological map are connected (Fig. 2) and determine a topological (neighborhood) relationship among the different neurons (groups): close neurons on the map represent similar subsets of data (classes presenting similarities). In the present study, we dealt with a two-dimensional topological map with a

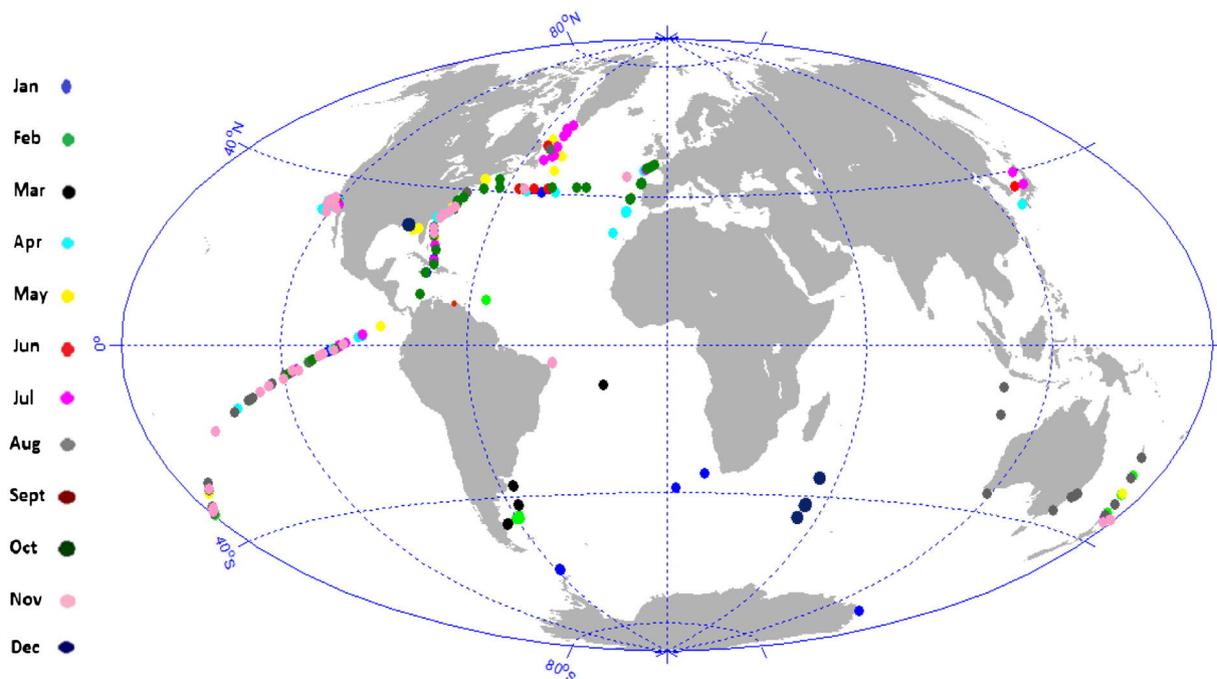


Fig. 1. Geographic distribution of the 1068 matchups, each month has been colored differently in order to show the seasonal variability covered by the in situ data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

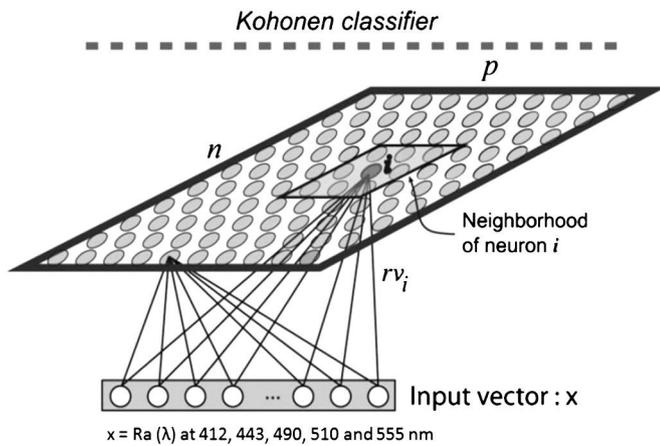


Fig. 2. Structure of the Self-Organizing Map. From Niang et al. (2003).

large number of neurons (10×10), providing a highly discriminated representation of the observations. The satellite Ra spectra are thus clustered into 100 groups. The choice of the shape and the number of neurons of the SOM have been established to minimize the quantization and topographic errors. The quantization error represents the average distance between each data vector (Ra spectrum) and its best matching unit (BMU) (closest reference vector in terms of Euclidian distance). The topographic error represents the proportion of all data vectors for which first and second BMUs are not adjacent units. It measures the topology preservation (Vesanto, Himberg, Alhoniemi, & Parhankangas, 1999). Several tests have been performed and the choice of a 10×10

SOM has proved to be the best to optimize these parameters and to have a detailed discrimination of the training dataset.

The training data set was made of 714,264 SeaWiFS Ra spectra at 412, 443, 490, 510 and 555 nm corresponding to a subsampling of one year satellite measurements with two days per month and one pixel over ten, with the previously established criteria in terms of oceanic and atmospheric conditions. The two days were homogeneously sampled at regular intervals of 15 days. This sampling was done so that the learning set is as representative as possible of the spatio-temporal variability of the Ra spectra. Note that this sampling was done for each year between 1997 and 2010. Each subset was then separately tested for the SOM training in order to see if there are significant differences in terms of shape and amplitude between the 100 reference vectors over the different years. Since it was not the case, the training phase was made with only one year of data (2002, the central year) in order to reduce the computing time and memory.

The data clustering we obtained by using 100 rv s gives a rational partition of the data set: each reference vector represents 7142 pixels on average, at most 14,245, and at least 2941. The rv s of the SOM are displayed in Fig. 3, which shows the topological order given by the SOM algorithm after the training phase. It is noted that the reference vectors which are close together in the two-dimensional topological map have similar spectra thanks to the neighborhood procedure. Each rv represents a synthesis of the variability of all the Ra spectra captured by each neuron.

The 10 two-dimensional scatter plots (with respect to our five-dimensional data space) representing the projection of the rv s ensemble are displayed in Fig. 4. We recall that the rv s represents a compression of the information contained in the initial Ra spectra data set. It is shown that Ra (490), Ra (510) and Ra (555) have similar distributions. However, they are significantly different from Ra

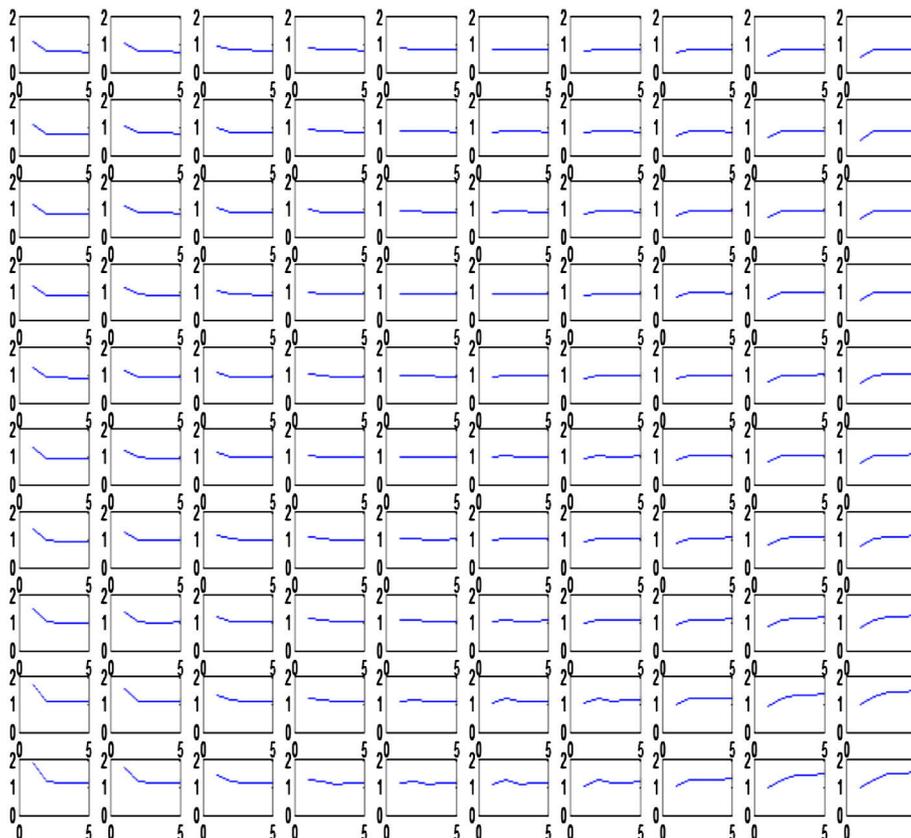


Fig. 3. Representation of the reference vectors corresponding to the Ra spectrum associated to each neuron, at the end of the training phase. For each box, the x axis represents the 5 wavelengths and the y axis represents Ra. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

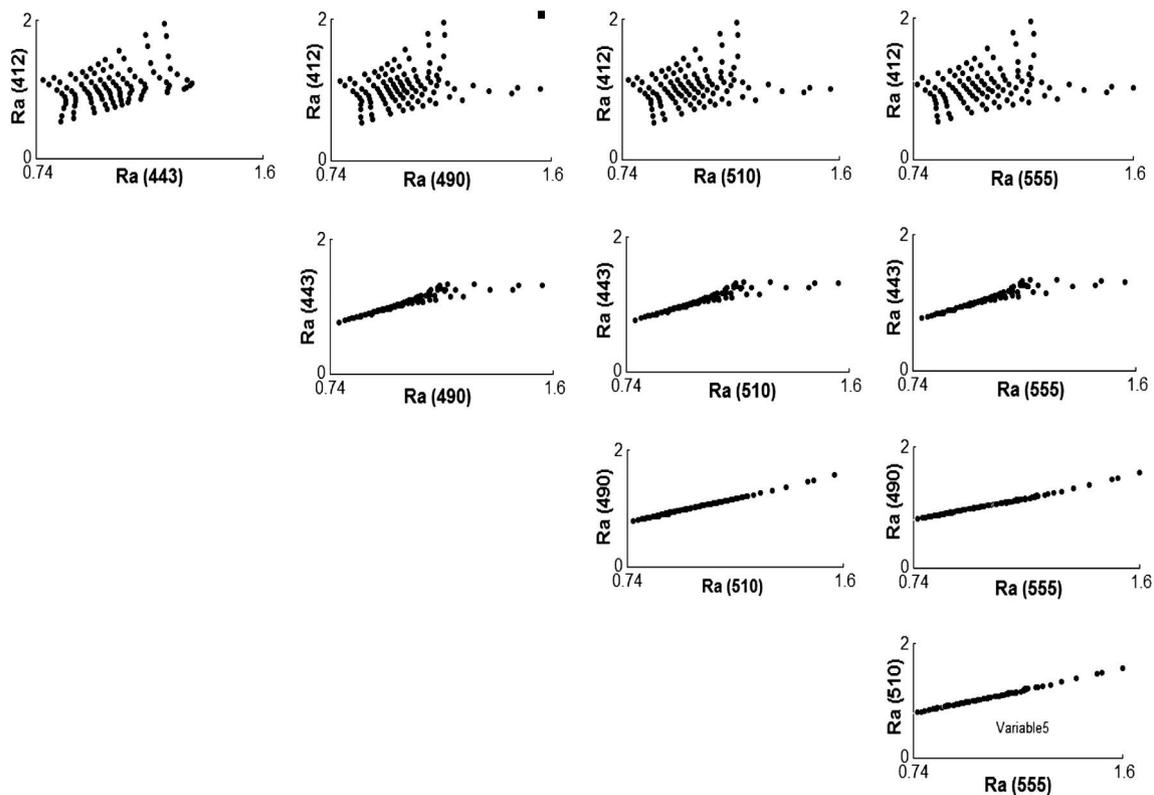


Fig. 4. The 10 two-dimensional scatter plots with respect to the data space: projection of the *rvs* ensemble.

(412) and Ra (443). This is also evidenced by the spectra displayed in Fig. 3, where we can see that, for a given spectrum, the highest Ra variations occur between the shortest wavelengths (412 and 443 nm) and the three other wavelengths (where the Ra remains relatively constant). The shape and amplitude of the Ra spectra mainly depend on the bio-optical environment, represented by different values of particulate scattering (b_p), absorption by the colored dissolved organic matter (a_{cdom}) and the absorption by the phytoplankton (a_{phy}), as it has been proven by the sensitivity study of Ra made by Alvain et al. (2012). As shown by the authors, the a_{cdom} strongly controls the spectral shape of Ra in the blue part of the spectrum. For low a_{cdom} values, the Ra spectra decrease from the blue to red wavelengths, while they increase with the wavelengths for high a_{cdom} values. For a_{phy} , the highest Ra sensitivity is found at 443 nm. The b_p has a restricted effect on the spectral shape but represents the main factor affecting the Ra absolute values at all the wavelengths.

These relationships between the five wavelengths show that we could have enough information about the Ra spectral variability by taking into account only one wavelength among 490, 510 and 555 nm. Nevertheless, in this study, the five wavelengths were taken into account so that the performance of SOM can be comparable to the previous approach of PHYSAT-v2008 (Section 4.1). Furthermore, Ra (490), Ra (510) and Ra (555) do not exhibit the same variability for some phytoplankton groups as it was demonstrated by the theoretical studies of Ra made by Alvain et al. (2012). Therefore, keeping the entire spectral information may be useful for future studies.

These above results show that the classification into 100 groups gives a balanced distribution of the input Ra spectra on the SOM as well as a high and precise discrimination of these spectra both in shape and amplitude. The second step of the classification, which is a supervised one, consists in making the link between the reference vectors and the phytoplankton groups using in situ data in order to decode the SeaWiFS images and to produce global phytoplankton group distribution maps.

4. Results

4.1. Improvement of Ra spectra retrieval with SOM

Before using the extended in situ pigment dataset (described in Section 2.2), the performance of SOM in terms of Ra spectra retrieval needs to be compared to the previous approach, based on the set of 41 Ra spectra used to develop PHYSAT-v2008.

This section aims at demonstrating the ability of SOM to cover a higher spectral variability with respect to the PHYSAT-v2008 approach. Forty-one GeP&CO pigment inventories collocated with satellite Ra spectra have been used by Alvain et al. (2005) to define the spectral signature of each phytoplankton group. However, the original PHYSAT approach was based on a basic characterization of these spectra. Thus, in order, to define a characteristic range of Ra values for each phytoplankton group, it was necessary to introduce several thresholds and shape criteria on the spectra. These criteria should be broad enough to maximize the number of Ra spectra classified in each range of values and narrow enough to avoid any overlapping between phytoplankton groups. Consequently, 21 out of 41 individual spectra were missed using these criteria (Fig. 5, top panel).

The set of 41 Ra spectra was projected on the trained SOM. Each spectrum was assigned to a specific neuron corresponding to the nearest reference vector, based on the Euclidean distance (see Appendix A). As shown in the bottom plots of Fig. 5, all the 41 spectra of the dataset were attached to a neuron and therefore classified by SOM. This means that, with SOM, it is possible to label more spectra from the same dataset that was used to develop the original method and so to account for the entire variability of the Ra spectra. This comparison demonstrates the advantage to use the neural network tool with respect to the initial approach. This advantage comes with many others such as the computing time-saving, the statistical reliability and the flexibility of the method, giving the possibility to easily work with much larger datasets.

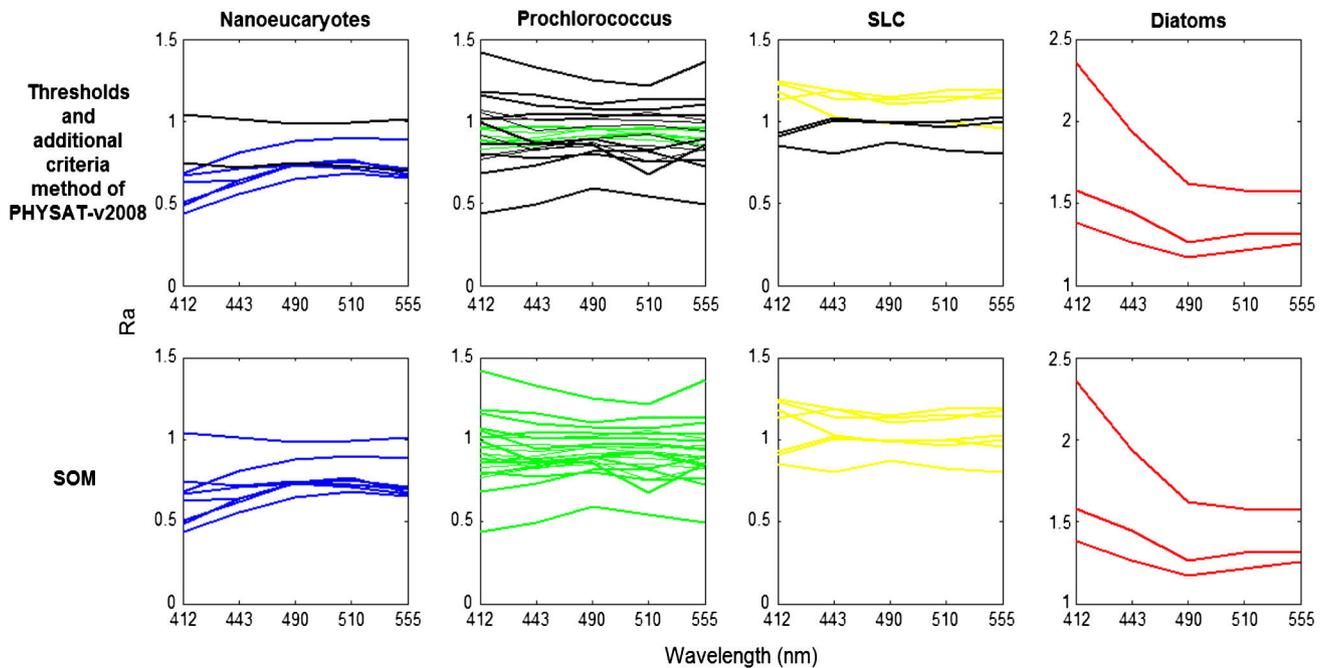


Fig. 5. Representation of the 41 Ra spectra used to develop the original PHYSAT method for each phytoplankton group. The individual Ra spectra unclassified by the thresholds and additional criteria approach used by PHYSAT-v2008 (top) are depicted by the black lines. Twenty-one out of 41 spectra are missed by the original approach while all the spectra are covered by the SOM approach (bottom).

4.2. Improvements of phytoplankton group detection with PHYSAT-SOM through the integration of a larger in situ database

The extended in situ pigment inventory dataset described in Section 2.2 has been used in this part of the study. The advantages and improvements brought by the ability of SOM to handle a larger dataset are demonstrated through a detailed analysis of the Ra spectra associated to each phytoplankton group as well as the visualization and analysis of a monthly climatology of phytoplankton groups over the 1997–2010 SeaWiFS archives, with an emphasis on fine scale spatial structure detection.

4.2.1. The labeling procedure

The labeling procedure consists in affecting a phytoplankton group to each class of Ra spectra obtained by SOM. For this purpose, the in situ pigment inventories were matched to SeaWiFS products at 9 km resolution over a 3×3 window. This match-up criterion, although less restricting than the usual 3-hour window for data and algorithm validation (Bailey & Werdell, 2006), was adopted to maximize the number of match-ups, as done in Alvain et al. (2008). Each collocated Ra spectrum was projected on the SOM, which assigned it to a specific neuron corresponding to the nearest \mathbf{rv} (reference vector), as explained in previous sections. After the projection, the Ra spectra are distributed among the 100 neurons of SOM. As a consequence, each neuron may correspond to a certain amount of spectra and may get:

- One or more spectra, all with the same label (case 1),
- More than one spectra with different labels (case 2),
- No spectra (case 3).

For case 1, the neuron takes the label of the affected spectra. For case 2, an additional step is needed: if at least 50% of the Ra spectra assigned to the \mathbf{rv} have the same label, the \mathbf{rv} takes this label; if not, it takes the label “no specific dominance”. For case 3, which is the rarest, the \mathbf{rv} s were not associated with any Ra spectrum. These \mathbf{rv} s likely correspond to spectra not represented in the match-up data set at this time. The corresponding neurons are not labeled but this could change as more in situ data becomes available in the future.

According to the thresholds on biomarkers used in this study (Alvain et al., 2005), four phytoplankton groups were successfully identified in 759 of 1068 collocated samples (i.e. 71% of the collocated samples): Nanoeucaryotes (216 samples: 28%), *Prochlorococcus* (213 samples: 28%), *Synechococcus-like* cyanobacteria (SLC) (278 samples: 37%) and diatoms (52 samples: 7%). For each phytoplankton group, 20% of the samples were randomly selected for the validation stage of the automatic classification. The 609 remaining samples (corresponding to 609 coincident Ra spectra) were used to allocate a phytoplankton group to each neuron of SOM.

The relative frequency of the Ra spectra captured by each neuron gives information about the proportion of in situ data associated to each one. This relative frequency is computed separately for each phytoplankton group by dividing, for each neuron, the number of spectra corresponding to a given phytoplankton group and affected to this neuron by the total number of spectra corresponding to this group. This method has been used in order to take into account the label rather than the absolute number of spectra, which significantly differs from one phytoplankton group to another. These differences may introduce a bias in the neurons labeling if the relative frequency is not computed separately for each phytoplankton group (for instance, SLC are associated to 37% of the in situ measurements, while diatoms correspond to only 7% of the data. This is due to the dominance of cyanobacteria over large spatial scales while diatoms represent a local blooming group). In order to take into account only neurons associated with a significant part of in situ pigment inventories, a given neuron is labeled only if the relative frequency of the Ra spectra captured by its \mathbf{rv} is equal or higher than 0.025 (which represents 3 spectra on average).

The relative frequency of the labeled Ra spectra associated to the 100 neurons of SOM is displayed in Fig. 6. Note that all the 609 coincident Ra spectra have been captured by at least one neuron. As we can see in this figure, 14% of the neurons have not captured spectra (case 3). The spectra associated to only one label (case 1) were captured by 12% of the neurons. The neurons whose reference vector has captured spectra with different labels (case 2) are the most frequent (74%). After applying all the criteria described in the labeling procedure, 44% of the neurons have been assigned either to one of the 4 phytoplankton groups, or to a “no specific

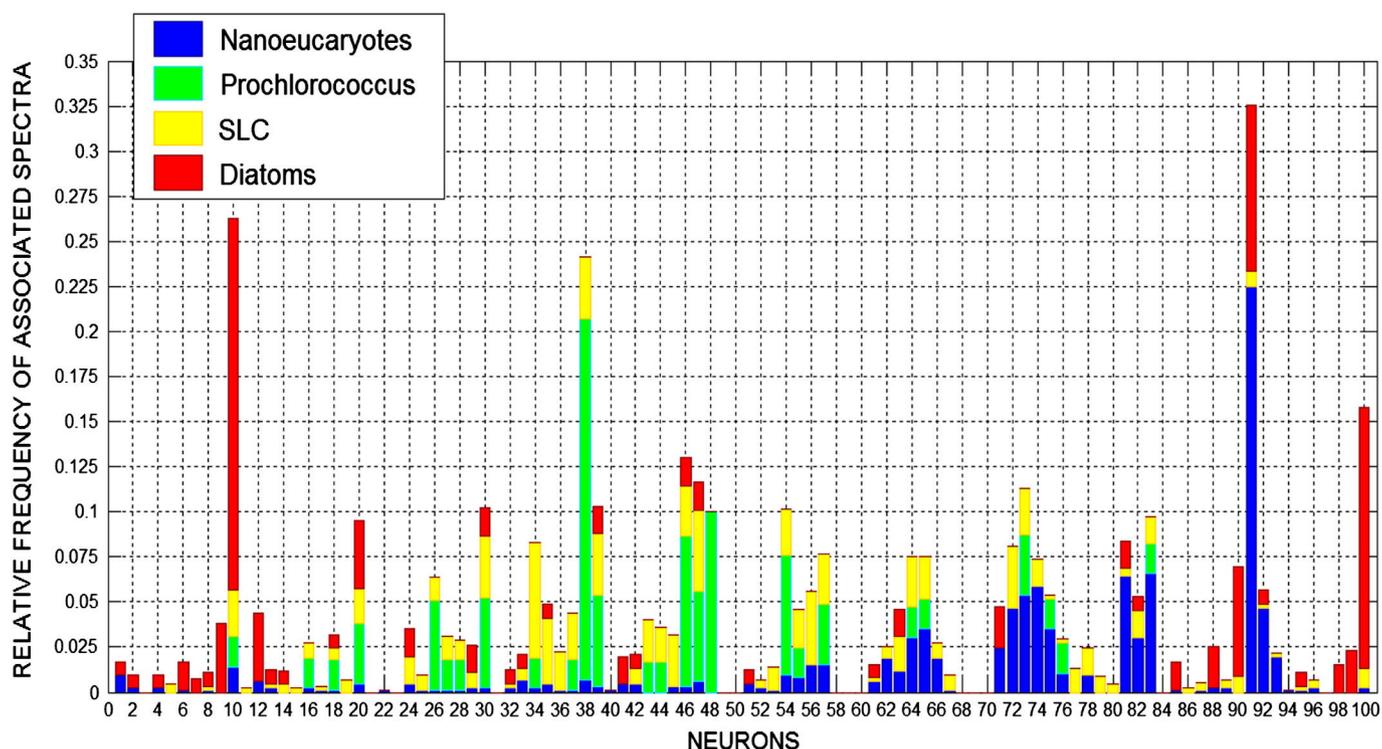


Fig. 6. Relative frequency of the labels associated with the Ra spectra for the 100 neurons highlighted by the SOM approach.

dominance". The 56% remaining belong either to the "case 3" neurons or have a relative frequency of captured Ra spectra lower than 0.025. In this study, these neurons were not labeled but they represent potentially exploitable spectral information and could be labeled as soon as additional in situ data become available. Five neurons are clearly identified as "diatoms" and eight neurons as "Nanoecaryotes". SLC are more often found in "no specific dominance" neurons, mixed with *Prochlorococcus*. Pure assemblages of SLC have been affected to six neurons. A good agreement between the distribution of the phytoplankton groups and the topological order imposed by the SOM on the radiance anomaly spectra is found, since adjacent neurons often characterize the same phytoplankton group (Fig. 6).

In addition to the four groups identified based on pigment inventories, the maps previously produced based on PHYSAT-v2008 (Alvain et al., 2008) showed *Phaeocystis-like* detection based on coherent specific spatial structures associated to this group based on literature. This additional group was included in the PHYSAT-SOM method. However, due to the absence of in situ pigment observations and specific biomarkers of *Phaeocystis*, specific Ra spectra were empirically identified. The Ra spectra corresponding in PHYSAT-v2008 to pixels identified as *Phaeocystis-like* in blooming conditions (January in the Southern Ocean) were projected on the trained SOM. These spectra have been captured by two adjacent neurons and labeled as *Phaeocystis-like*.

After carrying out the labeling of the neural network, each pixel of the daily Ra maps, computed from the nLw over the SeaWiFS archive, can be assigned to its nearest neuron (hence reference vector) of the SOM. Each captured pixel takes therefore the label of its nearest neuron in order to produce global daily phytoplankton group's maps. A certain discrepancy is possible between the referent vector of a neuron and the Ra spectra captured by this neuron since some spectra are poorly represented in the training data set. Therefore, only the spectra that are within a range of plus or minus one standard deviation from their corresponding reference vector were taken into account. This standard deviation was computed from the set of Ra spectra affected to each neuron at the end of the training phase. The standard deviation was considered separately for each wavelength.

4.2.2. Ra spectral signatures of the labeled phytoplankton groups

The spectral signatures of Ra corresponding to the four PHYSAT-SOM labeled classes are shown in Fig. 7, which represents the \mathbf{rv} s of the labeled neurons for the four phytoplankton groups identified from the extended in situ pigment inventories. Note that the *Phaeocystis-like* group was not included in this analysis since it was identified from PHYSAT-v2008 maps (based on literature).

The Nanoecaryote (Fig. 7a) label is characterized by Ra values lower than the unit at any wavelength, with a slight increase of Ra between 412 and 443 nm. The *Prochlorococcus* (Fig. 7b) and SLC groups (Fig. 7c) are associated to more flattened Ra spectra with minimal and maximal values around unit. The similarities between the Ra spectral signatures of *Prochlorococcus* and SLC can be explained by the close characteristics of these picoplanktonic cyanobacteria in terms of size and bio-optical environment (Alvain et al., 2012).

The diatoms (Fig. 7d) are characterized by relatively high Ra values compared to the three other groups and a decrease of Ra from 412 to 510 nm. These high values can be mainly explained by the relatively high backscattering associated to diatoms, given the large size of these cells (Alvain et al., 2012). However, two spectral behaviors are present for the diatom label, showing a second type of spectra, which is different in terms of shape and amplitude (Fig. 7e). This second type of spectra is characterized by a minimum value around 412 nm and higher values between 443 and 555 nm. It does not necessarily correspond to another species, but maybe also related to the same species located in a different bio-optical environment (Alvain et al., 2012). Indeed, as mentioned in Section 3, the shape and amplitude of the Ra spectra mainly depends on the bio-optical environment, represented by different values of b_p , a_{cdom} and a_{phy} (Alvain et al., 2012).

The relative frequencies of the pixels associated to the five neurons labeled as diatoms are displayed in Fig. 8 both in the Equatorial Pacific (left panel) and the Southern Atlantic Ocean (right panel). It is shown that these neurons are not all present in the same places. For instance, neuron 90 (represented by the second type of \mathbf{rv} , shown in Fig. 7e) is more frequently localized near the western coasts of South America in the Ecuador and Peru upwelling, while neuron 10 (represented by the

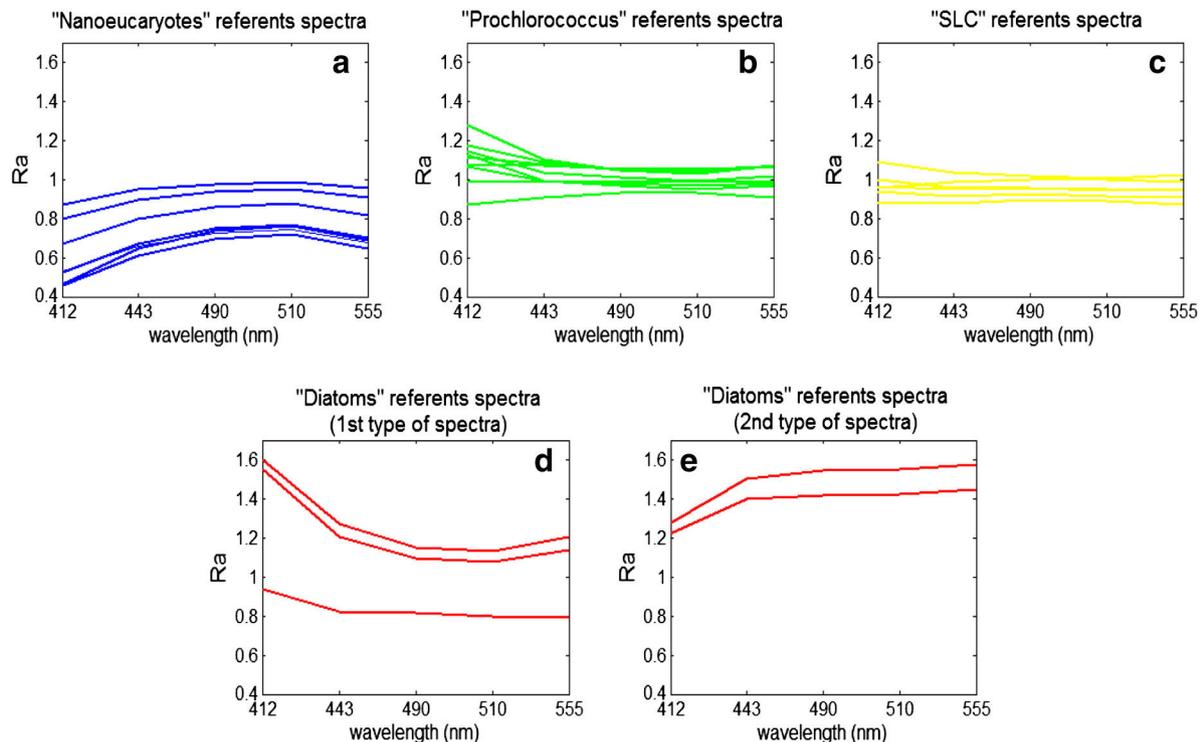


Fig. 7. Radiance anomaly spectra of the reference vectors (\mathbf{rvs}) for the four phytoplankton groups labeled from pigment inventories. Note the two different types of \mathbf{rv} for diatoms.

first type of \mathbf{rv} , shown in Fig. 7d) can also be observed offshore as well as in the Southern Atlantic. It can also be noted that neuron 100 is frequently localized in distinct areas with respect to the other neurons. The a_{cdom} , b_p and a_{phy} value ranges corresponding to the locations of the three most frequently observed neurons labeled as diatom (neuron 10: first type of \mathbf{rv} ; neurons 90 and 100: second type of \mathbf{rv}) are displayed in Fig. 9. The 3 inherent optical properties have been computed using theoretical reference relationships (Bricaud, Morel, Babin, Allali, & Claustre, 1998; Bricaud, Morel, & Prieur, 1981; Loisel, Nicolas, Sciandra, Stramski, & Poteau, 2006; Morel, 2009; Morel & Maritorena, 2001).

Significant differences for the 3 parameters are observed between the 3 neurons. The a_{cdom} values are higher for neurons 90 and 100 than neuron 10, especially at 412 nm (Fig. 9a). As the R_a values at 412 nm are mainly driven by the a_{cdom} (Alvain et al., 2012), this could be an explanation to the observed shift in this second type of spectra. The b_p values are also significantly higher for neurons 90 and 100 than neuron 10 at all the wavelengths (Fig. 9b). The b_p explains the largest part of the R_a absolute values at all the wavelengths (Alvain et al., 2012) and depends on several factors such as the cell-size or the aggregation effect (packaging). The aggregate formation may occur during or after the diatom blooms due to cell surface stickiness (Kiørboe et al., 1998). The a_{phy} values (Fig. 9c) are higher for neurons 90 and 100 than neuron 10 and likely compensate the high values of b_p (Alvain et al., 2012). These observations show that the different neurons, associated with the same phytoplankton group based on pigment information, are located in various environments where the inherent optical properties affecting the R_a spectra are likely to be significantly different.

The time factor may also play a significant role in the R_a spectral variability. The distributions of neurons 10, 90 and 100 are displayed in Fig. 10. It is shown that the occurrence of the different neurons varies between December 2001 and March 2002, which means that these neurons are not present in the same proportions throughout the bloom. This observation suggests that the blooming stage may also be an important factor affecting the R_a spectra. However, these observations are based on theoretical computations of the inherent optical properties and need to be validated by field data.

4.2.3. Validation of the PHYSAT–SOM maps

A validation of the global daily PHYSAT–SOM products is presented here using the 20% of samples which have been randomly selected for each group from the 759 successfully identified inventories (see Section 4.2.1). This validation data set contains 40 water samples dominated by Nanoecaryotes, 43 by *Prochlorococcus*, 52 by SLC and 10 by diatoms. A synthetic view of this validation exercise is provided in Table 1 by showing, for each group, the percentage of valid and wrong identifications. In this validation exercise, we discarded the “*Phaeocystis-like*” since they were identified based on literature only, as in PHYSAT–v2008. Table 1 shows that 66.7% of the pigment inventories corresponding to Nanoecaryotes are associated with the same phytoplankton group in the PHYSAT–SOM daily products. Wrong identifications are associated mostly with *Prochlorococcus* (19.05%) and SLC (9.52%) both in the Southern Hemisphere.

Prochlorococcus and SLC are correctly identified respectively in 58.07% and 66.67% of the in situ pigment inventories. Interestingly, most erroneous identifications of *Prochlorococcus* are associated with the SLC group (22.58%). This result is not surprising considering the close characteristics (geographical distribution and size as well as spectral shape and amplitude of \mathbf{rvs}) of these cyanobacteria. A high fraction (83%) of pigment inventories dominated by diatoms is properly identified in the 9×9 km PHYSAT–SOM daily products. Most erroneous identifications are associated with Nanoecaryotes (16.67%) and are found in waters which are alternatively dominated by Nanoecaryotes and diatoms, essentially in the North Atlantic, the Pacific and the Southern Ocean.

4.2.4. Monthly climatology of phytoplankton groups over the 1997–2010 SeaWiFS archives: global observations and improvements

In this section, the improvements brought by PHYSAT–SOM are shown through the analysis of the global monthly climatology maps of PHYSAT–SOM as well as statistics made on the pixel coverage. We will especially focus on the innovations made on the detection of fine scale spatial structures through a zoom on the diatom blooms in some upwelling regions and in the equatorial Pacific during *La Niña* event.

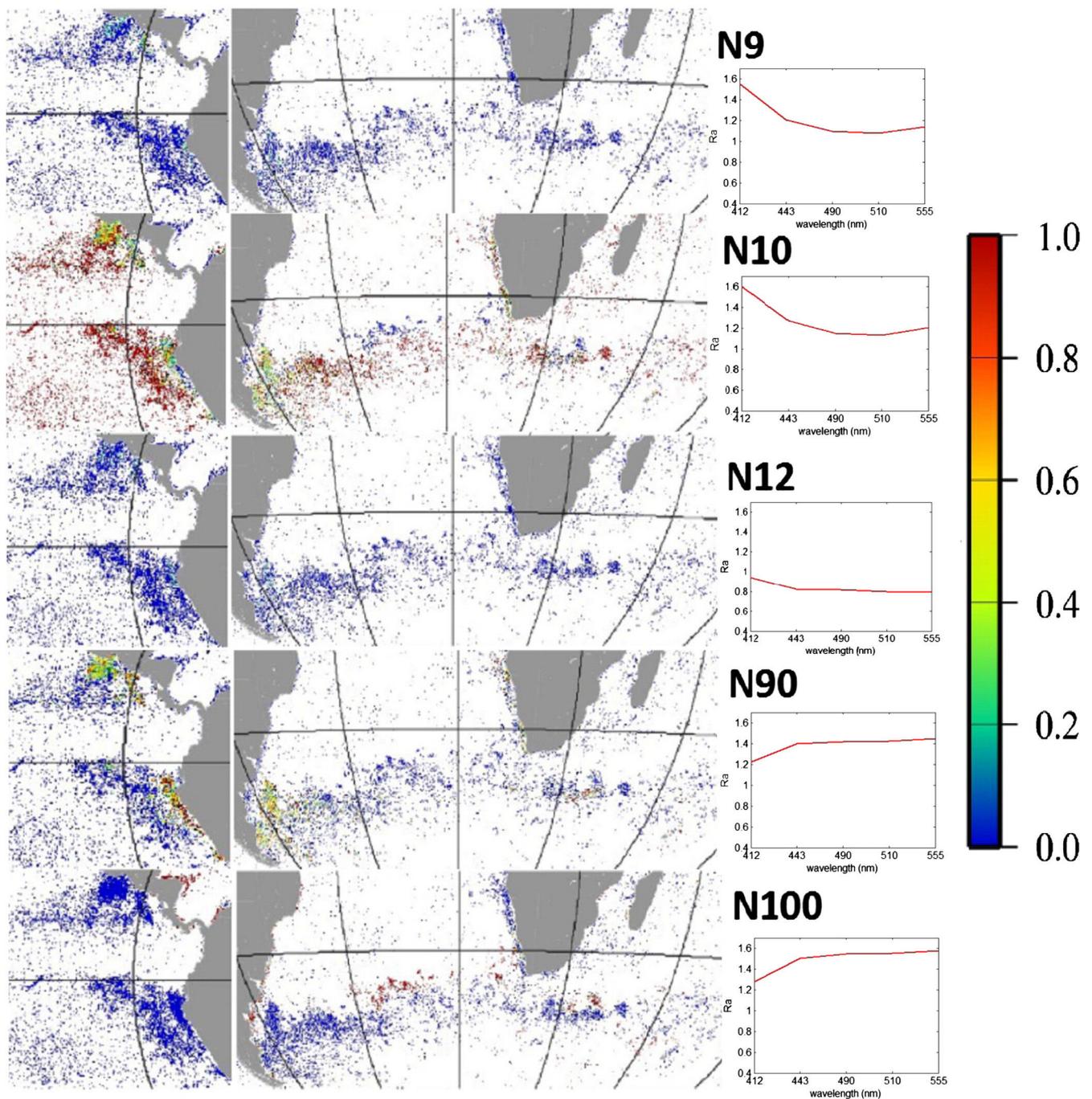


Fig. 8. Relative frequency of the pixels associated to the five neurons labeled as diatoms in the Equatorial Pacific (left panel) and the Southern Atlantic Ocean (right panel) for January 2002. Each neuron is represented with its corresponding reference vector.

Global monthly climatology maps of phytoplankton groups at 9 km resolution over the whole SeaWiFS archive (09/1997–10/2010) have been generated from the daily maps by selecting the most frequently detected group in each pixel during a month and at least for half of the total valid pixels. The climatology maps of the five phytoplankton groups made by PHYSAT–SOM for the months of January and June are displayed respectively in Figs. 11 and 12, together with the climatology maps generated by PHYSAT–v2008. These maps represent the most dominant phytoplankton group in each pixel over the 1997–2010 SeaWiFS imagery archives.

The maps obtained by PHYSAT–SOM shows well-defined patterns that are globally consistent with those obtained by Alvain et al. (2005), but significantly different in some regions, especially

in the upwelling areas. The global distribution of phytoplankton groups is characterized by the dominance of SLC and *Prochlorococcus* groups in oligotrophic tropical waters. Their small size makes them less sensitive to nutrient limitation than larger phytoplankton groups (Longhurst, 2007). Nanoecaryotes and diatoms prevail in the eutrophic waters of high latitudes. A biannual pattern appears in the Southern high latitudes showing a succession between Nanoecaryotes in winter and large-scale diatom blooms in summer, especially in the Southern Ocean during January. The geographical distributions of these 4 phytoplankton groups are coherent with previous studies (e.g. Dandonneau et al., 2004; DuRand, Olson, & Chisholm, 2001; Longhurst, 2007; Marty & Chiavérini, 2002; Steinberg et al., 2001; Zubkov, Sleigh, Burkill, & Leakey, 2000), as already discussed by Alvain et al. (2005). Large

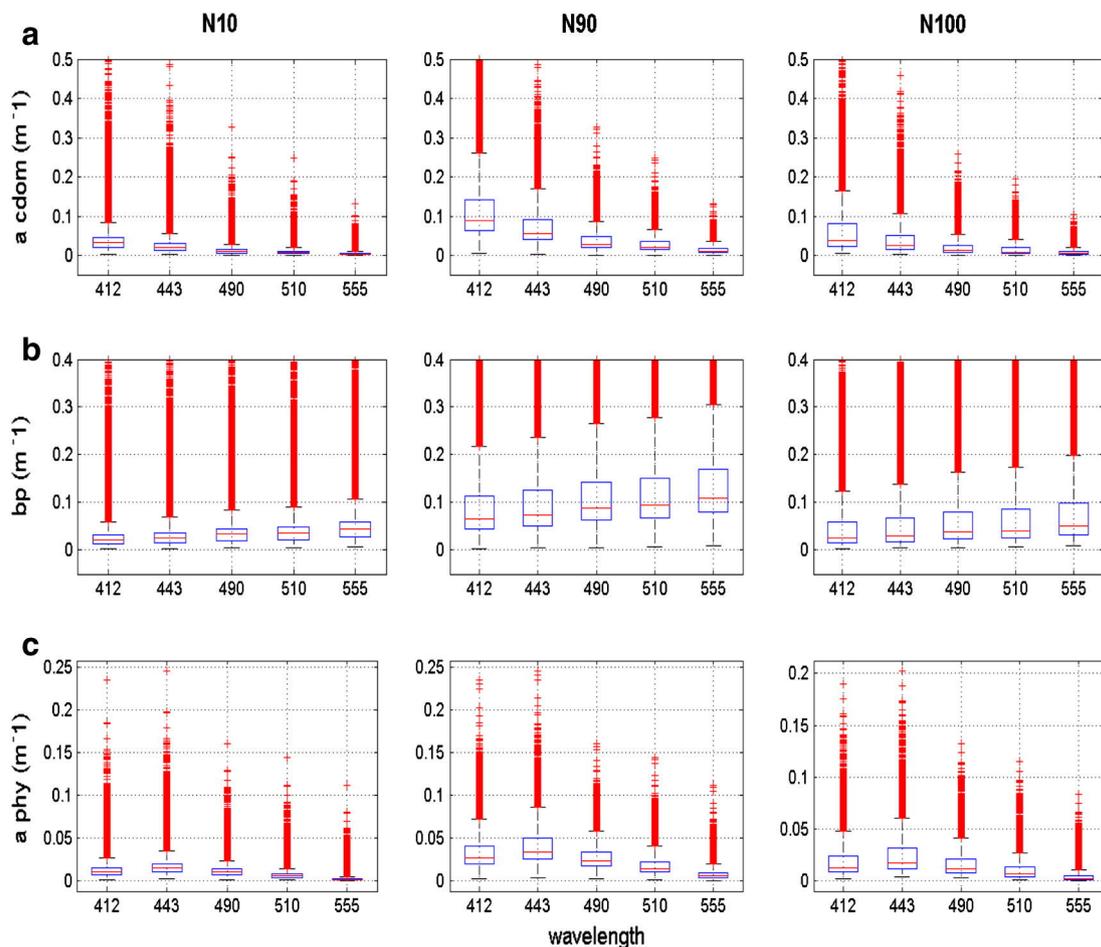


Fig. 9. Inherent optical properties value ranges corresponding to neurons 10 (first type of **rv**), 90 and 100 (second type of **rv**). The ends of the boxplots represent the standard deviation. The red crosses represent the outliers. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

patches of *Phaeocystis-like* are evidenced during the austral summer. *Phaeocystis-like* blooms are clearly visible in the PHYSAT-SOM maps in the English Channel and the Atlantic coasts of France in January as well as in the northern coasts of Australia in June. While, for instance, *Phaeocystis* blooms have been regularly observed in the English Channel coastal waters (e.g. Breton, Brunet, Sautour, & Brylinski, 2000; Brunet et al., 1996; Houliez, Lizon, Thyssen, Artigas, & Schmitt, 2012; Lancelot & Mathot, 1987; Rousseau, Lancelot, & Cox, 2006), our satellite estimates need to be validated with complimentary in situ identification in these complex waters. In the PHYSAT-SOM monthly climatology maps, diatom blooms are clearly visible in areas characterized by strong upwelling conditions. They are particularly dominant in the Benguela ecosystem (e.g. Kiørboe et al., 1998; Romero et al., 2002), in the South Atlantic (near Angola and Namibia coasts) or along the Equator and Peru coasts in the Equatorial Pacific, especially during the austral summer, which is in agreement with previous studies. Indeed, Jackson, Bouman, Sathyendranath, and Devred (2011) reported that a region of strong upwelling over the wide Peruvian shelf around 15°S was characterized by a strong diatom population, even during the El Niño event. DiTullio et al. (2005) found that, on the Peruvian continental shelf, upwelling of iron-rich subsurface waters led to massive diatom blooms, reaching 80–90% of the phytoplankton population with chl-*a* values $>45 \mu\text{g}\cdot\text{L}^{-1}$. Moreover, Bruland, Rue, Smith, and DiTullio (2005) found, through a fieldwork performed in the Peru upwelling regime during August and September 2000, that the extremely high-chlorophyll “brown waters of Peru” (chl-*a* between 20 and $45 \mu\text{g}\cdot\text{L}^{-1}$) result from massive diatom blooms occurring in iron-rich upwelling region observed over the broad continental shelf off

northern and central Peru. Besides, diatom blooms are clearly observed along the Humboldt Current (along the west coast of South America from the southern tip of Chile to northern Peru). Incubation experiments performed by Hutchins et al. (2002) stated that, in non-iron-limiting conditions, these waters include large pennate diatoms. The distribution of diatoms is strongly controlled by the thickness of the mixed layer and they are found preferentially in stratified waters, which allow the development of the bloom (Jochem, Mathot, & Queguiner, 1995; Weber & El-Sayed, 1987). Interestingly, such local phenomena are more visible in the climatology maps produced by PHYSAT-SOM than those generated by PHYSAT-v2008. This latter difference is due to the ability of SOM to classify a higher variability of *Ra* spectra than PHYSAT-v2008 (as explained previously in Section 4.1) as well as the contribution of a large in situ dataset, covering various environmental conditions and more easily manageable with SOM than it would be with the previous version of PHYSAT.

Another main advantage of using SOM lies in the larger spatial coverage comparing to PHYSAT-v2008, which is a direct consequence of the larger variability of *Ra* spectra taken into account (as demonstrated in Section 4.1). The global monthly detection ratio over the 1997–2010 SeaWiFS imagery archives is displayed in Table 2. This ratio represents the number of pixels covered by PHYSAT-SOM compared to those covered by PHYSAT-v2008 for the whole climatology. The results show that PHYSAT-SOM labels, on average, two times more pixels than PHYSAT-v2008 for the total amount of pixels. *Phaeocystis-like* is characterized by the highest detection ratios with a maximum ratio value around 30 followed by diatoms with a maximum ratio of 10. *Prochlorococcus* and SLC are detected two times more by PHYSAT-SOM on average with a

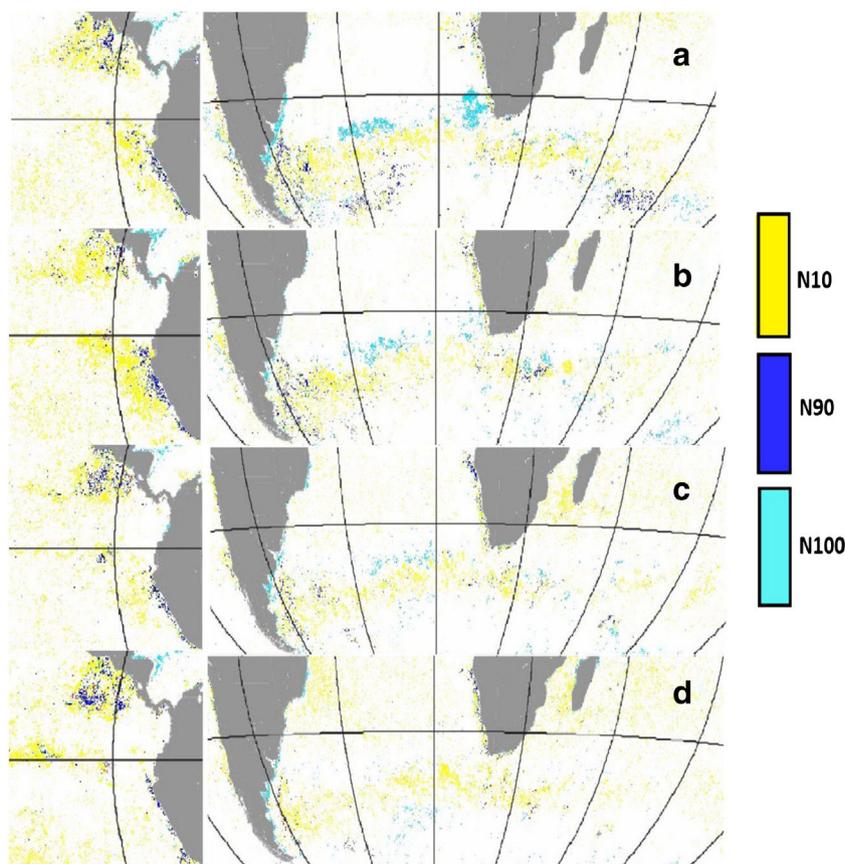


Fig. 10. Distribution of the three most frequent neurons labeled as diatoms in the Equatorial Pacific (left panel) and the Southern Atlantic Ocean (right panel) for December 2001 (a), January 2002 (b), February 2002 (c) and March 2002 (d).

maximum ratio of 3.21 for SLC in September. However, it might seem that the SLC are underestimated by PHYSAT–SOM in the climatology maps. This appearance is due to the fact that the pixels dominated by SLC and *Prochlorococcus* are homogeneously mixed in the PHYSAT–SOM maps while in the PHYSAT–v2008 maps, several areas are dominated only by SLC. The Nanoecaryotes are characterized by the lowest detection ratios between the 2 versions, with a minimum in February (0.95) and a mean value of 1.5 over the whole climatology.

In January, large patches of Nanoecaryotes are detected in the PHYSAT–v2008 map in the southern Atlantic Ocean, from the eastern Patagonian coastal waters to the western coasts of South Africa (Fig. 11). However, in the PHYSAT–SOM map, diatoms are more dominant in these waters. This could be explained by the fact that the GeP&CO pigment inventory set used to develop PHYSAT–v2008 (Alvain et al., 2005) did not include samples from the southern Atlantic Ocean (such as samples taken offshore the Patagonian coasts). These waters are known to be characterized by strong diatom blooms in that period, especially in non-iron limiting regions such as the continental shelf Patagonian waters and near South Georgia and South Sandwich Islands (e.g., Borrione & Schlitzer, 2012; Korb, Whitehouse, & Ward, 2004).

Table 1

Validation results for the PHYSAT–SOM method. For each group, the percentage of valid (bold) and wrong (others) identifications are shown. Wrong identifications are separated by groups.

PHYSAT–SOM/in situ	Nanoecaryotes	<i>Prochlorococcus</i>	SLC	Diatoms
Nanoecaryotes	66.67	19.05	9.52	4.76
<i>Prochlorococcus</i>	19.35	58.07	22.58	–
SLC	20	13.33	66.67	–
Diatoms	16.67	–	–	83.33

The strong 1998 *La Niña* return produced very strong and unusual blooms (Chavez et al., 1999; Masotti et al., 2011; Ryan, Polito, Strutton, & Chavez, 2002). It is generally admitted that diatoms do not contribute to more than 20% of phytoplankton biomass in this region (Blanchot, André, Navarette, Neveux, & Radenac, 2001; Dandonneau et al., 2004; Kobayashi & Takahashi, 2002; Masotti et al., 2011). However, it has been shown that this bloom was dominated by diatoms. Their blooms may break out when equatorial upwelling is intense and non-limited to the coastal region. *La Niña* event of August 1998 brought large amounts of nutrients into the photic layer (Ryan et al., 2002; Strutton & Chavez, 2000), contributing to rapid diatom growth. This diatom bloom was also confirmed by Gorgues et al. (2010), using a biogeochemical model. Their observations were also consistent with those of PHYSAT–v2008. This unusual bloom has been revisited with PHYSAT–SOM and compared to PHYSAT–v2008. This was done by computing the relative frequency of the pixels associated to diatoms from a synthesis of August 1998. As shown in Fig. 13, this bloom is more clearly visible in the PHYSAT–SOM map, with a higher pixel density associated to the better spatial coverage. Well-defined spatial structures, characterized by large meanders, are clearly visible in the PHYSAT–SOM map. These structures may be due to the interactions between north and south equatorial currents. Satellite observations of Rossby waves (very low-frequency waves in the ocean's surface and thermocline) and chlorophyll anomalies propagating in subtropical gyres have suggested that the sea color variations are caused by the accumulation of floating particles by Rossby waves, acting as “marine hay rakes” in convergence zones (Dandonneau, Vega, Loisel, Du Penhoat, & Menkes, 2003). Such processes may be determinant for the distribution of living organisms in oligotrophic waters.

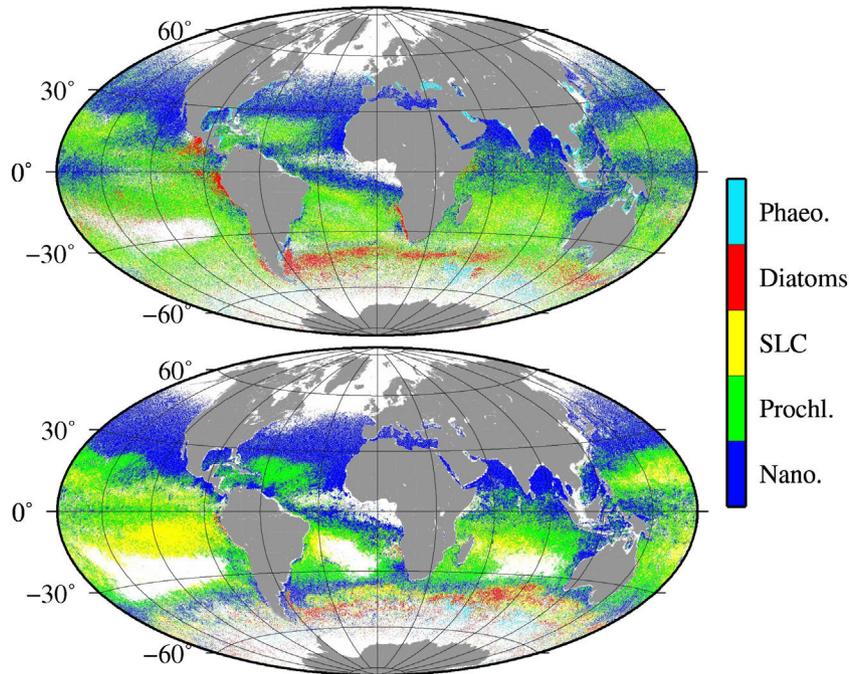


Fig. 11. Global climatology of most frequently dominant phytoplankton groups over the 1997–2010 SeaWiFS image archive made by PHYSAT-SOM (top) and PHYSAT-v2008 (bottom) for January month.

5. Conclusion

The global variability of phytoplankton groups is presented in this study based on PHYSAT spectral anomalies of water-leaving radiances (Ra) computed from SeaWiFS nLw. The PHYSAT method empirically related the Ra spectra with the presence of dominant phytoplankton groups. Thanks to the use of an objective automatic classification method, significant improvements and perspectives in terms of phytoplankton

groups remote sensing are brought to the original PHYSAT approach. In order to define a characteristic range of Ra for each phytoplankton group, it was previously necessary to introduce several thresholds and shape criteria on the spectra so that there is no overlapping between the different spectral signatures of the phytoplankton groups. This approach makes the method limited in terms of spectra diversity characterization accuracy, as shown in Section 4.1, and thus restricts its extension to a wide number of applications.

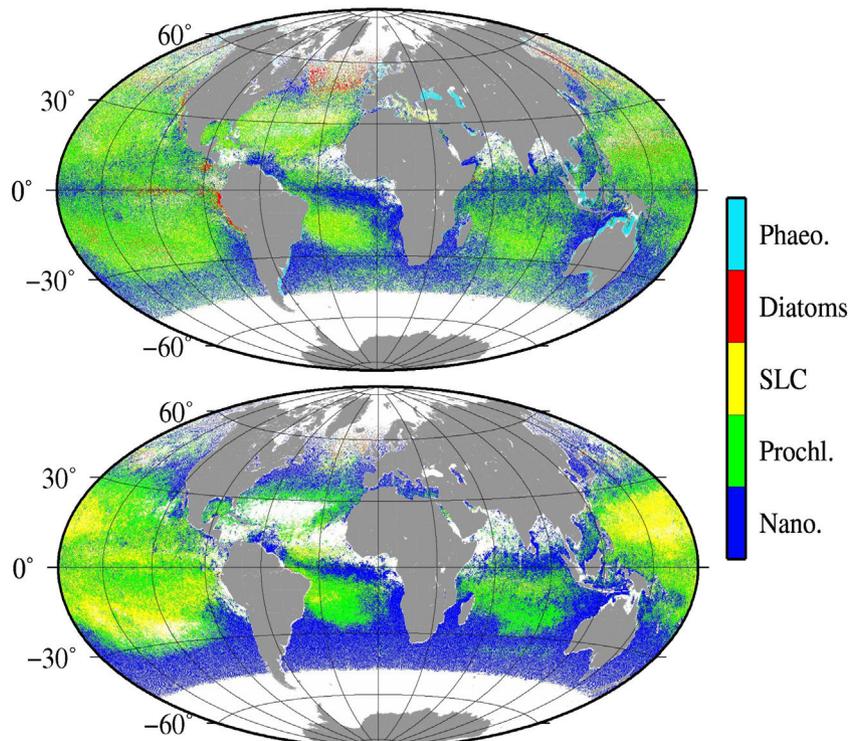


Fig. 12. Global climatology of most frequently dominant phytoplankton groups over the 1997–2010 SeaWiFS image archive made by PHYSAT-SOM (top) and PHYSAT-v2008 (bottom) for June month.

Table 2

Global monthly detection ratio (PHYSAT–SOM/PHYSATv–2008) over the 1997–2010 SeaWiFS imagery archives.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Nano	1.34	0.95	1.25	1.58	1.80	1.95	1.45	1.52	1.44	1.54	1.44	1.51
Porchlo	1.76	1.83	1.84	1.72	1.56	2.04	2.53	2.00	1.69	2.05	2.00	1.94
SLC	1.96	2.03	1.89	2.52	2.41	1.72	2.54	2.07	3.21	2.31	2.00	2.12
Diatoms	2.32	3.47	5.06	5.01	10.03	4.09	4.61	5.40	8.14	5.46	5.13	5.48
Phaeo	5.24	7.26	32.31	29.52	19.73	25.93	15.61	18.41	32.21	20.83	11.16	6.163
Total	1.80	1.54	1.83	2.04	2.02	2.07	2.24	2.01	1.98	2.15	2.06	1.99

A neural network tool (Self-Organizing Maps) was used to automatically classify in terms of shape and amplitude a set of Ra spectra computed from a large satellite data set in order to have the largest possible coverage of characteristic spectral anomalies. Our approach, called PHYSAT–SOM is a two-step method. The first step consists in training the neural network: the most pertinent spectral patterns are extracted from the SeaWiFS images by compressing the information of the images into 100 reference vectors using SOM (unsupervised step). The second step, which is a supervised one, consists in relating these spectral patterns to in situ pigment inventories in order to assign them to a phytoplankton group (labeling procedure).

Before the labeling of SOM, a comparison between the original PHYSAT approach and PHYSAT–SOM, based on a set of 41 Ra spectra collocated with field pigment inventories, demonstrates the improvement of the spectral retrieval with the neural network tool. It is shown that SOM allows retrieving all the Ra spectra. Indeed, due to the thresholds introduction to define the spectral signatures of each phytoplankton group, the previous method was able to classify only 20 individual spectra out of 41. This means that, with SOM, it is now possible to label more spectra from the same dataset that was used to develop the original method. This result underlines the advantage to use the neural network tool with respect to the initial approach, in addition to time-saving and to the possibility to easily work with a much larger dataset. Most importantly, SOM gives also the possibility to study the Ra global spectral variability and diversity regardless of the labeling possibilities, related to the in situ data availability.

In a second step, a large in situ pigment inventory dataset, gathering samples from various cruises, was used to allocate a phytoplankton group to each reference vector determined by the SOM method. The efficient integration of this dataset, taking into account large spatial and temporal scales as well as various environmental conditions, is a major advantage of the neural network tool. SOM allowed relating these spectra to phytoplankton groups in a timely manner, with a higher flexibility and reliability, which would be much more difficult to achieve with the initial approach.

Four phytoplankton groups (Nanoeucaryotes, *Prochlorococcus*, SLC and diatoms) have been identified from the in situ data, based on thresholds of biomarker pigment ratios, taken from the literature. The collocated Ra spectra corresponding to each group were then projected on the trained SOM and assigned to a specific neuron. The Ra spectral signatures of each group are now defined from the set of reference vectors of their corresponding neurons. Characteristic spectra in terms of shape and amplitude were distinguished for each phytoplankton group. It is also possible to associate several spectral signatures to the same phytoplankton group. These different spectra do not necessarily correspond to different species but may be also related to the same one in different blooming stages and/or bio-optical environments. In fact, the shape and amplitude of the Ra spectra mainly depend on the bio-optical environment, represented by different values of particulate scattering (b_p), absorption by the colored dissolved organic matter (a_{cdom}) and the absorption by the phytoplankton (a_{phy}), as it has been proven by the Ra sensitivity study made by [Alvain et al. \(2012\)](#). For example, it was found that the diatoms were represented by two different Ra spectra. The results show that the different neurons labeled as diatoms are associated to various environments where the inherent optical properties affecting the Ra spectra are likely to be significantly different. It was also shown that the occurrence of these neurons varies throughout the bloom. This observation suggests that the blooming stage may also be an important factor affecting the Ra spectra.

The classifier was then applied to the daily SeaWiFS images from 1997 to 2010: each Ra spectrum of the maps was assigned to its nearest neuron (hence reference vector) of the SOM. The monthly climatology maps obtained by PHYSAT–SOM show well-defined patterns, that are globally consistent with those obtained by [Alvain et al. \(2005\)](#), but significantly different in some regions. The visibility of local phenomena, especially in the upwelling zones, characterized by strong diatom blooms, is significantly improved. The diatom blooms in the Peru upwelling and the western coasts of South America have been described and discussed based on previous studies, which are consistent with our results in that region. Moreover, the diatom bloom following the *La Niña* event of August 1998 has been illustrated and revisited. This bloom is more clearly visible in the PHYSAT–SOM map, with a higher pixel density associated to the better spatial coverage. Based on the whole climatology, it was found that the PHYSAT–SOM method labels two times more pixels than PHYSAT–v2008 on average. High detection ratios are reached thanks to the ability of SOM to take into account a higher spectral variability, as shown by the comparison exercise in terms of Ra spectra retrieval ([Section 4.1](#)).

The results of this study show that PHYSAT–SOM gives the ability to significantly improve the original PHYSAT–v2008 approach. These improvements lie in an accurate and efficient retrieval of the Ra spectra. The integration of larger in situ datasets is a major advantage of our method. It gives the opportunity to improve fine scale spatial pattern detection and gives the possibility to explore other spectral signatures, even associated with the same label. These new features provide additional perspectives in phytoplankton group remote sensing, going beyond the simple dominance to the possibility of phytoplankton group detection in different bio-optical environments and blooming stages.

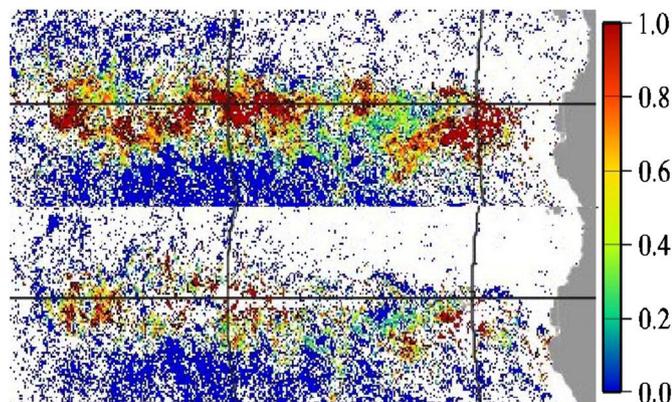


Fig. 13. Relative frequency of the pixels associated to diatoms in the Equatorial Pacific Ocean during *La Niña* event of August 1998 for PHYSAT–SOM (top) and PHYSAT–v2008 (bottom).

Furthermore, as a perspective, the Ra spectral signatures automatically classified by SOM could be labeled using other approaches than biomarker pigments. For example, we could try to combine the spectral response approach used in this study with ecological approaches, such as the method developed by Raitso et al. (2008), incorporating ecological and geographical knowledge together with ocean color, bio-optical characteristics, and remotely sensed physical parameters (e.g. sea surface temperature, wind stress, photosynthetically active radiation). This could be done by trying to label the Ra spectra in terms of phytoplankton groups derived from the ecological model results. This combination could be a way to explore a possible relationship between the spectral anomalies and the presence of different or same phytoplankton groups in various environmental and ecological conditions. Since the backscattering values determining the Ra spectra magnitude depend from the particle size, another perspective would be to explore the relationships between the radiance anomaly spectra and different phytoplankton size classes, determined with other spectral approaches, based, for instance, on the specific absorption of phytoplankton (e.g. Brewin, Lavender, Hardman-Mountford, & Hirata, 2010; Ciotti & Bricaud, 2006).

Acknowledgments

This work is funded (PhD fellowship) by the SMCO (Syndicat Mixte de la Côte d'Opale), the Région Nord-Pas de Calais and the CNES (Centre National d'Etudes Spatiales) through the TOSCA/PHYTOCOT project. The authors acknowledge NASA/GSFC/DAAC for providing access to daily L3 SeaWiFS binned products as well as all the contributors to the NOMAD dataset. We are grateful to all the organizers and observers of GeP&CO, HALIFAX, ICOTA and OOMPH campaigns (see details in the text).

Appendix A. The Self-Organizing Map algorithm

Self-Organizing Maps (SOM) are neural models, which were first introduced by Kohonen (1984), for visualizing and clustering multidimensional observations. The principal goal of SOM is to transform an incoming signal pattern of arbitrary dimension (5 dimensions in the present paper) into a one or two-dimensional discrete map. In the present case, we use a regular two-dimensional grid (10×10 neurons) (Fig. 1). Each neuron in the lattice is fully connected to all the source nodes in the input layer by its weight vector (\mathbf{w}) whose dimension is thus the dimension of the observation. Each neuron of the map (C) is characterized by its output value and its weight vector, which are respectively a real and an n -dimensional reference vector denoted (\mathbf{rv}), where n is the dimension of the input space. The graph structure induces a discrete distance d on C :

For any pair of neurons (c, r), $d(c, r)$ is defined as the length of the shortest path between c and r on C . For each neuron c , the discrete distance defines the notion of neighborhood of order d of c :

$$V_c(d) = \{r \in C, d(c, r) \leq d\}. \quad (A1)$$

The aim of the training phase is to construct a system of reference vectors, which represents a partition of the input space (observations) in such a way that they conserve the topological properties on C . Two similar observations are associated with two "Euclidean-similar" reference vectors in C .

There are three basic steps involved in the application of the algorithm after initialization: sampling, similarity matching, and updating. These three steps are repeated until formation of the feature map has completed. The algorithm is summarized as follows:

1. Initialization: choose random values for the initial weight $\mathbf{w}_j(0)$. The only restriction here is that the $\mathbf{w}_j(0)$ must be different for $j = 1, 2, \dots, l$, where l is the number of neurons of the map. It may be desirable to keep the magnitude of the weights small.

2. Sampling: draw a sample \mathbf{x} from the input space with a certain probability; the vector \mathbf{x} represents the activation pattern that is applied to the lattice. The dimension of \mathbf{x} is equal to the dimension of the input space.
3. Similarity matching: find the best-matching (winning) neuron $i(\mathbf{x})$ at time step n by using the minimum-distance Euclidean criterion:

$$i(\mathbf{x}) = \arg \min (j) \|\mathbf{x}(n) - \mathbf{w}_j\| \quad (A2)$$

with $j = 1 \dots 100$ as we use a two-dimensional 10×10 topological map.

4. Updating: adjust the synaptic weight vectors of all neurons by using the update formula

$$\mathbf{w}_j(n+1) = \mathbf{w}_j(n) + \eta(n) h(n) (\mathbf{x}(n) - \mathbf{w}_j(n)) \quad (A3)$$

where $\eta(n)$ is the learning-rate parameter and $h(n)$ is the neighborhood function centered on the winning neuron $i(\mathbf{x})$; both $\eta(n)$ and $h(n)$ are varied dynamically during learning for best results.

5. Continuation: continue with step 2 until no noticeable changes in the topological map are observed.

At the end of the training phase, SOM provides a partition of the observations together with a topological order. A given observation is assigned to a unique neuron of the map. The topological order on the lattice represents the existing order between the different subsets of the partitions (Vesanto et al., 1999). In SOM formalism, each neuron of the map is associated with its reference vector and, as mentioned before, represents a subset of observations. As the reference vectors and the observations have the same dimension, the set of neurons can be viewed as a set of prototype observations, which sums up the observations under study.

References

- Ackleson, S. G., Balch, W. M., & Holligan, P.M. (1994). Response of water-leaving radiance to particulate calcite and chlorophyll-*a* concentrations: A model for Gulf of Maine coccolithophore blooms. *Journal of Geophysical Research*, 99(C4), 7483–7499.
- Ainsworth, E. J., & Jones, S. F. (1999). Radiance spectra classification from the ocean color and temperature scanner on ADEOS. *IEEE Transactions on Geoscience and Remote Sensing*, 37, 1645–1656.
- Alvain, S., Loisel, H., & Dessailly, D. (2012). Theoretical analysis of ocean color radiances anomalies and implications for phytoplankton groups detection in case 1 waters. *Optics Express*, 20, 1070–1083.
- Alvain, S., Moulin, C., Dandonneau, Y., & Breon, F. M. (2005). Remote sensing of phytoplankton groups in case 1 waters from global SeaWiFS imagery. *Deep-Sea Research I*, 52, 1989–2004.
- Alvain, S., Moulin, C., Dandonneau, Y., & Loisel, H. (2008). Seasonal distribution and succession of dominant phytoplankton groups in the global ocean: A satellite view. *Global Biogeochemical Cycles*, 22, GB3001. <http://dx.doi.org/10.1029/2007GB003154>.
- Astoreca, R., Rousseau, V., Ruddick, K., Knechtick, C., Van Mol, B., Parent, J. -Y., et al. (2009). Development and application of an algorithm for detecting *Phaeocystis globosa* blooms in the Case 2 Southern North Sea waters. *Journal of Plankton Research*, 31, 287–300.
- Bailey, S. W., & Werdell, P. J. (2006). A multi-sensor approach for the on-orbit validation of ocean color satellite data products. *Remote Sensing of Environment*, 102, 12–23.
- Barlow, R. G., Aiken, J., Holligan, P.M., Cummings, D.G., Maritorena, S., & Hooker, S. (2002). Phytoplankton pigment and absorption characteristics along meridional transects in the Atlantic Ocean. *Deep-Sea Research I*, 47, 637–660.
- Belviso, S., Moulin, C., Bopp, L., & Stefels, J. (2004). Assessment of a global climatology of oceanic dimethylsulfide (DMS) concentrations based on SeaWiFS imagery (1998–2001). *Canadian Journal of Fisheries and Aquatic Sciences*, 61, 804–816.
- Blanchot, J., André, J. M., Navarette, C., Neveux, J., & Radenac, M. H. (2001). Picophytoplankton in the equatorial Pacific: Vertical distributions in the warm pool and in the high nutrient low chlorophyll conditions. *Deep-Sea Research I*, 48, 297–314.
- Borrione, I., & Schlitzer, R. (2012). Distribution and recurrence of phytoplankton blooms around South Georgia, Southern Ocean. *Biogeosciences Discuss*, 9, 10087–10120. <http://dx.doi.org/10.5194/bgd-9-10087-2012>.
- Brajard, J., Niang, A., Sawadogo, S., Fell, F. R., Santer, R., & Thiria, S. (2007). Estimating aerosol parameters above ocean from MERIS observation using topological maps. *International Journal of Remote Sensing*, 28, 781–795.
- Breton, E., Brunet, C., Sautour, B., & Brylinski, J. -M. (2000). Annual variations of phytoplanktonic biomass in the Eastern English Channel: Comparison by pigment signatures and microscopic counts. *Journal of Plankton Research*, 22, 1423–1440.

- Brewin, R. J. W., Lavender, S. J., Hardman-Mountford, N. J., & Hirata, T. (2010). A spectral response approach for detecting dominant phytoplankton size class from satellite remote sensing. *Acta Oceanologica Sinica*, 29(2), 14–32.
- Bricaud, A., Morel, A., Babin, M., Allali, K., & Claustre, H. (1998). Variations of light absorption by suspended particles with chlorophyll a concentration in oceanic (case 1) waters: Analysis and implications for bio-optical models. *Journal of Geophysical Research*, 103(C13), 31033–31044.
- Bricaud, A., Morel, A., & Prieur, L. (1981). Absorption by dissolved organic matter of the sea (yellow substance) in the UV and visible domains. *Limnology and Oceanography*, 26, 43–53.
- Brown, C. W., & Podesta, G. P. (1997). Remote sensing of coccolithophore blooms in the Western South Atlantic Ocean. *Remote Sensing of Environment*, 60, 83–91.
- Brown, C. W., & Yoder, J. A. (1994). Coccolithophorid blooms in the global ocean. *Journal of Geophysical Research*, 99, 7467–7482.
- Brunland, K. W., Rue, E. L., Smith, G. J., DiTullio, G. R. (2005). Iron, macronutrients and diatom blooms in the Peru upwelling regime: Brown and blue waters of Peru. *Marine Chemistry*, 93, 81–103.
- Brunet, C., Brylinski, J.-M., Bodineau, L., Thoumelin, G., Bentley, D., & Hilde, D. (1996). Phytoplankton dynamics during the spring bloom in the South-Eastern English Channel. *Estuarine, Coastal and Shelf Science*, 43, 469–483.
- Chavez, F. P., Strutton, P. G., Friederich, G. E., Feely, R. A., Feldman, G. C., Foley, D. G., et al. (1999). Biological and chemical response of the equatorial Pacific ocean to the 1997–98 El Niño. *Science*, 286, 2126–2131. <http://dx.doi.org/10.1126/science.288.5447.2126>.
- Chazottes, A., Bricaud, B., Crépon, M., & Thiria, S. (2006). Statistical analysis of a database of absorption spectra of phytoplankton and pigment concentrations using self-organizing maps. *Applied Optics*, 45, 8102–8115.
- Chazottes, A., Crépon, M., Bricaud, A., Ras, J., & Thiria, S. (2007). Statistical analysis of absorption spectra of phytoplankton and of pigment concentrations observed during three POMME cruises using a neural network clustering method. *Applied Optics*, 46, 3790–3799.
- Ciotti, A.M., & Bricaud, A. (2006). Retrievals of a size parameter for phytoplankton and spectral light absorption by coloured detrital matter from water-leaving radiances at SeaWiFS channels in a continental shelf off Brazil. *Limnology and Oceanography: Methods*, 4, 237–253.
- Claustre, H. (1994). Phytoplankton pigment signatures of the trophic status in various oceanic regimes. *Limnology and Oceanography*, 39, 1206–1211.
- Dandonneau, Y., Deschamps, P. Y., Nicolas, J. M., Loisel, H., Blanchot, J., Montel, Y., et al. (2004). Seasonal and interannual variability of ocean color and composition of phytoplankton communities in the North Atlantic, Equatorial Pacific and South Pacific. *Deep-Sea Research II*, 51, 303–318.
- Dandonneau, Y., Vega, A., Loisel, H., Du Penhoat, Y., & Menkes, C. (2003). Oceanic Rossby waves acting as a “hay rake” for ecosystem floating by-products. *Science*, 302, 1548–1551.
- DiTullio, G. R., Geesey, M. E., Maucher, J. M., Alm, M. B., Riseman, S. F., & Bruland, K. W. (2005). Influence of iron on algal community composition and physiological status in the Peru upwelling system. *Limnology and Oceanography*, 50, 1887–1907.
- Dupouy, C., Neveux, J., Dirberg, G., Röttgers, R., Tenório, M. B., & Ouilou, S. (2008). Bio-optical properties of the marine cyanobacteria *Trichodesmium* spp. *Journal of Applied Remote Sensing*, 2, 1–17. <http://dx.doi.org/10.1117/1.2839036>.
- DuRand, M.D., Olson, R. J., & Chisholm, S. W. (2001). Phytoplankton population dynamics at the Bermuda Atlantic time-series station in the Sargasso Sea. *Deep-Sea Research II*, 48, 1983–2003.
- Ehsani, A. H., & Quiel, F. (2008). Application of self organizing map and SRM data to characterize yardangs in the Lut desert, Iran. *Remote Sensing of Environment*, 112, 3284–3294.
- Falcitore, A., d’Alcalá, M. R., Croot, P., & Bowler, C. (2000). Perception of environmental signals by a marine diatom. *Science*, 288(5475), 2363–2366.
- Gorgues, T., Menkes, C., Slemmons, L., Aumont, O., Dandonneau, Y., Radenac, M. H., et al. (2010). Revisiting the La Niña 1998 phytoplankton blooms in the equatorial Pacific. *Deep-Sea Research I*, 57, 567–576.
- Gorricha, J., & Lobo, V. (2012). Improvements of the visualization of clusters in geo-referenced data using self-organizing maps. *Computers & Geosciences*, 43, 177–186.
- Houliez, E., Lizon, F., Thyssen, M., Artigas, L. F., & Schmitt, F. G. (2012). Spectral fluorometric characterization of Haptophyte dynamics using the FluoroProbe: an application in the Eastern English Channel for monitoring *Phaeocystis globosa*. *Journal of Plankton Research*, 34, 136–151.
- Hu, X., & Weng, Q. (2009). Estimating impervious surfaces from medium spatial resolution imagery using the self-organizing map and multi-layer perceptron neural networks. *Remote Sensing of Environment*, 113, 2089–2102.
- Hutchins, D. A., Hare, C. E., Weaver, R. S., Zhang, Y., Firme, G. F., DiTullio, G. R., et al. (2002). Phytoplankton iron limitation in the Humboldt Current and Peru Upwelling. *Limnology and Oceanography*, 47, 997–1011.
- Iskandar, I. (2010). Seasonal and interannual patterns of sea surface temperature in Banda Sea as revealed by self-organizing map. *Continental Shelf Research*, 30, 1136–1148.
- Jackson, T., Bouman, H. A., Sathyendranath, S., & Devred, E. (2011). Regional-scale changes in diatom distribution in the Humboldt upwelling system as revealed by remote sensing: implications for fisheries. *ICES Journal of Marine Science*, 68, 729–736.
- Jochem, F. J., Mathot, S., & Queguiner, B. (1995). Size fractionated primary production in the open Southern Ocean in austral spring. *Polar Biology*, 15, 381–392.
- Kjørboe, T., Tiselius, P., Mitchell-Innes, B., Hansen, J. L. S., Visser, A. W., & Mari, X. (1998). Intensive aggregate formation with low vertical flux during an upwelling-induced diatom bloom. *Limnology and Oceanography*, 43, 104–116.
- Kobayashi, F., & Takahashi, K. (2002). Distribution of diatoms along the equatorial transect in the western and central Pacific during the 1999 La Niña conditions. *Deep-Sea Research II*, 49, 2801–2821.
- Kohonen, T. (1984). *Self-organization and associative memory* (2nd ed.) Berlin, Heidelberg: Springer-Verlag.
- Kohonen, T. (2013). Essentials of the self-organizing map. *Neural Networks*, 37, 52–65.
- Korb, R. E., Whitehouse, M. J., & Ward, P. (2004). SeaWiFS in the southern ocean: Spatial and temporal variability in phytoplankton biomass around South Georgia. *Deep-Sea Research II*, 51, 99–116.
- Lancelot, C., & Mathot, S. (1987). Dynamics of a Phaeocystis-dominated spring bloom in Belgian coastal waters. 1. Phytoplankton activities and related parameters. *Marine Ecology Progress Series*, 37, 239–248.
- Loisel, H., Nicolas, J.-M., Sciandra, A., Stramski, D., & 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. <http://dx.doi.org/10.1029/2005JC003367>.
- Longhurst, A. (2007). *Ecological geography of the sea*. (2nd ed.) San Diego, California: Academic press.
- Lubac, B., Loisel, H., Guiselin, N., Astoreca, R., Artigas, L. F., & Mériaux, X. (2008). Hyperspectral and multispectral ocean color inversions to detect *Phaeocystis globosa* blooms in coastal waters. *Journal of Geophysical Research*, 113, C06026. <http://dx.doi.org/10.1029/2007JC004451>.
- Mackey, M.D., Higgins, H. W., & Wright, S. W. (1996). CHEMTAX – A program for estimating class abundances from chemical markers: application to HPLC measurements of phytoplankton. *Ecology Progress Series*, 144, 265–283.
- Marty, J. C., & Chiavérini, J. (2002). Seasonal and interannual variations in phytoplankton production at DYFAMED time-series station, northwestern Mediterranean Sea. *Deep-Sea Research II*, 49, 2017–2030.
- Masotti, I., Moulin, C., Alvain, S., Bopp, L., Tagliabue, A., & Antoine, D. (2011). Large-scale shifts in phytoplankton groups in the Equatorial Pacific during ENSO cycles. *Biogeosciences*, 8, 539–550. <http://dx.doi.org/10.5194/bg-8-539-2011>.
- Mobley, C. D., & Sundman, L. (2003). Effects of optically shallow bottoms on upwelling radiances: Inhomogeneous and sloping bottoms. *Limnology and Oceanography*, 48, 329–336.
- Mobley, C. D., Zhang, H., & Voss, K. J. (2003). Effects of optically shallow bottoms on upwelling radiances: Bidirectional reflectance distribution function effects. *Limnology and Oceanography*, 48, 337–345.
- Morel, A. (1997). Consequences of a *Synechococcus* bloom upon the optical properties of oceanic (Case 1) waters. *Limnology and Oceanography*, 42(8), 1746–1754.
- Morel, A. (2009). Are the empirical relationships describing the bio-optical properties of case 1 waters consistent and internally compatible? *Journal of Geophysical Research*, 114(C1), C01016. <http://dx.doi.org/10.1029/2008JC004803>.
- Morel, A., & Maritorena, S. (2001). Bio-optical properties of oceanic waters: A reappraisal. *Journal of Geophysical Research*, 106(C4), 7163–7180.
- Nelson, D.M., Treguer, P., Brzezinski, M.A., Leynaert, A., & Queguiner, B. (1995). Production and dissolution of biogenic silica in the ocean: Revised global estimates, comparison with regional data and relationships to biogenic sedimentation. *Global Biogeochemical Cycles*, 9, 359–372.
- Niang, A., Badran, F., Moulin, C., Crépon, M., & 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, 82–94.
- Niang, A., Gross, L., Thiria, S., Badran, F., & Moulin, C. (2003). Automatic neural classification of ocean colour reflectance spectra at the top of the atmosphere with introduction of expert knowledge. *Remote Sensing of Environment*, 86, 257–271.
- Raitos, D. E., Lavender, S. J., Maravelias, C. D., Haralambous, J., Richardson, A. J., & Reid, P. C. (2008). Identifying four phytoplankton functional types from space: An ecological approach. *Limnology and Oceanography*, 53(2), 605–613.
- Richardson, A. J., Risien, C., & Shillington, F. A. (2003). Using self-organizing maps to identify patterns in satellite imagery. *Progress in Oceanography*, 59(223), 239.
- Romero, O., Boeckel, B., Donner, B., Lavik, G., Fisher, G., & Wefer, G. (2002). Seasonal productivity dynamics in the pelagic central Benguela System inferred from the flux of carbonate and silicate organisms. *Journal of Marine Systems*, 37, 259–278.
- Rousseau, V., Lancelot, C., & Cox, D. (2006). *Current status of eutrophication in the Belgian coastal zone*. Brussels: Presses Universitaires de Bruxelles.
- Ryan, J. P., Polito, P.S., Strutton, P. G., & Chavez, F. P. (2002). Unusual large-scale phytoplankton blooms in the equatorial Pacific. *Progress in Oceanography*, 55, 263–284.
- Sathyendranath, S., Watts, L., Devred, E., Platt, T., Caverhill, C., & Maass, H. (2004). Discrimination of diatoms from other phytoplankton using ocean color data. *Marine Ecology Progress Series*, 272, 59–68.
- Steinberg, D. K., Carlson, C. A., Bates, N. R., Johnson, R. J., Michaels, A. F., & Knap, A. H. (2001). Overview of the USJGOFS Bermuda Atlantic Time-series Study (BATS): A decade-scale look at ocean biology and biogeochemistry. *Deep-Sea Research II*, 48, 1405–1447.
- Strutton, P. G., & Chavez, F. P. (2000). Primary productivity in the equatorial Pacific during the 1997–1998 El Niño. *Journal of Geophysical Research*, 105, 26089–26101.
- Subramaniam, A., Brown, C. W., Hood, R. R., Carpenter, E. J., & Capone, D.G. (2002). Detecting *Trichodesmium* blooms in SeaWiFS imagery. *Deep-Sea Research II*, 49, 107–121.
- Uitz, J., Claustre, H., Morel, A., & Hooker, S. B. (2006). Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll. *Journal of Geophysical Research*, 111, C08005. <http://dx.doi.org/10.1029/2005JC003207>.
- Vesanto, J., Himberg, J., Alhoniemi, E., & Parhankangas, J. (1999). Self-organizing map in Matlab: The SOM Toolbox. Proceedings of the Matlab DSP conference 1999. Finland: Espoo.
- Vidussi, F., Claustre, H., Manca, B. B., Luchetta, A., & 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), 19939–19956.
- Wassermann, P. D. (1989). *Neural computing theory and practice*. New York: Von Nostrand Reinhold.

- Weber, L. H., & El-Sayed, S. (1987). Contributions to the net, nano- and picoplankton standing crop and primary production in the Southern Ocean. *Journal of Plankton Research*, 9, 973–994.
- Werdell, P. J., & Bailey, S. W. (2005). An improved in-situ bio-optical data set for ocean color algorithm development and satellite data product validation. *Remote Sensing of Environment*, 98, 122–140.
- Zorn, S. R., Drewnick, F., Schott, M., Hoffmann, T., & Borrmann, S. (2008). Characterization of the 15 South Atlantic marine boundary layer aerosol using an aerodyne aerosol mass spectrometer. *Atmospheric Chemistry and Physics*, 8, 4711–4728.
- Zubkov, M. V., Sleigh, M.A., Burkill, P. H., & Leakey, R. J. G. (2000). Picoplankton community structure on the Atlantic Meridional transect: A comparison between seasons. *Progress in Oceanography*, 45, 369–386.

Annexe 2 :

**Différences entre les vecteurs référents correspondants aux
apprentissages testés sur différentes années (1998 à 2010)
(chapitre 3)**

Différences entre les vecteurs référents correspondants aux apprentissages testés sur différentes années (1998 à 2010)

Différences entre les 5 longueurs d'ondes des différentes années échantillonnées pour l'apprentissage de la carte auto-organisatrice. Les années sont comparées deux à deux pour chaque longueur d'onde (test t, p-values avec seuil de signification à 0,025). A noter qu'il n'y a pas de différences significatives

	412 nm	443 nm	490 nm	510 nm	555 nm
1998/1999	0,79198293	0,68629373	0,6906644	0,66203525	0,67963525
1998/2000	0,74477555	0,70291252	0,72841568	0,71705961	0,70355629
1998/2001	0,61726394	0,53310027	0,56144491	0,50419609	0,52846291
1998/2002	0,69308522	0,643617	0,65252438	0,65290895	0,65200256
1998/2003	0,88935439	0,90519816	0,88855052	0,89369117	0,8721341
1998/2004	0,8534754	0,8570381	0,8475779	0,85582372	0,83201367
1998/2005	0,70942366	0,93591158	0,54285227	0,70677807	0,99227346
1998/2006	0,7304043	0,72034144	0,74591619	0,75214689	0,72896607
1998/2007	0,61494077	0,53218644	0,56466573	0,51017069	0,53386996
1998/2008	0,6782441	0,6196005	0,6231131	0,62221103	0,62861331
1998/2009	0,80951205	0,84582668	0,84672184	0,86787438	0,8268091
1998/2010	0,82634289	0,85907139	0,86201098	0,88029046	0,84429281
1999/2000	0,95235888	0,98208751	0,95759221	0,93868838	0,97056957
1999/2001	0,812377	0,82462384	0,86120039	0,82054955	0,83426866
1999/2002	0,89613237	0,94725774	0,95784389	0,98419071	0,9683957
1999/2003	0,89967328	0,77894079	0,79707197	0,76703484	0,80155223
1999/2004	0,9355944	0,82620939	0,83640133	0,803563	0,84037805
1999/2005	0,52285403	0,63609597	0,83230111	0,96178389	0,6766232
1999/2006	0,93781571	0,96342224	0,93962307	0,90301796	0,94363462
1999/2007	0,81066482	0,82406052	0,86532518	0,82801707	0,84112877
1999/2008	0,88035213	0,92010447	0,92558368	0,95043558	0,94248121
1999/2009	0,98076991	0,83828	0,83780192	0,79275707	0,84601641
1999/2010	0,96283252	0,82469324	0,82230844	0,78031878	0,82832895
2000/2001	0,85728173	0,8073053	0,81795032	0,75990411	0,80321915
2000/2002	0,94279088	0,92973386	0,91535909	0,92387025	0,93879069
2000/2003	0,8515309	0,79606246	0,83708166	0,82427375	0,82805264
2000/2004	0,88738518	0,84357395	0,87722045	0,86170057	0,86765791
2000/2005	0,48127228	0,65178255	0,79020685	0,97887225	0,70018866
2000/2006	0,98544798	0,9813281	0,98181632	0,96375198	0,97271966
2000/2007	0,85572373	0,80671596	0,82200392	0,76722009	0,8100511
2000/2008	0,92681666	0,90265692	0,88300743	0,89006302	0,91266205
2000/2009	0,93275687	0,85564542	0,87857255	0,85048439	0,87335962
2000/2010	0,91458969	0,84202258	0,86286827	0,83782189	0,85537979
2001/2002	0,91410162	0,87941329	0,90392401	0,8390611	0,86657694
2001/2003	0,71577184	0,61811413	0,66229811	0,60172939	0,64182284
2001/2004	0,74945143	0,66162476	0,69966506	0,6353082	0,6782083
2001/2005	0,38135168	0,49261242	0,966861	0,78841331	0,52775999
2001/2006	0,87100005	0,78932125	0,80033854	0,72628082	0,77675001
2001/2007	0,99906276	0,99972414	0,99556488	0,99216916	0,99263855

	412 nm	443 nm	490 nm	510 nm	555 nm
2001/2008	0,9297179	0,90643807	0,93651109	0,87231367	0,89266624
2001/2009	0,79282031	0,6732875	0,70126369	0,62578258	0,68380642
2001/2010	0,77503456	0,66047611	0,68616737	0,61411743	0,66683228
2002/2003	0,79674928	0,73299763	0,7569005	0,75576387	0,77187072
2002/2004	0,83181342	0,77864255	0,79550445	0,79154573	0,81009348
2002/2005	0,44026113	0,59656528	0,87385789	0,94703915	0,64939861
2002/2006	0,95704467	0,91149145	0,89756716	0,8889045	0,91205605
2002/2007	0,91278842	0,87894819	0,90812509	0,84645414	0,87350325
2002/2008	0,98412673	0,97319411	0,96777152	0,96678204	0,9741598
2002/2009	0,87649384	0,79044905	0,79695593	0,78098131	0,81570002
2002/2010	0,85843102	0,77724564	0,7816408	0,76878507	0,79821676
2003/2004	0,96357738	0,95178196	0,95894787	0,96241744	0,95964381
2003/2005	0,60646004	0,84496192	0,63994989	0,80989136	0,86623626
2003/2006	0,83684864	0,81398304	0,85504166	0,85968723	0,85454889
2003/2007	0,71370739	0,6172934	0,66588047	0,60819718	0,64788036
2003/2008	0,78120269	0,70788881	0,72600847	0,72397961	0,74706761
2003/2009	0,91833777	0,94023496	0,95788021	0,97411485	0,95416776
2003/2010	0,9359905	0,95367281	0,97349676	0,98658543	0,97206263
2004/2005	0,57441048	0,79876983	0,67593155	0,8459031	0,82670013
2004/2006	0,87267552	0,86173122	0,89533763	0,89732019	0,89445112
2004/2007	0,74749496	0,66085219	0,70337948	0,64195769	0,68446484
2004/2008	0,81610405	0,7529968	0,76408481	0,75931265	0,78488267
2004/2009	0,95455553	0,98829105	0,9988548	0,98839315	0,99444769
2004/2010	0,97240013	0,998182	0,98541873	0,97587161	0,9875599
2005/2006	0,46869549	0,66824971	0,7731845	0,94398212	0,72520716
2005/2007	0,37869898	0,49171617	0,96249385	0,79557003	0,53309483
2005/2008	0,42809809	0,57401061	0,90553574	0,91429541	0,62635682
2005/2009	0,53655219	0,7881879	0,67750346	0,83513549	0,82157716
2005/2010	0,55029244	0,80079771	0,66292297	0,82293718	0,83879809
2006/2007	0,86949085	0,78870526	0,80434336	0,73344582	0,78348321
2006/2008	0,94101247	0,88450351	0,86533754	0,85537296	0,88602625
2006/2009	0,91810051	0,87379911	0,89665977	0,88595183	0,9001432
2006/2010	0,89986453	0,86014354	0,88094367	0,87327058	0,88209515
2007/2008	0,92846883	0,90601556	0,94077345	0,87979271	0,89966183
2007/2009	0,79102134	0,67253237	0,70497699	0,63236419	0,69008181
2007/2010	0,77316207	0,65970432	0,68983667	0,62064522	0,67302961
2008/2009	0,86066819	0,76475074	0,7655767	0,7489586	0,79047711
2008/2010	0,84260237	0,75165497	0,75038649	0,73687128	0,77311987
2009/2010	0,98201457	0,98649678	0,98430031	0,98752432	0,98202827

Annexe 3:

Premier test d'application de PHYSAT-SOM au capteur MODIS-AQUA à l'échelle globale

A. 3.1. Problématique

La méthode PHYSAT-SOM a été principalement développée, dans le cadre de cette thèse, au capteur SeaWiFS. Cette étape était nécessaire afin de tester la méthode et la comparer à la version 2008 de PHYSAT (Alvain et al., 2005 ; Alvain et al., 2008), dont les produits ont principalement été générés à partir de ce capteur. De plus, SeaWiFS a offert 13 années d'observations allant de 1997 à 2010, permettant ainsi de faire des études de variabilités saisonnières, annuelles et décennales.

Plusieurs autres capteurs, dédiés à l'étude de la couleur de l'océan, ont été lancés dans le but de compléter et de poursuivre les observations de SeaWiFS. On dispose ainsi actuellement des observations provenant, par exemple, de MERIS, VIIRS et MODIS (voir introduction pour les détails relatifs à chaque capteur). Il existe deux capteurs MODIS, « TERRA » et « AQUA », en service au moment de la rédaction de ce manuscrit, ce dernier ayant été lancé en 2002. On utilisera plus simplement la notation unique « MODIS » dans la suite du texte pour MODIS-AQUA, le capteur utilisé pour effectuer ce test préliminaire.

Depuis sa mise en service, MODIS a subi une série de calibrations sur un intervalle d'environ 6 mois à une année. Pour tous les capteurs, ces calibrations sont généralement faites après leur lancement, en utilisant des cibles bien déterminées (artificielles ou naturelles) pour corriger d'éventuels dérives dans les mesures: c'est ce qu'on appelle la « vicarious calibration » (Eplee et al., 1998, 2001, 2006 ; Wang et Franz, 2000 ; Werdell et al., 2006, 2007 ; Bailey et al., 2008 ; Franz et al., 2007). D'autres types de calibrations peuvent cependant être effectués. Par exemple, pour MODIS, des comparaisons des mesures de ce capteur avec celles de SeaWiFS, dont le rapport signal/bruit est plus élevé et les mesures à certaines longueurs d'onde sont plus stables (Mélin et al., 2007 ; Meister et al., 2012 ; Rapport IOCCG n°6). Par exemple, en 2009, MODIS a été calibré avec les données SeaWiFS, dû à une tendance à la baisse des valeurs de nL_w à 412 nm qui rendait la précédente approche de calibration obsolète. Cette calibration a été faite en multipliant les mesures MODIS par un facteur donné, dans le but d'ajuster les tendances temporelles de la dérive des mesures, principalement à 412 et 443 nm (Eplee et al., 2009 ; Meister et al., 2012). Les différences radiométriques entre MODIS et SeaWiFS doivent donc être prises en considération dans l'interprétation des divergences potentielles, observées entre les valeurs des paramètres

biologiques calculés, tels que la chlorophylle *a* (Hu et al., 2003 ; Zhang et al., 2006 ; Gregg et Casey, 2007 ; Lee et al., 2010), en plus de la différence des algorithmes bio-optiques utilisés pour le calcul de ces paramètres (Darecki et Stramski, 2004 ; Pan et al., 2010) et des erreurs liées aux corrections atmosphériques, potentiellement différentes entre les deux capteurs (Jamet et al., 2011 ; Goyens et al., 2013).

Le capteur SeaWiFS n'étant plus en service actuellement, les futures études seront confrontées à la nécessité d'adapter la méthode PHYSAT-SOM à d'autres capteurs (e.g. MODIS, MERIS qui a les mêmes bandes spectrales que SeaWiFS, excepté 555 nm qui est remplacée par 560 nm, ou encore, le capteur géostationnaire de la couleur de l'océan GOCI) ou algorithmes bio-optiques. Pour le capteur MODIS, certaines longueurs d'onde diffèrent de quelques nanomètres avec celles de SeaWiFS (412, 443, 488 au lieu de 490, 531 au lieu 510 et 551 au lieu de 555 nm). Généralement, pour chaque changement de capteur, il est donc nécessaire de considérer les différences inhérentes aux capteurs en plus des problématiques intrinsèques à la méthode.

Afin d'effectuer un premier test d'application de PHYSAT-SOM à MODIS, des anomalies spectrales, calculées à partir des luminances marines normalisées journalières de niveau 3 de MODIS (à 4 km de résolution), ont été classées avec SOM suivant la même procédure décrite précédemment pour SeaWiFS. L'adaptation principale consiste en la nécessité de calculer une LUT (Look-up Table) des luminances marine normalisées moyennes par pas de concentration en chlorophylle *a*. Cette étape doit être réalisée à chaque changement de capteur ou d'algorithme de façon à ce que le signal moyen des luminances ($nL_{w_{ref}}$) intègre les spécificités éventuelles du capteur considéré. Les cartes auto-organisatrices ont par la suite été utilisées pour classer automatiquement (apprentissage) les anomalies spectrales (R_a) des cartes satellites journalières de MODIS. Un premier essai de labellisation de ces anomalies spectrales a été effectué, de la même façon que pour SeaWiFS, en utilisant une base de données *in situ* de pigments correspondant aux mesures MODIS.

A.3.2. Application de PHYSAT-SOM au capteur MODIS

A.3.2.1. Seuil sur l'épaisseur optique des aérosols

De nombreuses études, portant sur la comparaison des mesures des produits d'aérosols issus des différents capteurs de la couleur de l'océan (i.e. coefficient d'Ångström et épaisseur optique des aérosols ou AOT), ont été effectuées. Certaines d'entre elles se sont focalisées sur la comparaison des données SeaWiFS et MODIS et leur évaluation respective à partir des données *in situ* (principalement à partir des stations de mesure AERONET ou Aerosol Robotic Network, se trouvant dans diverses zones du globe). Ainsi Mélin et al. (2007) ont montré que les valeurs d'AOT de MODIS à 869 nm (longueur prise en compte pour fixer un seuil dans le cadre de ce test) et celles de SeaWiFS à 865 nm étaient toutes deux caractérisées par un coefficient de corrélation (R^2) de 0,7 en comparaison avec les données AERONET à l'échelle globale. La même étude a montré que les données d'AOT des 2 capteurs étaient globalement en accord sur toutes les longueurs d'onde avec des différences inférieures à 20% avec les mesures *in situ*, aussi bien pour MODIS que SeaWiFS. Il est toutefois nécessaire de préciser que les pourcentages d'erreurs entre les mesures des deux capteurs et les données *in situ* varient selon la région considérée (de l'ordre de 12% à 26%, toutes longueurs d'onde moyennées), tels que le montrent Mélin et al. (2013). En revanche, ces différentes études convergent vers le fait que les données d'AOT des deux capteurs sont globalement en accord sur toutes les longueurs d'onde, en comparaison avec les données *in situ*.

Compte tenu des analyses précédentes, on a fixé le même seuil maximal pris en compte pour SeaWiFS (0,15) sur l'épaisseur optique des aérosols de MODIS à 869 nm. Toutefois, précisons que les différences de mesures d'AOT entre les deux capteurs, plus importantes dans certaines régions que d'autres (Mélin et al., 2007, 2013), engendrent des différences dans le calcul des anomalies spectrales. De ce fait, lors d'une application future de la méthode PHYSAT-SOM dans une région donnée, il conviendra de prendre en compte ces différences afin de fixer un seuil adéquat.

A.3.2.2. Apprentissage de SOM sur les données MODIS

Des cartes journalières de R_a , à 4 km de résolution (niveau 3), ont été générées à partir des mesures de nL_w en utilisant les valeurs de la LUT ($nL_{w_{ref}}$), calculées spécifiquement pour MODIS. Le jeu de données d'apprentissage sélectionné pour MODIS comprend 806 855 spectres de R_a à 5 longueurs d'onde du domaine visible : 412, 443, 488, 531 et 551 nm. Ce jeu de données a été obtenu en procédant, comme pour SeaWiFS, à la sélection de 2 jours par mois sur un intervalle d'une année de mesures, mais en prenant 1 pixel sur 20 (au lieu d'1 pixel sur 10 pour SeaWiFS). Cet échantillonnage vise à avoir un nombre de spectres dans le même ordre de grandeur de celui utilisé pour l'apprentissage avec les données de SeaWiFS, et ce, dans le but d'être proche en termes de précision de la variabilité spectrale des R_a , décrite dans le chapitre 4, et de pouvoir ainsi comparer les formes et amplitudes des spectres référents d'anomalies spectrales résultant de l'apprentissage de SOM avec les données MODIS et SeaWiFS. Dans le cadre de ce test, les pixels de MODIS présentant le maximum de recouvrement avec ceux de SeaWiFS ont été pris en compte (passage de 9 km à 4 km de résolution). Cependant, l'hétérogénéité spatiale entre les pixels engendre des différences entre les données prises en compte pour les deux capteurs (Zibordi et al., 2006 ; Mélin et Zibordi, 2007 ; Djavidnia et al., 2010 ; Mélin, 2011). D'autres techniques d'atténuation des erreurs liées à la variabilité spatiale, utilisées, par exemple, lors des exercices de fusionnement des données de plusieurs capteurs (data merging) pourraient être utilisées dans le futur, lors d'études plus poussées (voir le Rapport IOCCG n°6, Mélin et Zibordi, 2007 ; Mélin et al., 2007 et le projet Globcolour, par exemple).

Une carte auto-organisatrice carrée de 10x10 neurones a également été utilisée lors de ce test. Comme pour SeaWiFS, les spectres de R_a sont donc classés en 100 groupes distincts selon leur forme et amplitude, permettant ainsi d'avoir une discrétisation très précise de l'ensemble des observations. Tel que précisé dans la problématique, les longueurs d'onde utilisées avec MODIS diffèrent de quelques nanomètres avec celles de SeaWiFS. D'une façon générale, ceci engendre une différence entre les données satellite qui doivent subir une correction (appelée « band shift error correction ») quand il s'agit d'exploiter simultanément des données de plusieurs capteurs (Mélin et al., 2011). Dans le but de quantifier les différences globales entre les valeurs d'anomalies spectrales issues des mesures des deux capteurs, des corrélations entre les valeurs des 100 spectres référents résultant de l'apprentissage de SOM avec les R_a de MODIS et de SeaWiFS ont été effectuées sur les 5

longueurs d'onde prises en compte deux à deux (412-412, 443-443, 490-488, 510-531 et 555-551). Les spectres référents, résumant toute l'information contenue dans le jeu de données, permettent ainsi d'avoir une vue globale sur les différences existant entre les différentes longueurs d'onde des deux capteurs. Les résultats sont affichés sur la figure A.3.1. Cette dernière montre que les valeurs de Ra sont corrélées entre les deux capteurs. Le plus faible coefficient de corrélation n'est toutefois pas associé aux longueurs d'onde décalées des deux capteurs (490 et 488 ; 510 et 531 ; 555 et 551), mais à 443 nm (0,86).

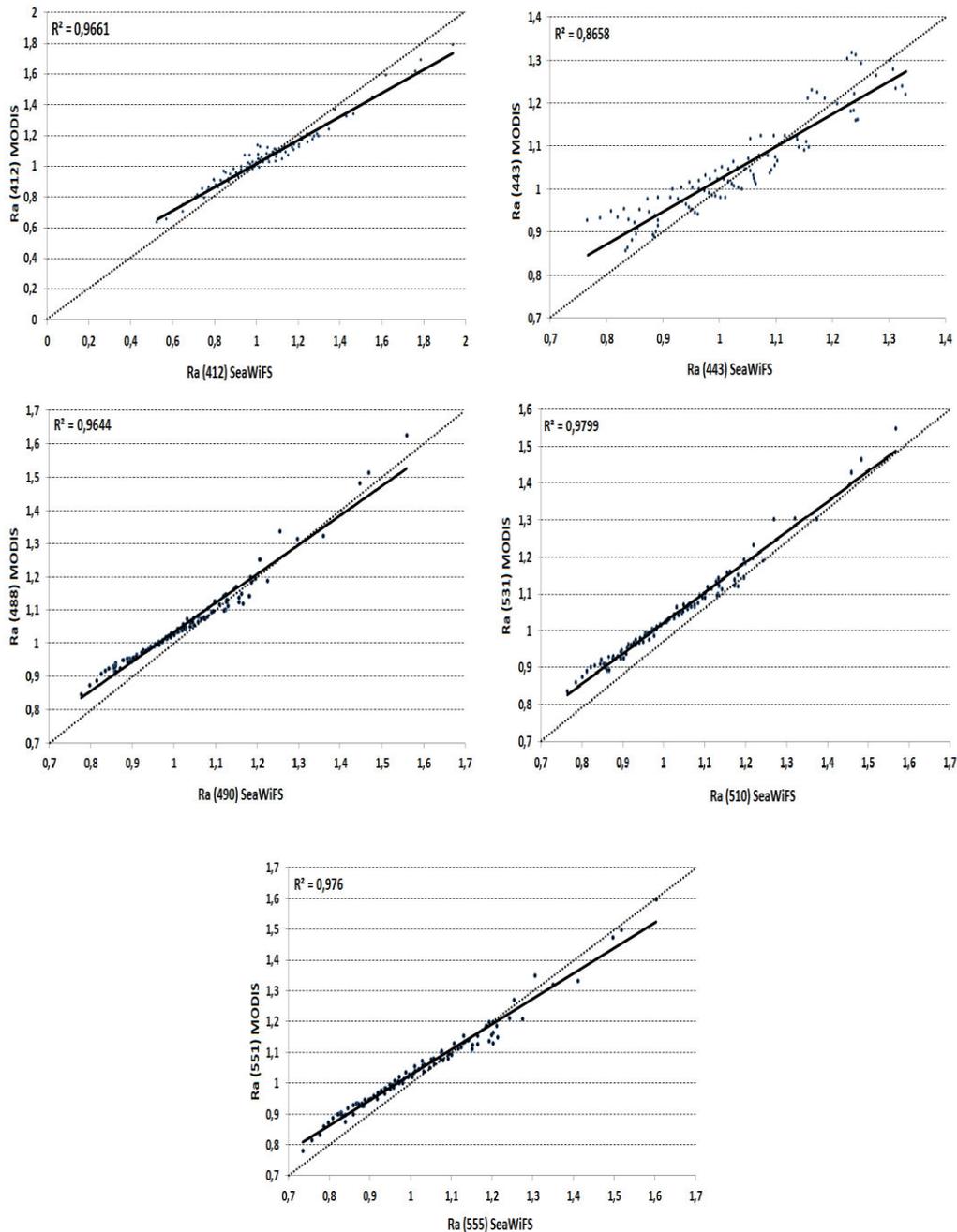


Figure A.3.1. Comparaison entre les valeurs des spectres référents résultant de l'apprentissage avec MODIS (412, 443, 488, 531 et 551 nm en ordonnées) et SeaWiFS (412, 443, 490, 510 et 555 nm en abscisses) pour chacune des 5 longueurs d'ondes prises en compte deux à deux

Les faibles différences observées entre les spectres référents peuvent être liées à de nombreux facteurs inhérents aux propriétés des capteurs et aux différents algorithmes utilisés lors des chaînes de traitements (e.g. algorithmes de corrections atmosphériques, algorithmes bio-optiques utilisés pour calculer la concentration en chlorophylle *a*). Par exemple, la divergence des mesures de réflectances des deux capteurs engendrent des écarts entre les concentrations en chlorophylle *a* et luminances marines normalisées mesurés par les deux capteurs (Hu et al., 2003; Zibordi et al., 2006 ; Gregg et al., 2007 ; Lee et al., 2010 ; Pan et al., 2010). Ces différentes études montrent toutefois que les biais entre les mesures des capteurs dépendent sensiblement de la zone considérée. La comparaison des anomalies de R_a permettra, en perspective, d'envisager des corrections de biais, nécessaires lorsque l'on cherchera à relier un spectre de R_a du capteur SeaWiFS à un spectre MODIS. La figure A.3.2 montre les intervalles de valeurs des spectres référents résultants de l'apprentissage de SOM avec SeaWiFS et MODIS sur les 5 longueurs considérées pour chaque capteur.

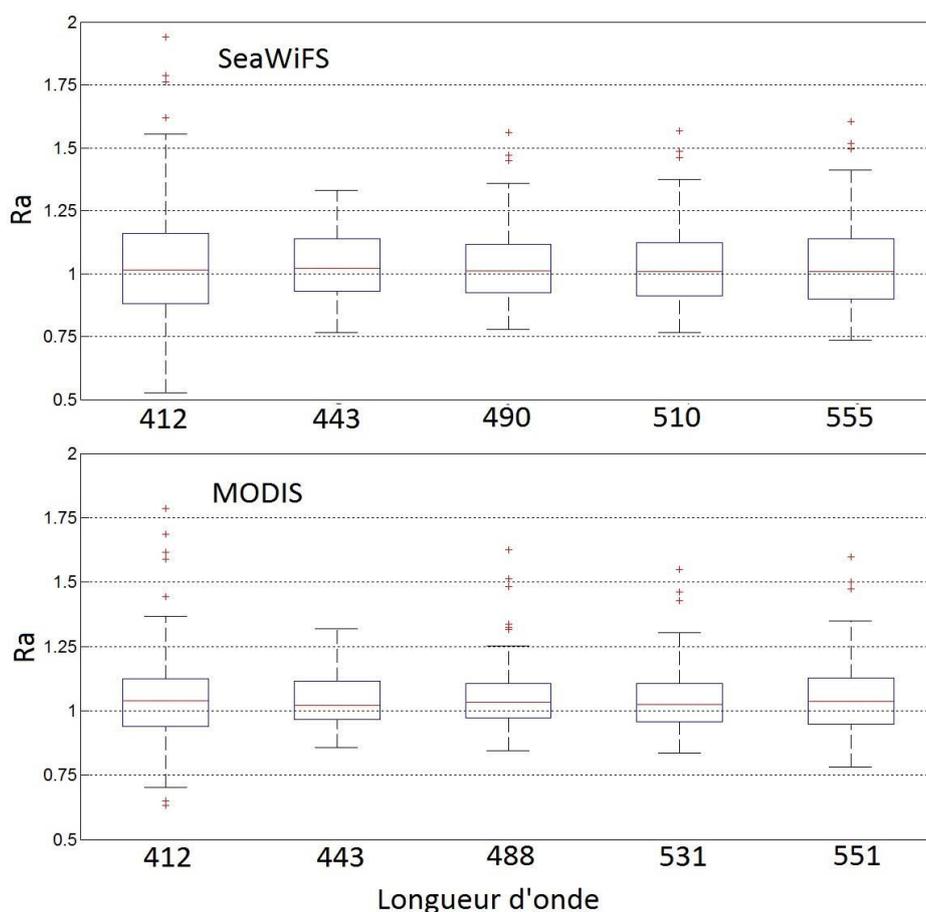
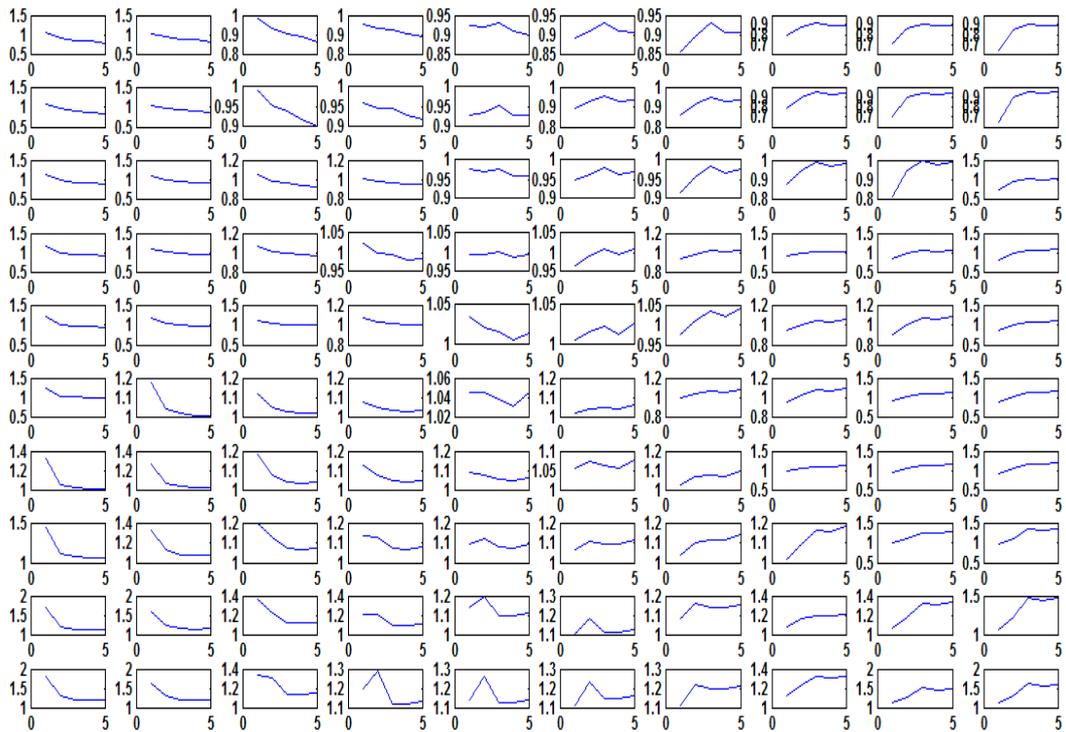


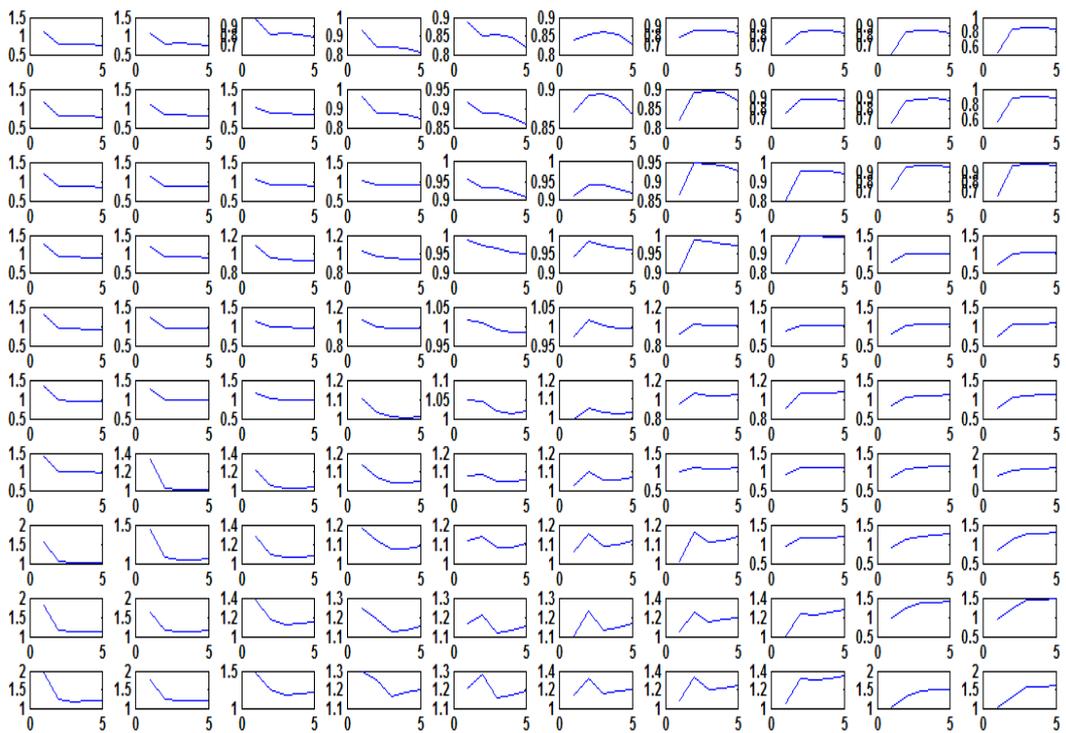
Figure A.3.2. Intervalles de valeurs des spectres référents résultants de l'apprentissage de SOM avec SeaWiFS et MODIS sur les 5 longueurs d'onde considérées pour chaque capteur. Les barres verticales représentent les écarts-type, les lignes rouges horizontales désignent les valeurs médianes, les extrémités inférieures et supérieures représentent respectivement les quartiles 25 et 75% et les croix désignent les valeurs aberrantes

On observe que les valeurs de Ra présentent un écart relativement grand à 412 nm pour SeaWiFS par rapport à MODIS (entre 0,5 et 1,55 pour SeaWiFS ; entre 0,75 et 1,35 pour MODIS). Cet écart de valeurs est relativement plus faible aux autres longueurs d'onde. Par exemple, à 443 nm, les valeurs de Ra sont situées entre 0,75 et 1,33 pour SeaWiFS et entre 0,80 et 1,31 pour MODIS. On observe aussi que, pour les deux capteurs, la plus grande variabilité des valeurs de Ra se situe à 412 nm. Ceci est dû à la plus grande sensibilité des anomalies spectrales dans les courtes longueurs d'onde (Alvain et al., 2012).

Bien que dépendante de la dimension (5 dimensions pour 5 longueurs d'onde dans le cas étudié) des données et de la configuration de la carte auto-organisatrice choisie, la phase d'apprentissage demeure généralement rapide (quinze minutes environ avec cette configuration de SOM). Un nouvel apprentissage semble donc judicieux à faire lorsqu'on souhaite appliquer la méthode à un autre capteur afin de tenir compte des changements des valeurs de Ra sur certaines longueurs d'onde et de prendre en compte les différences radiométriques inhérentes aux différents capteurs. Les spectres référents résultant de l'apprentissage de SOM avec les Ra de MODIS sont affichés sur la figure A.3.3 qui montre, comme pour SeaWiFS, l'ordre topologique induit par l'algorithme SOM à la fin de la phase d'apprentissage. La classification des spectres de Ra de MODIS avec la carte SOM de 10x10 neurones donne une répartition équilibrée des données sur la carte, ce qui dénote d'une bonne qualité d'apprentissage. Les vecteurs référents classent un maximum de 15075 (neurone 91) et un minimum de 2630 spectres (neurone 19). La fonction de voisinage impose que les spectres voisins au niveau de la carte auto-organisatrice soient similaires en termes de forme et d'amplitude et vice-versa (figure A.3.3).



MODIS



SeaWiFS

Figure A.3.3. Représentation des spectres référents correspondant aux Ra associés à chaque neurone à la fin de la phase d'apprentissage pour MODIS (en haut) et SeaWiFS (en bas). Pour chaque unité, l'axe des abscisses représente les longueurs d'onde et l'axe des ordonnées représente les Ra. Les axes des ordonnées de chaque case n'ont pas été mis à la même échelle afin de mettre en évidence les différentes formes de spectres.

A.3.2.3. Labellisation des anomalies MODIS selon la procédure définie précédemment pour SeaWiFS

L'apprentissage de la carte auto-organisatrice à partir des anomalies spectrales, calculées pour le capteur MODIS, a montré que les plus grandes différences entre les spectres référents (par rapport à ceux trouvés pour SeaWiFS) se situent dans les courtes longueurs d'onde (essentiellement à 443 nm) ($R^2=0,86$). Dans les plus grandes longueurs d'onde, les anomalies spectrales sont caractérisées par des valeurs relativement proches entre les 2 capteurs ($R^2 > 0,95$). En appliquant la même méthode suivie précédemment pour le capteur SeaWiFS (identification des groupes à partir des inventaires de pigments et méthode de labellisation, décrites dans le chapitre 4), les spectres référents MODIS ont été labellisés en utilisant des informations *in situ* colocalisées avec les mesures de ce capteur. Cette base de données diffère donc par rapport à celle utilisée précédemment (période d'observation, résolution des images et horaires de passage du satellite différents). . De nombreuses campagnes ont été prises en compte, parmi lesquelles, NOMAD, OOMPH, OISO, ICOTA, HALIFAX (décrites précédemment dans le chapitre 4) mais aussi, KEOPS (Kerguelen Ocean and compared Plateau Study, Armand et al., 2008) dans la région des Kerguelen, de même que la campagne MELVILLE-2011 dans l'Atlantique Sud (Scripps Institution of Oceanography). Les observations coïncidentes comprennent 1124 mesures correspondant aux nanoeucaryotes, 450 pour *Prochlorococcus*, 151 pour SLC et 59 pour les diatomées

Bien que le nombre de mesures coïncidentes soit plus élevé que celui trouvé pour SeaWiFS selon les critères utilisés (1784 pour MODIS contre 759 pour SeaWiFS), elles sont géographiquement moins étendues (figure A.3.4). Elles ne couvrent par exemple pas le Pacifique Sud et la Mer des Caraïbes du fait de l'absence de mesures spatiales MODIS coïncidentes avec les données GeP&CO. L'Atlantique Nord est, de ce fait aussi, moins représenté par ces données avec 41,3 % de données de moins. D'autres régions du globe, telles que les côtes Ouest et Est Nord-Américaines, l'Atlantique Sud près des côtes de Patagonie ou l'Océan Indien, sont relativement bien couvertes par les mesures *in situ* colocalisées avec celles de MODIS.

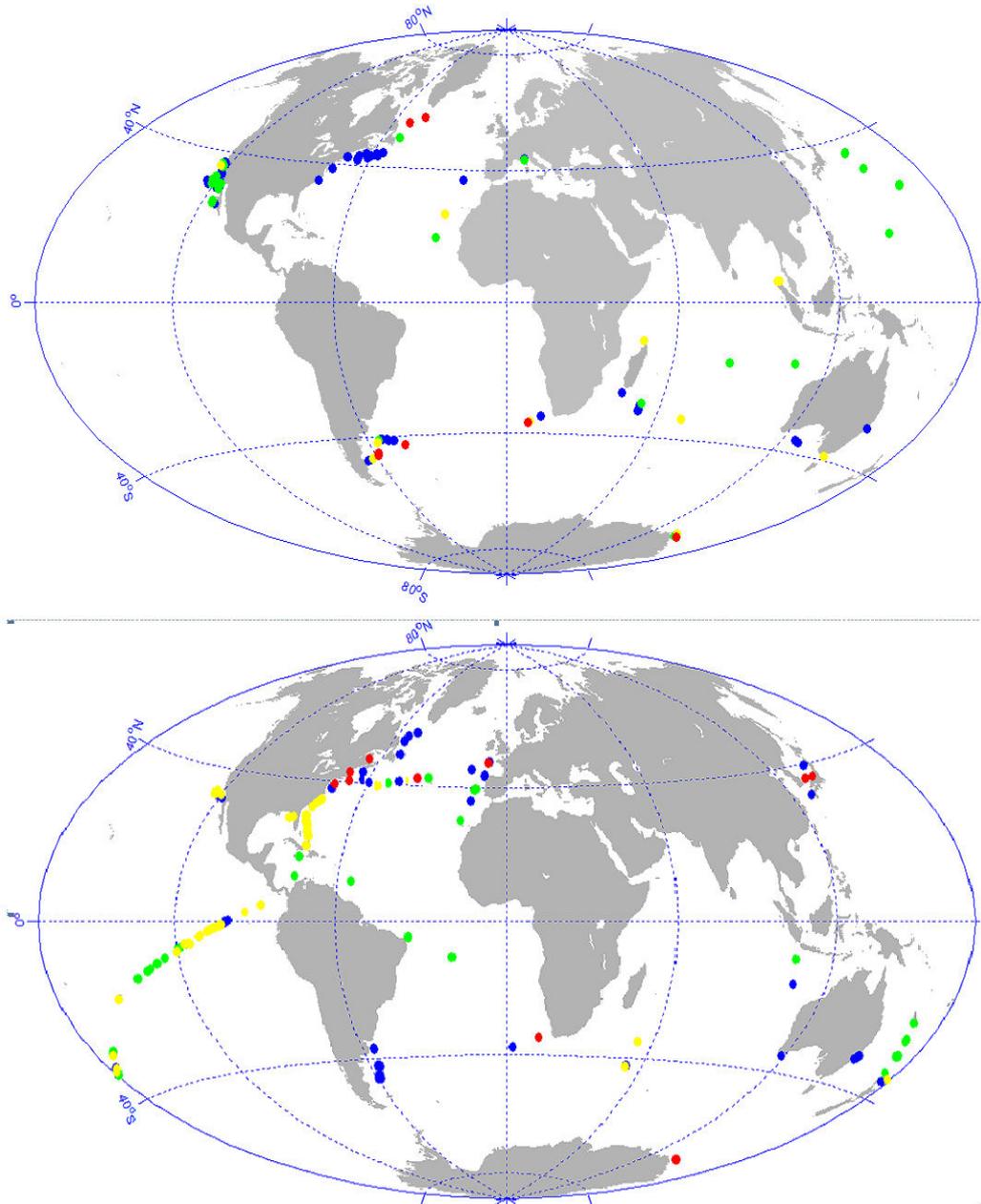


Figure A.3.4. Distribution géographique des mesures *in situ* coïncidentes aux données MODIS (en haut) et SeaWiFS (en bas). Le même code couleur utilisé dans les chapitres précédents désigne les 4 groupes de phytoplancton dominants dans chaque échantillon.

La fréquence relative des spectres de Ra affectés à chaque neurone (calculée suivant la même méthode décrite dans le chapitre 4) est montrée sur la figure A.3.5 (en haut) avec, pour comparaison, les fréquences relatives des spectres de SeaWiFS (en bas). Avec cette base de labellisation, 38 neurones au total sont labellisés (contre 44 avec la base de données *in situ* utilisée précédemment), dont 33 portant un label correspondant à l'un des 4 groupes de phytoplancton dominants.

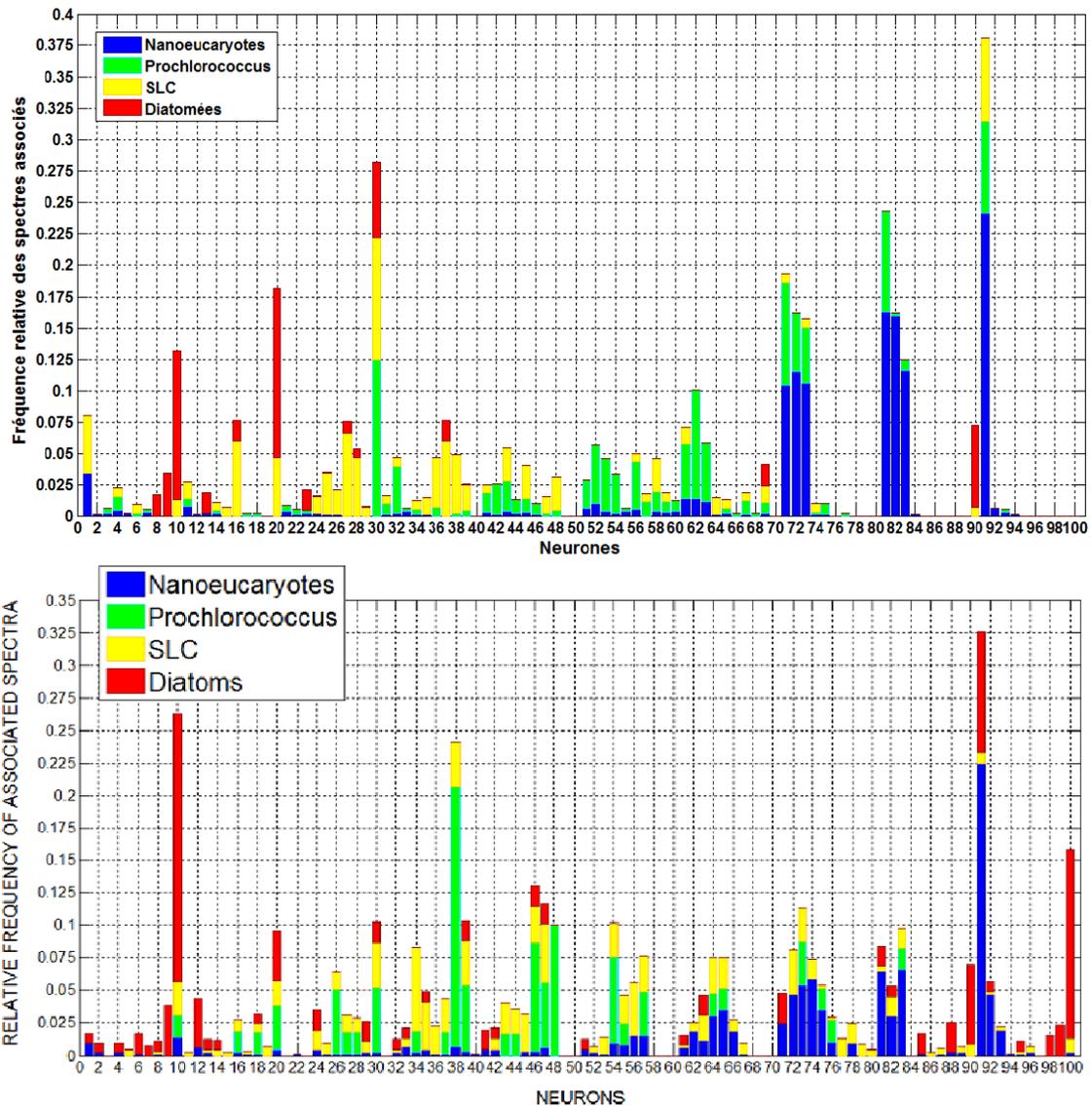


Figure A.3.5. Fréquence relative des labels associés aux spectres de Ra répartis sur les 100 neurones de la carte auto-organisatrice pour MODIS (en haut) et SeaWiFS (en bas)

Une vue d'ensemble des deux parties de la figure A.3.5 révèle que, malgré la différence en termes de couverture géographique des bases de données utilisées et les légères différences entre les valeurs des spectres référents résultants de l'apprentissage avec les données de chacun des deux capteurs, les labels restent globalement situés dans les mêmes zones de la carte auto-organisatrice, même si les fréquences relatives des spectres associés aux différents neurones changent. Ce résultat n'est pas surprenant puisque, tel qu'expliqué dans le chapitre 4, c'est la présence, l'absence et la proportion des signaux spécifiques à chaque groupe phytoplanctonique qui vont déterminer à la fois si le neurone est labellisé ou pas et, si c'est le cas, quel sera son label. Une information « manquante » ou « additionnelle » (selon si on a des données *in situ* ou pas dans une zone géographique donnée) peut

significativement changer l'information globale classée par chaque neurone, ce qui va modifier son étiquette (label). Les signatures spectrales des 4 groupes de phytoplancton identifiés à partir de la base de données in situ MODIS sont affichées sur la figure A.3.6.

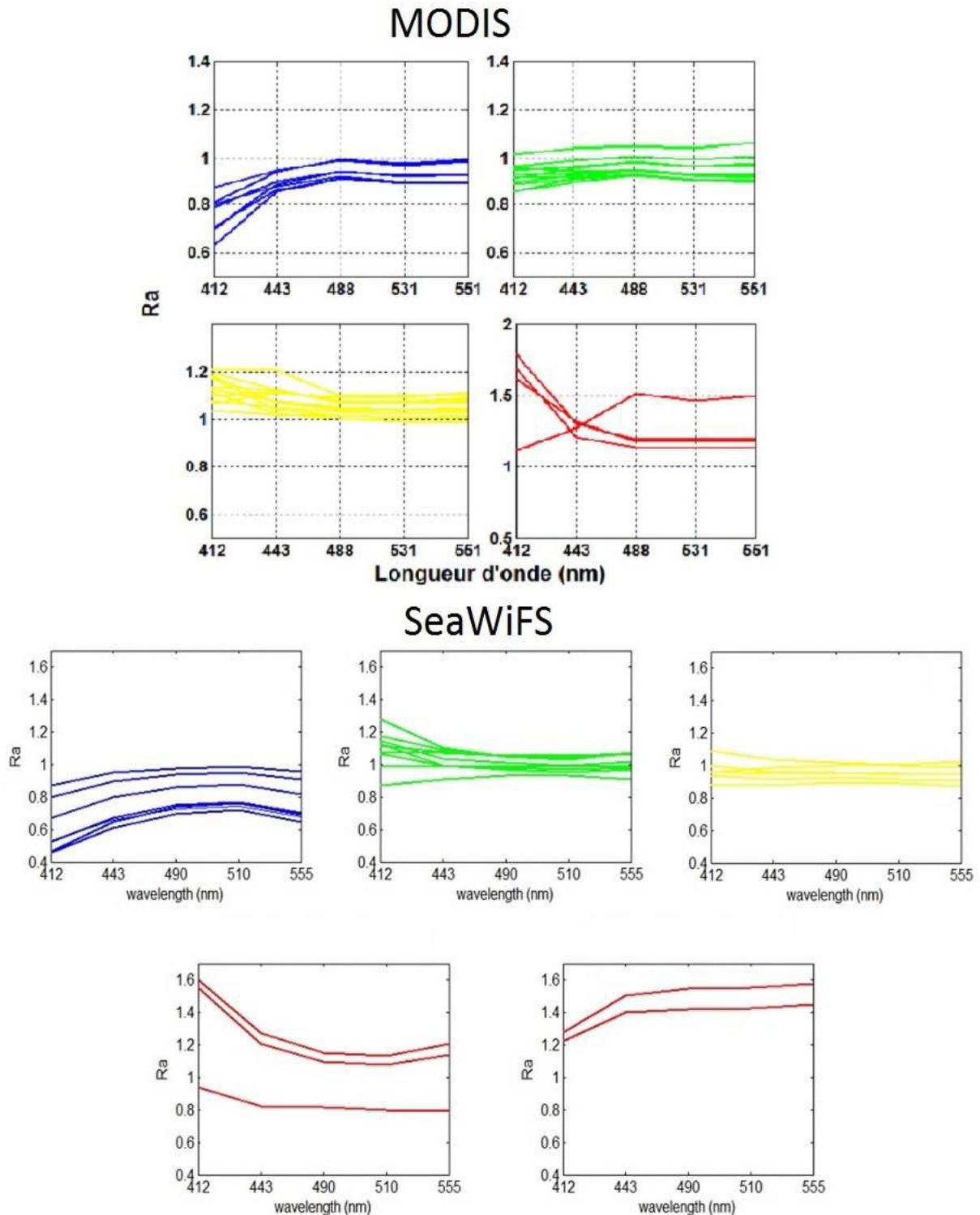


Figure A.3.6. Signatures spectrales des groupes de phytoplancton identifiés à partir de la base de données in situ coïncidentes avec les mesures du capteur MODIS (en haut) et SeaWiFS (en bas), via la projection sur la carte SOM. Chaque groupe est représenté par les spectres référents des neurones auxquels il a été affecté. Le même code couleur utilisé dans les chapitres précédents désigne les 4 groupes de phytoplancton

Les formes générales des spectres (enveloppes spectrales) restent sensiblement similaires à celles trouvées avec SeaWiFS pour les 4 groupes de phytoplancton, même si de faibles écarts d'amplitude peuvent être observés étant donné qu'on change de capteur et de base de données de labellisation. Les nanoeucaryotes sont toujours caractérisés par des valeurs de R_a relativement faibles (en dessous de 1) avec une hausse entre 412 et 443 nm. De même, *Prochlorococcus* et SLC sont affectés à des spectres relativement neutres (plats). On remarque toutefois un écart dans l'amplitude des spectres référents de SLC qui sont, cette fois, supérieurs à 1 sur toutes les longueurs d'onde (alors que certains spectres affichaient des valeurs en dessous de 1 avec l'apprentissage et la labellisation précédents). Les diatomées ont été affectées à 4 neurones, dont 3 sont caractérisés par un spectre référent ayant une forte décroissance entre 412 et 443 nm et un autre caractérisé par un spectre référent similaire aux deuxième type de spectre de diatomées discuté précédemment (caractérisé par une hausse de R_a entre 412 et 488 nm). En suivant la même méthode que celle utilisée précédemment pour SeaWiFS, ces signatures spectrales peuvent à présent être exploitées pour étudier la distribution spatio-temporelle des groupes de phytoplancton à l'échelle globale à partir des cartes journalières de R_a .

A.3.2.4. Observations globales préliminaires des groupes de phytoplancton avec MODIS

La carte auto-organisatrice, labellisée avec la base de données *in situ* correspondant aux mesures MODIS, a été utilisée pour produire des cartes journalières, synthèses mensuelles et climatologies pluriannuelles de la distribution des principaux groupes de phytoplancton dominants à l'échelle globale. Les mêmes méthodes, utilisées pour élaborer des synthèses à partir des produits journaliers (chapitre 4), ont été appliquées pour MODIS. Des climatologies de distribution des groupes de phytoplancton ont été générées sur une période d'observation commune de SeaWiFS et MODIS entre 2003 et 2009. Les résultats sont présentés sur la figure A.3.7. Cette dernière montre des cartes représentant des synthèses climatologiques de la distribution des 5 groupes de phytoplancton, générées entre 2003 et 2009 à partir de l'archive MODIS (à gauche) et SeaWiFS (à droite) pour 4 mois. En termes de couverture, on observe que les pixels non labellisés (du fait des critères sur la concentration en chlorophylle a et l'épaisseur optique des aérosols) sont généralement situés au niveau des mêmes zones entre les deux capteurs (Principalement au niveau des gyres subtropicales oligotrophes et au large des côtes Ouest Africaines du fait de la présence massive d'aérosols désertiques). Par ailleurs, les cartes générées avec MODIS sont caractérisées par des

structures relativement plus compactes (denses) par rapport à celles générées avec SeaWiFS, du fait de l'augmentation de la résolution spatiale.

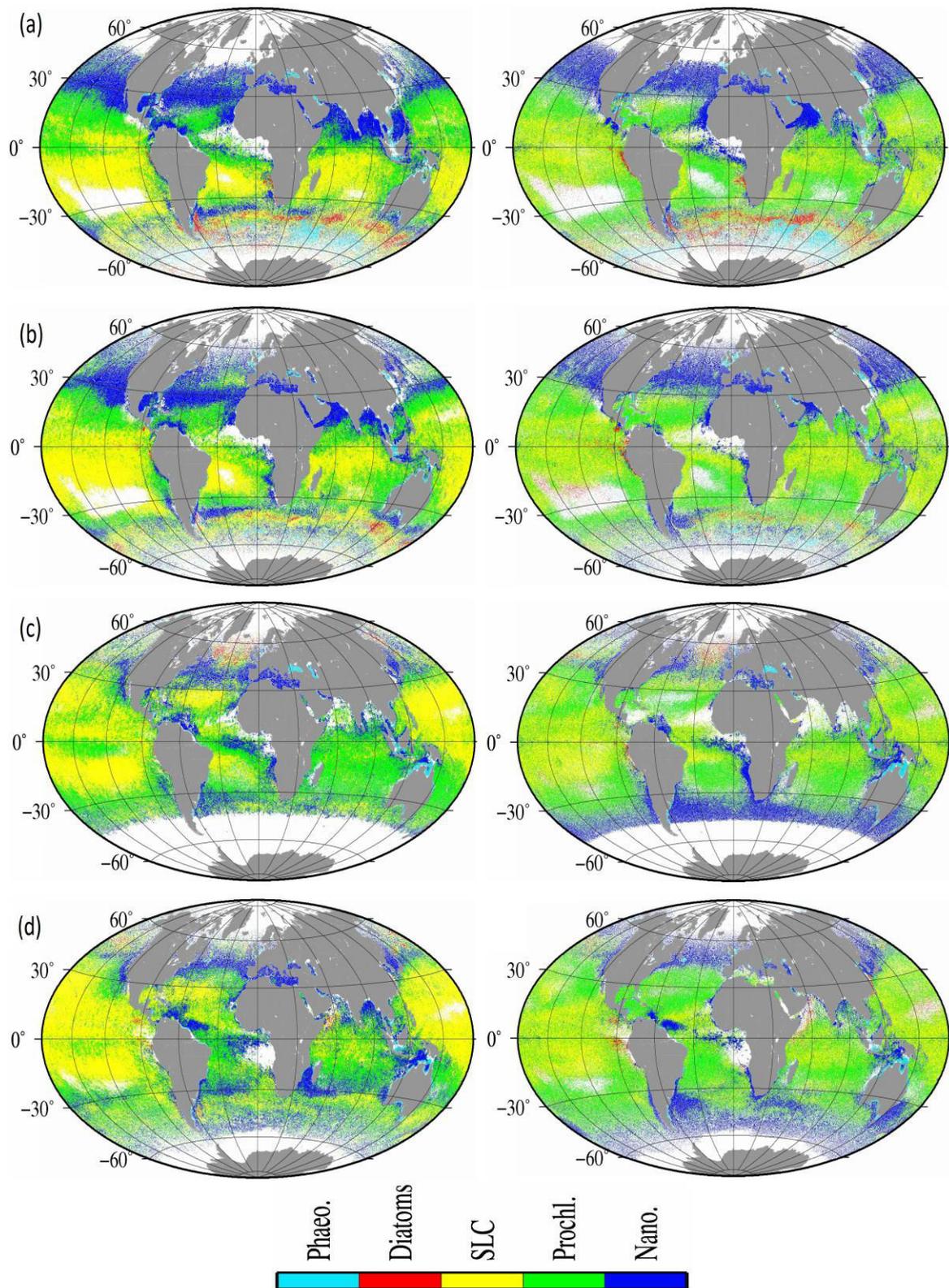


Figure A.3.7. Cartes représentant des synthèses mensuelles de la distribution des 5 groupes de phytoplancton générées entre 2003 et 2009 à partir de l'archive MODIS (à gauche) et SeaWiFS (à droite) pour les mois de Janvier (a), Mars (b), Juin (c) et Septembre (d)

On observe globalement un bon accord entre les résultats obtenus par PHYSAT-SOM à partir des mesures des deux capteurs, avec toutefois des différences dans certaines zones. Les nanoeucaryotes sont toujours présents dans les moyennes et hautes latitudes. Ils sont particulièrement abondants au niveau de l'hémisphère Nord durant l'hiver. Leur répartition affiche toutefois quelques différences entre les deux capteurs dans l'hémisphère Sud durant les mois de Juin et Septembre (figure A.3.7 c et d). En effet, ils sont relativement plus abondants dans cette région durant le mois de Juin d'après les cartes SeaWiFS par rapport à MODIS. Durant le mois de Septembre, ils sont plus dominants dans les cartes de MODIS vers 30° Sud, notamment au niveau de l'Atlantique et de l'Océan Indien. Les cyanobactéries (*Prochlorococcus* et SLC) sont présentes dans les basses à moyennes latitudes selon la saison pour les deux capteurs. La dominance des SLC est toutefois relativement plus marquée dans les cartes générées avec MODIS par rapport à SeaWiFS dans certaines régions telles que le Pacifique subéquatorial. Notons que la base de données de labellisation utilisée pour les deux capteurs est significativement différentes, en particulier dans le Pacifique Equatorial et Sud, zones pour lesquelles on a peu d'information *in situ* pour MODIS par rapport à SeaWiFS (absence des données GeP&CO). Les diatomées et *Phaecystis-like* sont particulièrement détectés dans les eaux de l'Océan Austral par les deux capteurs durant le mois de Janvier (figure A.3.7 a).

A l'échelle globale, les structures de groupes sont relativement bien conservées entre les deux capteurs, malgré la différence des bases de données *in situ* utilisées pour labelliser SOM pour chaque capteur. Cependant, cette étude très préliminaire devra être complétée à l'avenir pour répondre en profondeur aux problématiques suivantes :

- Quels sont les biais engendrés par l'utilisation d'une nouvelle base d'apprentissage sur les spectres référents (MODIS par rapport à SeaWiFS) ? Serait-il possible, en corrigeant ces biais, de retrouver des spectres référents plus proches de ceux trouvés précédemment pour SeaWiFS ? Pour répondre à ces questions, on pourrait, par exemple, utiliser des mesures *in situ* de nLw, calculer les spectres de Ra correspondants, puis les comparer aux spectres de Ra correspondants aux mesures coïncidentes des deux capteurs (sur leurs périodes de recouvrement).
- Même si les spectres référents générés par l'apprentissage de SOM avec SeaWiFS et MODIS montrent des différences dans les courtes longueurs d'onde

(essentiellement à 443 nm), leur topologie (emplacement et ordre des spectres sur la carte) reste similaire entre les deux apprentissages (figure A.3.3). Est-il donc nécessaire de refaire une labellisation spécifique à chaque capteur ou peut-on, sans faire d'erreurs trop importantes, utiliser la labellisation des spectres du capteur initial ?

- Quel serait alors l'impact d'un changement de base de labellisation (due à l'absence de données sur les mêmes périodes) sur la détection des groupes de phytoplancton à l'échelle globale ? cet impact pourrait, par exemple, être quantifié en procédant, pour les deux capteurs, à une labellisation des spectres référents avec une base de données *in situ* de pigments commune aux deux capteurs (sur leur période de recouvrement).
- Quel est l'impact du changement de la résolution spatiale sur le taux de couverture en pixels labellisés et sur la détection des groupes de phytoplancton ?
- Quelles erreurs fait-on selon le choix de la procédure choisie, c'est-à-dire, nouvel apprentissage ou non, nouvelle labellisation ou non ?

ANNEXE 4 :

**Tables de calcul (Look-up Tables) des anomalies spectrales
correspondant à chacune des classes d'eau étudiées en
Manche et Mer du Nord**

**Exemples de distributions spatio-temporelles de quelques
anomalies spectrales en Manche et Mer du Nord**

Table des valeurs de luminances marines normalisées correspondant à chaque valeur de concentration en chlorophylle a (nLw ref) pour la classe d'eau n°= 2 (matériel particulaire d'origine biologique + matières dissoutes)

Chl-a (mg.m ⁻³)	nLw (412)	nLw (443)	nLw (490)	nLw (510)	nLw (555)	Nbr mesures
1,5	0,02003333	0,31456668	0,46750002	0,57270003	0,49396669	30
1,75	0,04707143	0,37835716	0,56810241	0,68263098	0,61386431	420
2	0,08895514	0,39532587	0,59838702	0,70109818	0,65855528	1783
2,25	0,15529495	0,48027275	0,73977353	0,84673012	0,82558048	3916
2,5	0,34533963	0,77254278	1,22691626	1,37302524	1,40507992	15453
3	0,70257551	1,25505559	1,91336202	2,13677468	2,29271699	75111
3,5	0,86813315	1,43823227	2,11977316	2,38131872	2,66243512	146009
4	0,7978448	1,28995244	1,88990565	2,14696173	2,53147821	299652
6	0,45749259	0,7136536	1,07095676	1,24367894	1,59239139	144714
8	0,3375935	0,47184534	0,70336826	0,82959086	1,15265926	39332
10	0,29705479	0,37907567	0,58092597	0,68659399	1,02016618	14368
15	0,31792063	0,30580266	0,47621888	0,52605736	0,8512007	1814
20	0,45850973	0,33359225	0,53456313	0,53150002	0,95514568	206
25	0,38501564	0,25601564	0,40384377	0,39220315	0,74270316	64
30	0,28991668	0,15825001	0,25825001	0,22691668	0,48050002	12

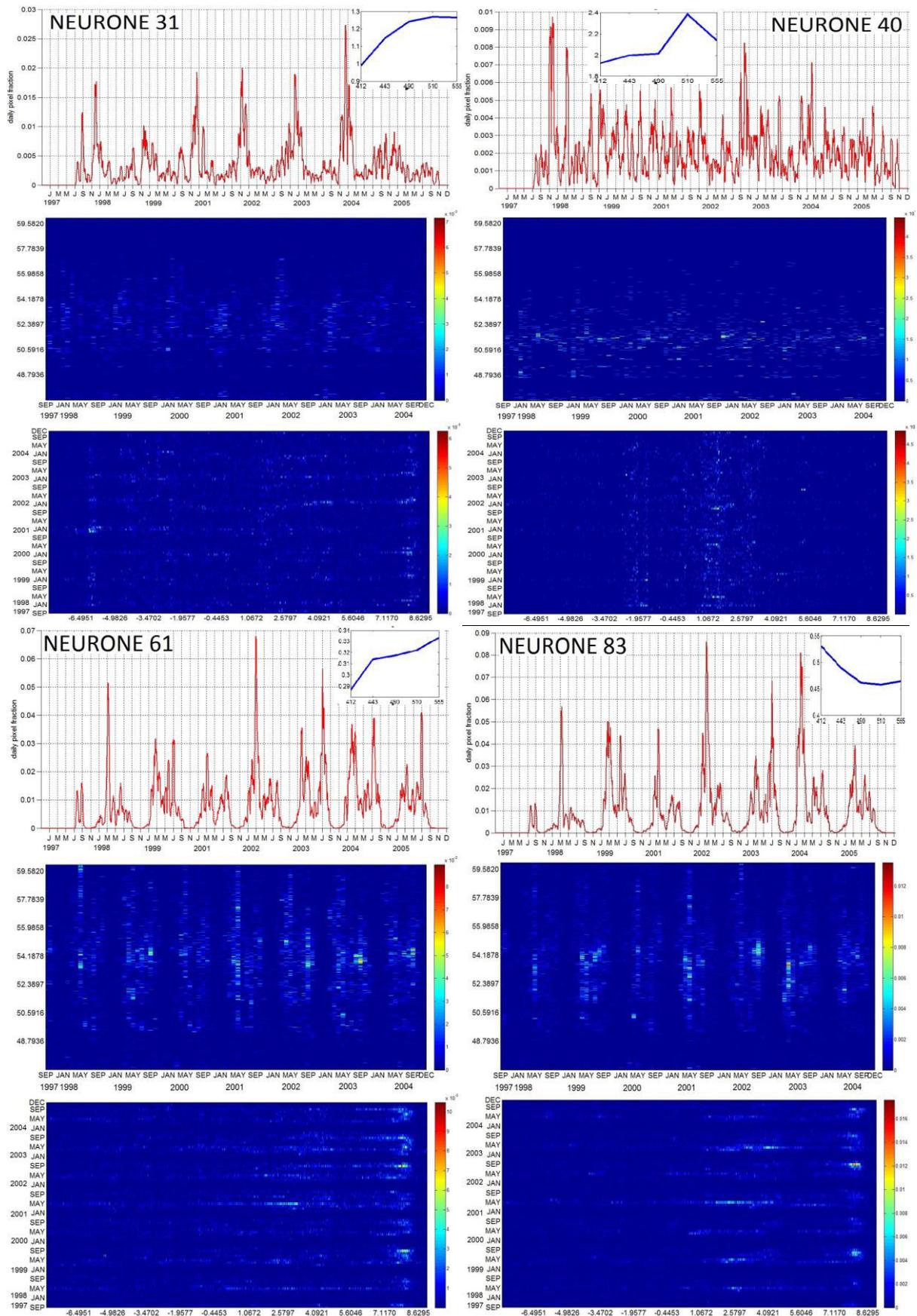
Table des valeurs de luminances marines normalisées correspondant à chaque valeur de concentration en chlorophylle a (nLw ref) pour la classe d'eau n°= 4 (phytoplancton + matières dissoutes)

Chl-a (mg.m ⁻³)	nLw (412)	nLw (443)	nLw (490)	nLw (510)	nLw (555)	Nbr mesures
0,7	0,02445	0,34895001	0,46795003	0,56920003	0,35515002	20
0,8	0,05381901	0,325543	0,45980545	0,54009505	0,35566065	221
0,9	0,14972077	0,40592844	0,57899719	0,61757267	0,43515239	1411
1	0,41681957	0,70616463	1,01477942	0,97885411	0,75719935	49228
1,25	0,57823742	0,92384465	1,305381	1,274867	1,03797017	367271
1,5	0,640085	0,9910418	1,38313965	1,38161938	1,1802863	580706
1,75	0,68742691	1,04211263	1,44466466	1,4715156	1,31840979	650444
2	0,73678886	1,10359195	1,5243261	1,57719943	1,47653834	632806
2,25	0,76768714	1,146488	1,58492472	1,66204068	1,61708444	566684
2,5	0,78455514	1,16025111	1,60431803	1,70319112	1,72511054	626882
3	0,71595166	1,01371894	1,38508659	1,48114661	1,56917251	355912
3,5	0,4900912	0,68053718	0,93977839	1,01111858	1,12556295	108043
4	0,3141039	0,46479676	0,66764876	0,72504278	0,85356443	72030
6	0,21747024	0,34552624	0,50622378	0,54702444	0,70373348	23718
8	0,26481057	0,34579118	0,49832163	0,51052776	0,71905822	2217
10	0,45654079	0,46026183	0,67263093	0,63809874	0,98924039	233
15	0,38900003	0,266	0,38000003	0,287	0,59100002	1

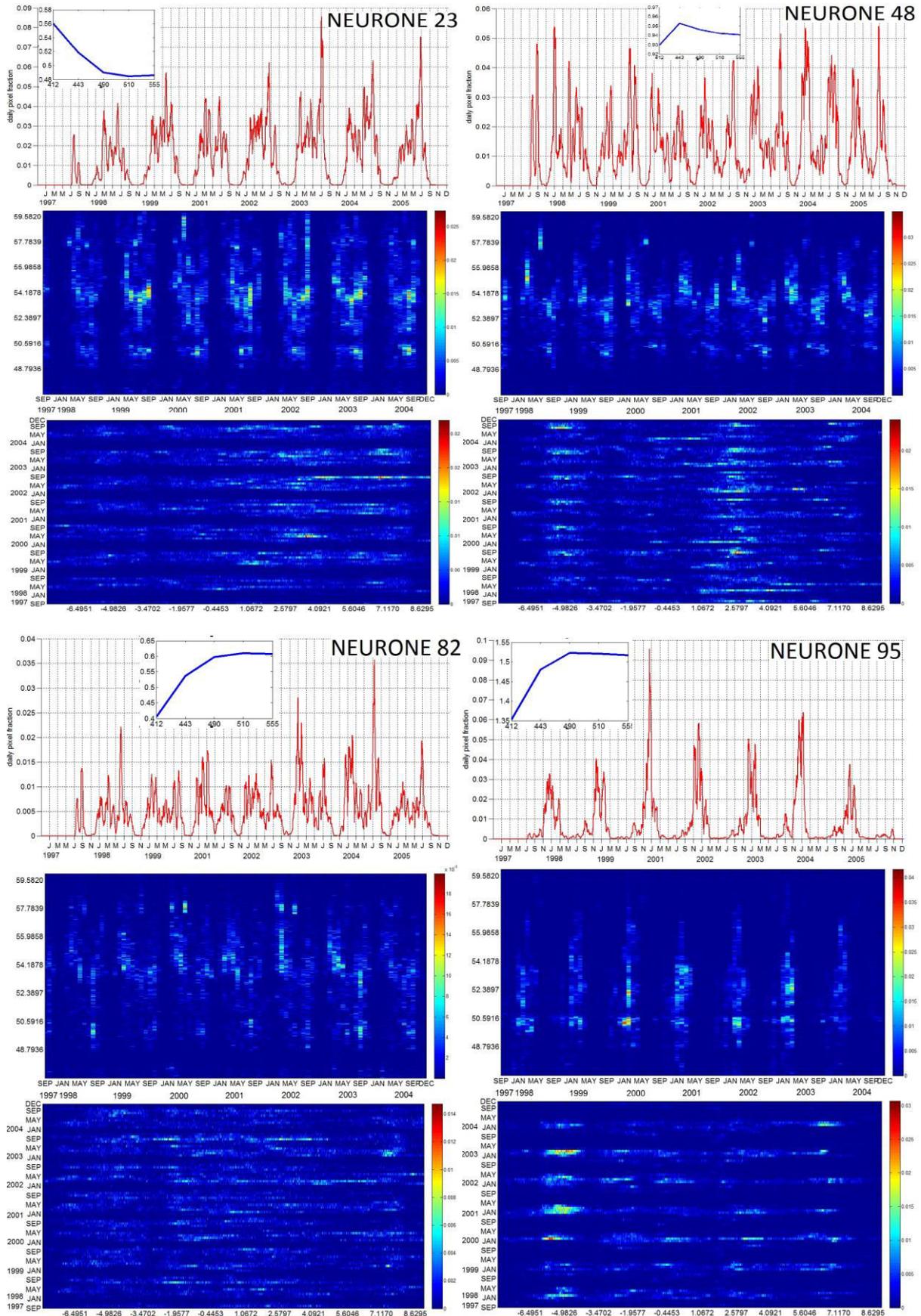
Table des valeurs de luminances marines normalisées correspondant à chaque valeur de concentration en chlorophylle a (nLw ref) pour la classe d'eau n°=5 (eaux du large turbides)

Chl-a	nLw (412)	nLw (443)	nLw (490)	nLw (510)	nLw (555)	Nbr mesures
0,01	1,05238081	1,13154991	1,13199319	1,01440389	0,11640042	16328
0,03	1,08192657	1,14468041	1,09920958	0,969672	0,13953995	15583
0,04	1,07696271	1,12831835	1,07254392	0,94456487	0,15750281	9642
0,05	1,08670521	1,14775984	1,08721096	0,96520209	0,17469952	10270
0,06	1,10711473	1,15252106	1,08283739	0,95682849	0,18856643	10900
0,07	1,1154234	1,1530511	1,08387544	0,95303208	0,20645171	18200
0,09	1,10763109	1,15025587	1,07422639	0,94320967	0,22617433	20761
0,1	1,12745904	1,16619497	1,07652177	0,94164946	0,24909673	24308
0,12	1,11658751	1,15567299	1,05746584	0,91708056	0,26930394	39286
0,14	1,13928857	1,17807517	1,0600797	0,90942292	0,29885626	51058
0,16	1,18135494	1,21128533	1,07410009	0,90936758	0,33116106	68254
0,18	1,19112195	1,22441402	1,07272689	0,89514378	0,35817678	93794
0,2	1,18532173	1,22558864	1,07553764	0,88378175	0,38153792	132747
0,22	1,16180703	1,21221083	1,0725797	0,86876572	0,39945346	189303
0,24	1,11913004	1,18288086	1,05867995	0,85264197	0,41143469	257778
0,26	1,07857123	1,1518524	1,04160754	0,83987279	0,42154339	326493
0,28	1,03307444	1,11616767	1,0211824	0,82563415	0,42851261	394583
0,3	0,910541	1,01713298	0,97039292	0,79836838	0,44986869	2,97E+06
0,5	0,74945751	0,88421444	0,91083836	0,7847502	0,49798364	5,91E+06
0,7	0,64466943	0,79289788	0,86982672	0,7858195	0,54662181	3,57E+06
0,8	0,60378645	0,7573434	0,85688901	0,79454883	0,58104277	1,98E+06
0,9	0,58059962	0,7374625	0,84965992	0,80105443	0,60533496	1,75E+06
1	0,57200684	0,73532622	0,86269711	0,82783976	0,65031761	2,62E+06
1,25	0,57556943	0,7372283	0,86920432	0,85249579	0,70470036	2,78E+06
1,5	0,5738044	0,73741017	0,88045749	0,8820758	0,76863643	1,88E+06
1,75	0,56019345	0,72929587	0,88702756	0,90374267	0,82490728	1,35E+06
2	0,54486545	0,719299	0,89163679	0,92212804	0,87719806	983185
2,25	0,52700131	0,70243254	0,8842488	0,9277617	0,91574789	714590
2,5	0,51064423	0,67956772	0,86453526	0,92300157	0,94792129	738088
3	0,51289816	0,67900456	0,8713707	0,94940871	1,02171029	625286
3,5	0,51149837	0,67715184	0,87813685	0,97389668	1,09668853	402549
4	0,47364753	0,65538401	0,88881201	1,01138399	1,20958235	576930
6	0,42164658	0,67595565	1,01360036	1,19932821	1,55778107	499566
8	0,32165518	0,53617794	0,84519993	1,02577799	1,43599004	251731
10	0,26508413	0,40309137	0,62863164	0,77993313	1,17239517	212168
15	0,24523286	0,31239067	0,45934191	0,58100949	0,94842354	133006
20	0,2198312	0,25742345	0,38499257	0,50223596	0,88066356	58665
25	0,20406818	0,2150136	0,34120452	0,45107286	0,83748824	25672
30	0,20105756	0,18375705	0,30325309	0,40575986	0,78355424	6429

Exemples de variations spatio-temporelles de quelques anomalies spectrales pour la classe d'eau n°2 (Matériel particulaire d'origine biologique+dissous)



Exemples de variations spatio-temporelles de quelques anomalies spectrales pour la classe d'eau n°4 (phytoplancton + matières dissoutes)



Exemples de variations spatio-temporelles de quelques anomalies spectrales pour la classe d'eau n°5 (eaux turbides du large)

