



**HAL**  
open science

# Vers l'identification d'assemblages phytoplanctoniques depuis l'espace à travers le couplage innovant d'analyses des luminances et de données in-situ, perspectives pour l'étude des habitats et des réponses des assemblages

Anne-Hélène Rêve-Lamarche

► **To cite this version:**

Anne-Hélène Rêve-Lamarche. Vers l'identification d'assemblages phytoplanctoniques depuis l'espace à travers le couplage innovant d'analyses des luminances et de données in-situ, perspectives pour l'étude des habitats et des réponses des assemblages. Océan, Atmosphère. Université de Lille, 2018. Français. NNT: . tel-02404562

**HAL Id: tel-02404562**

**<https://hal.science/tel-02404562>**

Submitted on 11 Dec 2019

**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.



Distributed under a Creative Commons Attribution 4.0 International License



Université de Lille

Laboratoire d'Océanologie et de Géosciences UMR LOG 8187

Équipe Télédétection et Hydrodynamique

École doctorale Sciences de la Matière,  
du Rayonnement et de l'Environnement

Spécialité Sciences de la Terre et de l'Univers

Vers l'identification d'assemblages phytoplanctoniques depuis  
l'espace à travers le couplage innovant d'analyses des  
radiances et de données *in-situ*, perspectives pour l'étude des  
habitats et des réponses des assemblages

Thèse de doctorat présentée par

**Anne-Hélène RÊVE-LAMARCHE**

Soutenue publiquement, le 12 février 2018, devant le jury composé de

Directeurs de thèse	Grégory BEAUGRAND	DR, UMR LOG 8187 CNRS, ULNE, ULCO
	Séverine ALVAIN	CR, UMR LOG 8187 CNRS, ULNE, ULCO
Rapporteurs	Pascal MORIN	DR, Institut polaire français Paul-Emile Victor
	David ANTOINE	PR, Curtin University
Examineurs	Mélilotus THYSSEN	CR, MIO Institut Méditerranéen d'Océanologie
	Bertrand LUBAC	MCF, Université de Bordeaux
Invité	Antoine MANGIN	Directeur ACRI-HE

et présidé par Pascal MORIN



Université de Lille

Laboratoire d'Océanologie et de Géosciences UMR LOG 8187

Équipe Télédétection et Hydrodynamique

École doctorale Sciences de la Matière,  
du Rayonnement et de l'Environnement

Spécialité Sciences de la Terre et de l'Univers

Vers l'identification d'assemblages phytoplanctoniques depuis  
l'espace à travers le couplage innovant d'analyses des  
radiances et de données *in-situ*, perspectives pour l'étude des  
habitats et des réponses des assemblages

Thèse de doctorat présentée par

**Anne-Hélène RÊVE-LAMARCHE**

Soutenue publiquement, le 12 février 2018, devant le jury composé de

Directeurs de thèse	Grégory BEAUGRAND	DR, UMR LOG 8187 CNRS, ULNE, ULCO
	Séverine ALVAIN	CR, UMR LOG 8187 CNRS, ULNE, ULCO
Rapporteurs	Pascal MORIN	DR, Institut polaire français Paul-Emile Victor
	David ANTOINE	PR, Curtin University
Examineurs	Mélilotus THYSSEN	CR, MIO Institut Méditerranéen d'Océanologie
	Bertrand LUBAC	MCF, Université de Bordeaux
Invité	Antoine MANGIN	Directeur ACRI-HE

et présidé par Pascal MORIN



*Je dédie cette thèse*

*À mon  Papa ,*  
*À ma famille.*



## Résumé

L'identification des groupes phytoplanctoniques depuis l'espace présente un fort intérêt pour de nombreux domaines de recherche en lien avec l'étude des océans. Afin de fournir des observations attendues par la communauté scientifique, différentes approches ont été développées, et permettent d'obtenir des informations en lien avec la taille ou des groupes fonctionnels de phytoplancton depuis les mesures effectuées par des radiomètres à bord de satellites. Cette thèse a eu pour cadre le développement d'une méthode initiée en 2005, nommée PHYSAT. Cette dernière est initialement basée sur la mise en évidence de relations empiriques entre des anomalies de luminance et la présence d'un groupe phytoplanctonique dominant. Alvain *et al.* (2008) ont ainsi permis l'identification de cinq groupes phytoplanctoniques lorsqu'ils sont dominants, *via* la détermination empirique d'anomalies de luminance dont les amplitudes et formes spectrales, aux longueurs d'ondes visibles des capteurs de couleur de l'océan, sont spécifiques à la présence de ces groupes. Une étude théorique (Alvain *et al.*, 2012) a, par ailleurs, permis d'apporter une première explication théorique à la présence de ces anomalies de luminance, en tenant compte des propriétés optiques inhérentes (coefficient d'absorption, de rétrodiffusion et absorption des matières dissoutes colorées). Ces résultats théoriques formulaient l'hypothèse que l'identification des cas de dominance uniquement représentait une sous-utilisation du potentiel d'informations porté par les anomalies de luminance. Dans cette thèse, nous avons donc évalué le potentiel de développement de la méthode PHYSAT au-delà de la dominance, vers l'identification empirique d'assemblages phytoplanctoniques. Cela a nécessité, en premier lieu, une adaptation de la chaîne de traitement de la méthode. En effet, les anomalies de luminance étaient classées jusqu'à la dernière version (Ben Mustapha *et al.*, 2014) uniquement à l'aide d'une carte auto-organisatrice des caractéristiques spectrales (classification mathématique). Cette approche ne prenait pas pleinement en compte les caractéristiques spatio-temporelles des anomalies. Il a donc été proposé une nouvelle approche, plus adaptée, basée sur les outils issus de la phénologie, associés à une approche innovante concernant l'analyse des données *in-situ* elle-même. Ensuite, le potentiel de développement de PHYSAT vers l'identification empirique des assemblages a été évalué dans une étude globale à partir de données de pigments biomarqueurs, ainsi que dans une étude régionale en Mer du Nord, à partir des données CPR. Nos résultats montrent, pour la première fois, la possibilité de détecter empiriquement des cas phytoplanctoniques mixtes, au-delà de la dominance. Enfin, des pistes de recherche pour poursuivre la compréhension et le développement de PHYSAT sont proposées.





## Abstract

Over the past years, a large number of new approaches in the domain of ocean color have been developed to study phytoplankton ocean processes based on phytoplankton size classes and functional types. One of these methods, named PHYSAT, currently allows for the qualitative identification of five main phytoplankton groups from remote sensing measurements (Alvain *et al.*, 2008). This method established empirical relationships between *in-situ* dominant phytoplankton groups and specific ocean color radiance anomalies in open ocean waters. These radiance anomalies are defined by specific shape and amplitude according to the dominant phytoplankton group in waters. The theoretical explanation of PHYSAT (Alvain *et al.*, 2012) showed that radiance anomalies are induced by a combination of inherent optical properties (absorption and backscattering coefficients, absorption of colored dissolved organic matters). This study suggested that the empiric identification of dominant cases represents an under-exploitation of the radiance anomalies. The work presented here addresses the potential development of the PHYSAT method toward the empiric identification of phytoplankton assemblages. A global ocean (based on biomarker pigments data) and a regional sea's (North Sea, based on CPR abundance data) applications are presented. Our results show, for the first time, the ability of PHYSAT to identify phytoplankton assemblages using empirical relationships with radiance anomalies, beyond the identification of dominance cases only. This work was facilitated by the use of previous mathematical classification of radiance anomalies in terms of shape and amplitude (using a self-organizing map, Ben Mustapha *et al.* (2014), coupled with a new classification that take into account the phenology of radiance anomalies types. This work emphasizes the potential of PHYSAT's radiance anomalies when using new classification tools, and detailed *in-situ* databases. Future directions are listed to understand the empirical relationships observed in this work beyond the current theoretical knowledge.

## Remerciements

---

Je remercie David ANTOINE et Pascal MORIN (Président de Jury) qui ont évalué ce travail de thèse en qualité de rapporteurs. Merci également aux examinateurs, Melilotus THYSSEN, Bertrand LUBAC et à la participation d'Antoine MANGIN.



Je remercie les organismes qui m'ont financée, à savoir la région Hauts-de-France et le Centre National d'Études Spatiales (CNES), sous la gestion du Centre National de la Recherche Scientifique (CNRS).



Je remercie le Centre National d'Études Spatiale (CNES) qui ont financé les travaux de cette thèse à travers le programme TOSCA/PHYTOHAB.



Je remercie le directeur de laboratoire, François SCHMITT, de m'avoir accueillie dans ses locaux pour réaliser ma thèse (Wimereux et Lille), ainsi que les tutelles du laboratoire, l'Université de Lille (ULNE), l'Université du Littoral Côte d'Opale (ULCO) et le Centre National de la Recherche Scientifique (CNRS). Merci également à l'équipe Télédétection et Hydrodynamique, à laquelle j'ai été intégrée, sous la direction de Hubert LOISEL et Cédric JAMET.

Je remercie l'École Doctorale Sciences de la Matière, du Rayonnement et de l'Environnement (SMRE), mon directeur des études doctorales, Nicolas TRIBOVILLARD, et l'Université de Lille de m'avoir permis de réaliser cette thèse dans de bonnes conditions.

Je remercie Grégory BEAUGRAND, mon directeur de thèse, et Séverine ALVAIN, mon encadrante, de m'avoir permis de réaliser cette thèse avec des conditions de travail impeccables. Merci pour votre patience, votre compréhension, vos conseils, et tout ce que je n'ai plus la force d'écrire, et surtout, pour votre soutien dans les moments difficiles de cette thèse.

Je remercie les personnes qui ont formé mon comité de thèse : Marie-Fanny RACAULT, Eric GOBERVILLE et Patrick LOHODEY. Vos questions et remarques ont été très utiles pour mûrir les projets de cette thèse.

Je remercie Christophe GUINET et Guillemette LABADIE, ainsi que toutes les personnes ayant collaborées au projet éléphant de mer au Centre d'Étude Biologique de Chizé, ou ailleurs. Merci pour l'accueil au CEBC, de m'avoir permis de présenter mes résultats lors des journées Éléphants de Mer. Je suis enthousiaste quant-à l'aboutissement de ce projet commun.

Je remercie la plate-forme de calcul scientifique CALCULCO, gérée par le SCoSI de l'ULCO (Service COmmun du Système d'Information de l'Université du Littoral Côte d'Opale).

Merci à toutes les personnes avec qui j'ai pu travailler de près ou de loin, et celles qui m'ont donné des conseils avisés.

---

Il paraît qu'il est coutume d'entamer cette section de remerciements par quelque chose comme "le manuscrit de thèse est le fruit d'un travail personnel long et fastidieux". Je ne déroge pas à la règle, et ne vais pas nier le dur labeur, mais ce travail de thèse est avant-tout un travail d'équipe. Sans toutes les personnes que je vais remercier ensuite (et celles que je vais oublier de remercier, désolée), cette thèse n'aurait pas été ce qu'elle est aujourd'hui.

Séverine (et hop!, la partie un peu moins formelle), je *vous* remercie pour tout (et ça commence à en faire beaucoup ...). D'abord, merci de m'avoir donné une chance, toute jeune et venant d'un monde parallèle, pour mon stage de Master 1 (merci Dorothée VINCENT pour le tuyau ☺), ce stage vers l'infini et au-delà qui m'a montré combien j'aime la programmation, le tricafoilage des données et tout ça (même si je suis un peu nulle). Cela s'en est suivi d'un stage de Master 2 très cool, où j'ai pu rencontrer Marie-Fanny RACAULT vers de nouvelles aventures phénologiques; pour en arriver là. J'ai beaucoup apprécié toutes ces années de travail, de discussions de tout et de rien au café-thé. J'ai vécu cette thèse comme n'importe qui devrait vivre une thèse (enfin, on se passerait bien de certaines épreuves quand même...), du travail, de l'humanité, de la compréhension, des encouragements, des chats, et plus récemment, des licornes et des caniches roses. 🦄. Pour tout ça, mais aussi je l'espère pour de nouvelles aventures, Merci.

Merci Michelle MARQUANT 🦄 et Valérie PLOUVIN pour toute la gestion administrative, et votre joie de vivre maternelle (bisous bisous!).

Ce travail de thèse n'aurait pas été possible sans l'aide informatique de Régis SION, qui, avant mon passage sur Mac récemment (huhu, moi aussi), m'a beaucoup aidé, surtout avec ma machine virtuelle qui était diablement possédée.

David DESSAILLY, que l'on m'a présenté il y a 5 ans comme "l'homme à cloner", n'a pas failli à sa réputation! Merci David de coder le jour ce dont j'ai rêvé la nuit, et pour ton aide tout au long de cette thèse (mais aussi avant, et sûrement encore après). Tu es le geek que j'aurais aimé être, peut être avec moins de barbe ☺ (au lieu de ça, je joue à *Plague*, où je détruis le monde pour me venger de ne pas être douée...).

Je dois également remercier beaucoup de personnes du laboratoire, pour ne pas dire TOUTES les personnes du laboratoire. Merci à Arnaud (👉!!), F-P, Lucie, Lucille, Maxime, Olmo, Mohammed, Bing, Alice, Dorothée, Felipe, Vincent et Vincent, Hubert, Cédric, Mahmoud, Natacha, Xavier, Michel, Laurent, Simon, Fabien, Charles, Coco, et d'autres, et toute une série de stagiaires sympatiques mais dont la liste est trop longue.

Un ÉNORME merci à William 🦄, avec qui j'ai partagé mon bureau un long moment (ou un trop (?) long moment?). Maintenant, il ne reste que *Némo*. Évidemment, je ne te remercie pas pour les discussions de football avec Maxime ⚽. Bien que tu ne sois qu'un "collègue", mon ami, j'ai hâte de pouvoir partager de nouveau de tout et de rien avec toi, autour de délicieux sushis, et de parler fille avec ta femme, Caroline (Monsieur "Je ne me marierai jamais", je t'en veux toujours, et je le grave ici à jamais ☺). (PS : tu as vu, maintenant je sais faire des ☺, 👉, 🦄 en L<sup>A</sup>T<sub>E</sub>X).

Merci à ma Poukette extrapionnelle Manon 🦄! Finalement, on y est arrivé toutes les deux, malgré les lapins géants bouffeurs de planète (c'est pas très malin de sa part)! Les observatOIres auraient pu nous prévenir que tout se passerait bien ☺, on aurait évité du stress inutile! Je t'aime ma Poukette d'amour.

Quand on commence à partir dans tous les sens, c'est qu'est venu le temps des distractions! Merci à Fabien Olicard (🇫🇷), Céline Holynski (🇫🇷), Nicole Ferroni (🇫🇷), Aldebert (🇫🇷), Monsieur Phi (🇫🇷), Nota Bene (🇫🇷), La statistique expliquée à mon chat (🇫🇷), Science Étonnante (🇫🇷), Micode (🇫🇷), et tous les autres qui prennent de leur temps pour nous distraire.

♡ Merci les copains d'être géniaux!! Mille **Mercis** à Alexandra Bébé Toons 🐱 & ton nounours 🐱, Amandine Chouchou 🐱, Antho 🐱, Chloé 🐱, Eve 🐱, Ju 🐱, Khomas 🐱, Mary Tomate 🐱, Seb 🐱, Thomas 🐱 ... (c'est un ordre alphabétique ça? *Maintenant battez-vous* ... :p). Vous êtes la famille que je me suis choisie! J'aurais préféré écrire un petit mot pour chacun d'entre vous, mais je n'ai plus de place alors ... Ah! et aussi ... je vous aime fort d'amour!! Vous êtes tous au top (chacun à sa manière ... :p)!

♡ Ma thèse n'est rien sans mon appui familial solide. C'est là que cela commence à devenir difficile... Allez, je commence par ma Maman 🐱. Maman, merci pour ta confiance inébranlable. Malgré les difficultés, nous avons toujours réussi à tout surmonter, car nous sommes un ensemble, un cocon bienveillant, avec ma "gentille" sœur Danaé 🐱 et mon adorable frère Vianey 🐱. Merci à vous, à mon Papou 🐱, Maxime 🐱 (big'up à Maman Raymonde 🐱), et Aurélie 🐱, de faire de ma vie une folie permanente ♡ remplie d'Amour. Je vous aime plus que tout!

Je remercie également ma belle-famille, très soudée, que j'aime et à qui je ne le dis pas assez (voire jamais? oups). Merci mille fois 🐱 Catherine 🐱, Jean-Luc, Élodie, Adeline, Bertrand, Ramzi, Timéo 🐱, Solène 🐱, Jade 🐱 et Méya 🐱, et à tous les tontons, taties, cousins, cousines et leurs enfants, et à tous les autres aussi.

Victor 🐱, mon petit loup-kangourou, toi que je ne peux classer ni dans la famille (quoique), ni dans les amis, je te réserve ce petit paragraphe. Merci de ta bonne humeur, de ta mauvaise humeur, de ta curiosité, de ta bienveillance, de ton amour. Je t'aime mon petit loup ♡.



Merci mes petits chats, Chronos 🐱 et Gaïa 🐱, pour vos ronrons, de dormir quand je travaille, de miauler quand j'ai besoin de silence, de manger des pages de thèse lorsqu'elles sont nulles, ou de vomir dessus .

Et comme on garde toujours Le Meilleur pour la fin...

Jonathan, toi qui es à mes côtés depuis très longtemps maintenant, je te remercie du fond du cœur. Pas seulement de m'avoir épaulé patiemment dans cette thèse, mais de m'épauler chaque jour, de m'avoir accompagné dans mes projets fous, et de continuer à soutenir mes lubies.

Le plus bel amour est celui qui nous fait nous surpasser, celui qui nous rend vulnérable et qui éveille l'âme. Tu es bien plus que ça. Merci pour le passé, et merci d'avance pour l'avenir. Je t'aime ∞.

*Trust I seek and I find in you, every day for us something new, open mind for a different view, and nothing else matters. Nothing Else Matters - Metallica*



Merci à tous!

---



# Table des matières

---

<b>Résumé</b>	<b>i</b>
<i>Abstract</i>	<b>iii</b>
<b>Remerciements</b>	<b>iv</b>
<b>Liste des Figures</b>	<b>xv</b>
<b>Liste des Tables</b>	<b>xix</b>
<b>Listes des définitions</b>	<b>xxii</b>
<b>1 Introduction générale</b>	<b>1</b>
1.1 Le phytoplancton : définition, rôles et diversité . . . . .	1
1.1.1 Le phytoplancton à la base du réseau trophique marin . . . . .	1
1.1.2 Le phytoplancton comme acteur de la régulation climatique . . . . .	1
1.1.3 Diversité du phytoplancton et classifications fonctionnelles . . . . .	3
1.1.3.1 Classification par taille cellulaire . . . . .	4
1.1.3.2 Classification par fonctions biogéochimiques . . . . .	5
1.2 Suivi du phytoplancton . . . . .	8
1.2.1 Introduction à la couleur de l’océan . . . . .	9
1.2.2 Le signal marin . . . . .	11
1.2.3 La couleur de l’océan observée par satellite et son apport à l’étude du phytoplancton . . . . .	13
1.3 Détection des communautés phytoplanctoniques depuis l’espace grâce à la couleur de l’océan . . . . .	15
1.3.1 État des lieux des algorithmes de détection du phytoplancton . . . . .	16
1.3.2 Principe de la méthode PHYSAT . . . . .	21
1.4 Problématique et objectifs de la thèse . . . . .	24
<b>2 Approche détaillée et historique de la méthode PHYSAT</b>	<b>27</b>
2.1 Les anomalies de luminance . . . . .	29
2.1.1 Calcul de l’anomalie de luminance . . . . .	29
2.1.2 Critères de qualité permettant le calcul de l’anomalie de luminance .	31
2.1.3 Caractéristiques des anomalies de luminance . . . . .	32
2.1.4 Classification des anomalies de luminance . . . . .	33



2.2	Objectifs de la labellisation empirique : donner un nom aux anomalies de luminance observées . . . . .	40
2.2.1	À partir d'une base de données de concentration en pigments biomarqueurs publiée par le passé (Alvain <i>et al.</i> , 2005) . . . . .	41
2.2.2	À partir d'une expertise de terrain (Alvain <i>et al.</i> , 2008) . . . . .	45
2.2.3	À partir de données de cytométrie (Thyssen <i>et al.</i> , 2015) . . . . .	46
2.2.4	État des lieux des données potentiellement utilisables pour la calibration de PHYSAT . . . . .	47
2.3	Analyse théorique du lien empirique observé entre la présence d'un groupe phytoplanctonique dominant et d'une anomalie de luminance spécifique . .	48
2.4	Limites actuelles et questionnement sur le potentiel de développement de la méthode PHYSAT . . . . .	51
<b>3</b>	<b>Utilisation des caractéristiques saisonnières des anomalies de luminance comme outil de classification complémentaire</b>	<b>55</b>
3.1	Définition et calcul de la phénologie des anomalies de luminance . . . . .	59
3.1.1	Définition de la phénologie et de ses indicateurs . . . . .	60
3.1.2	L'information temporelle portée par les classes de $Ra(\lambda)$ . . . . .	61
3.1.3	Choix du seuil statistique pour la détermination des indices de la phénologie des classes de $Ra(\lambda)$ . . . . .	63
3.1.4	En pratique : développement d'un algorithme de calcul des métriques de la phénologie de manière flexible . . . . .	65
3.2	Utilisation des métriques de la phénologie des anomalies de luminances comme nouvelles caractéristiques des ces spectres pour une classification . .	66
3.3	Résultats . . . . .	69
3.3.1	Les indices de la phénologie : exemple pour les classes de $Ra(\lambda)$ du capteur SeaWiFS sur la période 1998-2010 . . . . .	69
3.3.2	Classification des classes de $Ra(\lambda)$ grâce à leur phénologie . . . . .	75
3.3.2.1	Classification des 100 classes de $Ra(\lambda)$ issues de la carte auto-organisatrice . . . . .	75
3.3.2.2	Exemples de séries temporelles de fréquence des classes groupées . . . . .	77
3.3.2.3	Exemples de spectres des classes groupées . . . . .	80
3.3.2.4	Validation spatiale des regroupements de classes de $Ra(\lambda)$ .	80
3.3.2.5	Diminution du nombre de classes de $Ra(\lambda)$ <i>via</i> la phénologie	85
3.4	Discussion . . . . .	88
3.5	Conclusion et perspectives . . . . .	90
<b>4</b>	<b>Détection des assemblages phytoplanctoniques à l'échelle globale par PHYSAT</b>	<b>93</b>
4.1	Matériels et méthodes . . . . .	96

4.1.1	Analyse des données <i>in-situ</i> : Étape 1 . . . . .	96
4.1.1.1	Les concentrations pigmentaires . . . . .	96
4.1.1.2	Informations générales sur les données <i>in-situ</i> de cette étude	97
4.1.1.3	Analyse des échantillons <i>in-situ</i> en fonction de leur com- position pigmentaire, au-delà de la dominance . . . . .	101
4.1.2	Données satellites et pré-traitement de l'information satellite : Étape	2103
4.1.3	Établissement du lien empirique entre les données <i>in-situ</i> et les ano- malies de luminance : Étape 3 . . . . .	104
4.1.3.1	Statistiques sur les matchups . . . . .	105
4.1.3.2	Établissement du lien empirique entre classes de $Ra(\lambda)$ et information <i>in-situ</i> : étape 3 . . . . .	110
4.2	Résultats . . . . .	112
4.2.1	Les assemblages phytoplanctoniques déterminés <i>in-situ</i> . . . . .	112
4.2.2	Lien empirique entre les données <i>in-situ</i> et les anomalies de luminance	120
4.2.3	Spectres des classes de $Ra(\lambda)$ labellisées . . . . .	128
4.2.4	Validations moyennes des associations classes de $Ra(\lambda)$ et assemblages	129
4.2.5	Distribution spatiale des classes de $Ra(\lambda)$ labellisées . . . . .	134
4.3	Discussion . . . . .	154
4.3.1	Classification des concentrations pigmentaires <i>in-situ</i> . . . . .	154
4.3.2	Procédure(s) de calibration et de validation . . . . .	155
4.3.3	Distribution spatiale des assemblages <i>via</i> la distribution des classes de $Ra(\lambda)$ associées . . . . .	159
4.3.4	Spectres des classes de $Ra(\lambda)$ associées aux assemblages . . . . .	161
4.3.5	Contribution des assemblages à la concentration en chlorophylle- <i>a</i> estimée . . . . .	163
4.4	Conclusion . . . . .	166
<b>5</b>	<b>Potentiel d'information porté par les anomalies de luminance : étude en Mer du Nord</b>	<b>167</b>
5.1	Matériels et méthodes . . . . .	168
5.1.1	Caractérisation des assemblages CPR <i>in-situ</i> : étape 1 . . . . .	168
5.1.1.1	Structuration des données CPR . . . . .	170
5.1.1.2	La méthode Indicator Value de Dufrêne et Legendre (1997)	174
5.1.2	Calcul des $Ra(\lambda)$ dans la région de la Mer du Nord : étape 2 . . . . .	180
5.1.3	Essai de labellisation et validation des labels : étape 3 . . . . .	182
5.1.3.1	Procédure classique de calibration et validation . . . . .	182
5.1.3.2	Répétitions de la procédure de calibration et de validation	183
5.2	Résultats . . . . .	184
5.2.1	Espèces indicatrices & assemblages de diatomées . . . . .	184
5.2.2	Procédures successives de calibration . . . . .	187

5.2.3	Spectres d'anomalies de luminance associés aux assemblages de diatomées . . . . .	190
5.2.4	Validation . . . . .	191
5.2.5	Distribution spatio-temporelle des anomalies de luminance associées aux assemblages de diatomées . . . . .	199
5.3	Discussion . . . . .	210
5.3.1	Choix de la méthode <i>Indicator Value</i> pour la caractérisation des assemblages de diatomées <i>in-situ</i> . . . . .	210
5.3.2	Distribution des classes de $Ra(\lambda)$ labellisées . . . . .	212
5.3.3	Variabilité des caractéristiques spectrales des $Ra(\lambda)$ labellisés en amplitude, forme et phénologie . . . . .	215
5.3.3.1	Variations des caractéristiques de $Ra(\lambda)$ intra- et inter-assemblages . . . . .	215
5.3.3.2	Cas de "labellisations multiples" . . . . .	218
5.3.4	Validation des associations . . . . .	219
5.3.5	La base de données de $Ra(\lambda)$ en Mer du Nord, limitations de l'approche et perspectives . . . . .	219
5.4	Conclusion . . . . .	222
<b>6</b>	<b>Conclusion générale et perspectives</b>	<b>223</b>
6.1	Limites de la méthode PHYSAT : bilan actuel et propositions d'études <i>via</i> l'expérimentation (milieux contrôlés) et/ou l'approche théorique . . . . .	224
6.1.1	Limites liées à l'aspect empirique . . . . .	224
6.1.2	Expériences en milieu contrôlé et simulations pour aller plus loin dans la compréhension théorique du signal d'anomalie de luminance . . . . .	225
6.2	Intégration de nouveaux paramètres dans l'algorithme PHYSAT : vers l'évaluation des probabilités de présence des groupes et assemblages phytoplanc-toniques . . . . .	227
6.3	Évaluation du potentiel de couplage multicapteurs dans le cadre de l'utilisation de PHYSAT . . . . .	233
6.4	Conclusion . . . . .	236
	<b>Valorisation des travaux</b>	<b>239</b>
	<b>Références</b>	<b>241</b>
	<b>Annexes</b>	<b>255</b>
	Annexe 1 : Liste des radiomètres principaux à portée globale . . . . .	257
	Annexe 2 : Cartes de Kohonen pour les capteurs MODIS, VIIRS, et MERIS . . . . .	259

Annexe 3 : Seuils des $Ra(\lambda)$ définissant les signatures spectrales pour les groupes dominants dans la version Alvain <i>et al.</i> (2005) . . . . .	263
Annexe 4 : Séries temporelles de fréquences de détection des 100 classes de $Ra(\lambda)$ (capteur SeaWiFS) issues de la carte auto-organisatrice . . . . .	265
Annexe 5 : Algorithme " <i>FindPheno</i> " développé en R, R Core Team (2015) . . . . .	277
Annexe 6 : Cartes de fréquences des classes de $Ra(\lambda)$ regroupées grâce à leurs ressemblances spectrales (super-classes <i>som</i> ) . . . . .	279
Annexe 7 : Carte d'occurrence des classes de $Ra(\lambda)$ <i>som</i> (gauche) et des classes de $Ra(\lambda)$ <i>phénologie</i> (droite) les plus fréquentes par mois (Climatologie SeaWiFS 1998-2010) . . . . .	281
Annexe 8 : Les campagnes en mer ayant collectés des échantillons pour analyses pigmentaires . . . . .	285
Annexe 9 : Nom des provinces de Longhurst . . . . .	287
Annexe 10 : Cartes de distribution des fréquences de détection mensuelles des assemblages v2017 en climatologie (PHYSAT version 2017, SeaWiFS, de 1998 à 2010) . . . . .	289
Annexe 11 : Cartes de distribution des fréquences de détection mensuelles des assemblages vMixte en climatologie (PHYSAT version Mixte (2008-2017), SeaWiFS, de 1998 à 2010) . . . . .	297
Annexe 12 : <i>Estimation of the potential detection of diatom assemblages based on ocean color radiance anomalies in the North Sea - Frontiers in Marine Science, 2017</i> . . . . .	309
Annexe 13 : Le <i>Continuous Plankton Recorder</i> . . . . .	321
Annexe 14 : Liste des espèces de diatomées des données CPR en Mer du Nord . . . . .	323
Annexe 15 : Méthode <i>Indicator Value</i> : résultats pour toutes les diatomées du CPR . . . . .	325



# Liste des Figures

(Légendes courtes)

---

<b>Chapitre 1 : Introduction générale</b>	<b>1</b>
1.1 Schéma de la pompe biologique d'après Passow et Carlson (2012). . . . .	3
1.2 Images de cellules phytoplanctoniques en microscopie illustrant la diversité en formes et en taille des cellules. . . . .	4
1.3 Estimations de concentrations moyennes en Chlorophylle- <i>a</i> pour l'océan mondial des capteurs SeaWiFS, MODIS, VIIRS. . . . .	14
1.4 Enveloppes statistiques de décision pour la labellisation des groupes de PHYSAT, version 2005 (Alvain <i>et al.</i> , 2005). . . . .	22
1.5 Exemples de produits cartographiques de la méthode PHYSAT issus de la calibration du capteur SeaWiFS. . . . .	23
<b>Chapitre 2 : Approche détaillée et historique de la méthode PHYSAT</b>	<b>27</b>
2.1 Schéma du processus empirique utilisé pour la calibration de PHYSAT. . .	28
2.2 Table de référence de PHYSAT (SeaWiFS, L3) . . . . .	30
2.3 Exemples de caractéristiques des spectres de $R_a(\lambda)$ . . . . .	33
2.4 Enveloppes statistiques de décision pour la labellisation des groupes de PHYSAT, version 2005 (Alvain <i>et al.</i> , 2005). . . . .	34
2.5 Schéma d'une structure de carte auto-organisatrice, adapté de Kohonen (2013). . . . .	36
2.6 Spectres de références (ligne rouge) de la carte auto-organisatrice du capteur de couleur de l'océan SeaWiFS (échelle globale). . . . .	39
2.7 Exemples de cartes d'occurrence et de fréquence des classes de $R_a(\lambda)$ du capteur SeaWiFS, à l'échelle globale et en climatologie. . . . .	40
2.8 Cartes de synthèses mensuelles de la distribution des cinq groupes dominants de la méthode PHYSAT (versions 2008 & 2014) pour le capteur SeaWiFS. . . . .	44
2.9 Carte des groupes de phytoplancton dominants de Janvier 2002, avec l'ajout du groupe dominant <i>Phaeocystis spp.</i> (Alvain <i>et al.</i> , 2008). . . . .	46

2.10	Formes et amplitudes des anomalies de luminance théoriques en tenant compte de l'absorption phytoplanctonique $a_{phy}$ , la rétrodiffusion et l'absorption des matières dissoutes colorées (Alvain <i>et al.</i> , 2012). . . . .	50
2.11	Anomalies de luminance théoriques et observées pour les groupes de PHY-SAT (Alvain <i>et al.</i> , 2012). . . . .	51

**Chapitre 3 : Utilisation des caractéristiques saisonnières des anomalies de luminance comme outil de classification complémentaire** **55**

3.1	Répartition des groupes phytoplanctoniques dominants définis par Alvain <i>et al.</i> (2005, 2008) sur la carte auto-organisatrice de Ben Mustapha <i>et al.</i> (2014) pour le capteur de couleur de l'océan SeaWiFS. . . . .	56
3.2	Exemple de classes de $Ra(\lambda)$ dont les caractéristiques spectrales sont similaires mais la fréquence d'apparition dans le temps et la distribution spatiale sont différentes. . . . .	59
3.3	Indicateurs de la phénologie adapté de Platt et Sathyendranath (2008). . .	60
3.4	Exemples de fréquences d'occurrence relatives de classes de $Ra(\lambda)$ issues du traitement PHYSAT des données du capteur SeaWiFS, à l'échelle globale. .	62
3.5	Initiations de l'augmentation de fréquence des 100 classes de $Ra(\lambda)$ du capteur de couleur SeaWiFS entre 1998 et 2010. . . . .	71
3.6	Répartitions des périodes d'initiation, de fin et de maximum pour le pic de fréquence principal des classes de $Ra(\lambda)$ sw. . . . .	72
3.7	Fin des augmentations de fréquence des 100 classes de $Ra(\lambda)$ du capteur de couleur SeaWiFS entre 1998 et 2010. . . . .	73
3.8	Maxima de l'augmentation de fréquence des 100 classes de $Ra(\lambda)$ du capteur de couleur SeaWiFS entre 1998 et 2010. . . . .	74
3.9	Dendrogramme de la classification des indices de la phénologie des classes de $Ra(\lambda)$ SeaWiFS. . . . .	76
3.10	Position dans la carte auto-organisatrice de quelques classes de $Ra(\lambda)$ regroupées par la phénologie. . . . .	77
3.11	Séries temporelles des fréquences d'occurrences des classes d'anomalies de luminance SeaWiFS classées en fonction de leur phénologie. . . . .	79
3.12	Spectres de références des classes de $Ra(\lambda)$ SeaWiFS qui groupées par leur phénologie. . . . .	80
3.13	Fréquences relatives des pixels associés aux classes de $Ra(\lambda)$ n°21 et n°61, considérées comme similaires au point de vue de leur phénologie. . . . .	82
3.14	Fréquences relatives des pixels associés aux classes de $Ra(\lambda)$ n°23 et n°87, considérées comme similaires au point de vue de leur phénologie. . . . .	83
3.15	Fréquences relatives des pixels associés aux classes de $Ra(\lambda)$ n°18 et n°28, considérées comme similaires au point de vue de leur phénologie. . . . .	84

3.16 Spectres référents des classes de $Ra(\lambda)$ SeaWiFS (calculés par la carte auto-organisatrice) et groupées grâce à leur phénologie. . . . .	86
3.17 Carte de classes de $Ra(\lambda)$ la plus fréquente pour les pixels disponibles du mois de Décembre en climatologie de SeaWiFS (1998-2010) pour les classes de $Ra(\lambda)$ <i>som</i> et de $Ra(\lambda)$ <i>phénologie</i> . . . . .	87
3.18 Comparaison des classifications "phénologie" et "spectrales" (super-classes).	90

#### **Chapitre 4 : Détection des assemblages phytoplanctoniques à l'échelle globale par PHYSAT 93**

4.1 Diagramme de la répartition des cas de dominances des données <i>in-situ</i> brutes selon les seuils établis sur les concentrations relatives en pigments par Alvain <i>et al.</i> (2005). . . . .	98
4.2 Provinces écologiques de Longhurst (Longhurst, 2010). . . . .	99
4.3 Répartition des données <i>in-situ</i> utilisées dans cette étude pour définir des profils pigmentaires. . . . .	100
4.4 Classification des échantillons <i>in-situ</i> d'après leur composition pigmentaire grâce à une méthode classique (distance de Mahalanobis, <b>UPGMA</b> ). . . . .	102
4.5 Topologie d'une mesure coïncidente selon les recommandations de Werdell et Bailey (2005) . . . . .	105
4.6 Concentrations relatives <i>in-situ</i> des pigments biomarqueurs dans les 25 classes de la carte auto-organisatrice, regroupée ensuite en 10 classes. . . . .	113
4.7 Concentrations relatives <i>in-situ</i> des cinq pigments biomarqueurs au sein des 10 assemblages. . . . .	116
4.8 Cartes de localisation des points <i>in-situ</i> appartenant aux groupes pigmentaires issus de la CAH post-carte auto-organisatrice . . . . .	117
4.9 Spectres de référence des classes de $Ra(\lambda)$ labellisés en fonction des 10 assemblages déterminés <i>in-situ</i> . . . . .	128
4.10 Distribution climatologique des fréquences de détection des $Ra(\lambda)$ associés aux Nanoecaryotes dominants <i>via</i> une calibration empirique (v2017). . . . .	135
4.11 Distribution climatologique des fréquences de détection des $Ra(\lambda)$ associés aux diatomées dominantes <i>via</i> une calibration empirique (v2017). . . . .	136
4.12 Distribution climatologique des fréquences de détection des $Ra(\lambda)$ associés aux <i>Prochlorococcus spp.</i> non dominants <i>via</i> une calibration empirique (v2017). . . . .	137
4.13 Distribution climatologique des fréquences de détection des $Ra(\lambda)$ associés à l'assemblage de diatomées et nanoecaryotes <i>via</i> une calibration empirique (v2017). . . . .	138



4.14	Distribution climatologique des fréquences de détection des Ra( $\lambda$ ) associés à l'assemblage de <i>Synechococcus spp.-like</i> et <i>Prochlorococcus spp. via</i> une calibration empirique (v2017). . . . .	139
4.15	Distribution climatologique des fréquences de détection des Ra( $\lambda$ ) associés aux dinoflagellés dominants <i>via</i> une calibration empirique (v2017). . . . .	140
4.16	Cartes de localisation des assemblages déterminés <i>in-situ</i> dans cette étude (v2017). . . . .	143
4.17	Cartes de localisation des assemblages déterminés <i>in-situ</i> dans cette étude (vMixte). . . . .	150
4.18	Contributions mensuelles moyennes à la concentration en Chlorophylle- <i>a</i> estimée des assemblages phytoplanctoniques détectés par PHYSAT. . . . .	164

## **Chapitre 5 : Potentiel d'information porté par les anomalies de luminance : étude en Mer du Nord** **167**

5.1	Cartes de la localisation des 5977 points d'échantillonnage du <i>Continuous Plankton Recorder</i> entre 1998 et 2010, disponibles pour cette étude. . . . .	169
5.2	Dendrogramme des données CPR basé sur une dissimilarité de Bray-Curtis et un algorithme UPGMA calculé pour les données d'abondance de diatomées	184
5.3	Taux de labellisation des classes de Ra( $\lambda$ ) durant les procédures successives de calibration qui incluent 1200 sous-sets de calibration. . . . .	189
5.4	Spectres d'anomalies de luminance labellisés (spectres de référence des classes de Ra( $\lambda$ )) pour l'assemblage caractérisé par a) <i>Thalassionema nitzschoides</i> , b) <i>Thalassiosira sp.</i> et c) <i>Rhizosolenia imbricata</i> . . . . .	191
5.5	Cycles saisonniers moyens de deux classes de Ra( $\lambda$ ) ayant des caractéristiques spectrales similaires, mais étant différents dans leur phénologie. . . . .	191
5.6	Moyenne des scores de validation pour les 1200 procédures de labellisation des classes de Ra( $\lambda$ ). . . . .	197
5.8	Cycle saisonnier moyen des six assemblages CPR issus de la méthode IndVal.	210
5.9	Carte des zones écologiques de la Manche Mer du Nord définies à partir des régimes de stratification (Van Leeuwen <i>et al.</i> , 2015). . . . .	213
5.10	Exemple d'arrangements de cellules de l'espèce de diatomée <i>Thalassionema nitzschoides</i> . . . . .	218

## **Chapitre 6 : Discussion générale et perspectives** **223**

6.1	Hypervolume de la niche écologique de trois paramètres environnementaux et types de gamme de variation des préférences. . . . .	229
-----	---	-----

6.2	Représentations graphiques de tables de références PHYSAT des capteurs de couleur de l'océan a) SeaWiFS, b) MODIS, c) MERIS et d) VIIRS. . . .	234
6.3	Inter-comparaison des distributions spatiales de Janvier 2004 pour SeaWiFS et MODIS. . . . .	235



# Liste des Tables

(Légendes courtes)

---

<b>Chapitre 1 : Introduction générale</b>	<b>1</b>
1.1 Caractéristiques des trois principales méthodes permettant d'identifier simultanément plusieurs types phytoplanctoniques : OC-PFT, PhytoDOAS et PHYSAT. . . . .	19
<b>Chapitre 2 : Approche détaillée et historique de la méthode PHYSAT</b>	<b>27</b>
2.1 Seuils sur les concentrations relatives en pigments biomarqueurs pour définir la dominance <i>in-situ</i> (Alvain <i>et al.</i> , 2005). . . . .	42
2.2 Scores de validation des associations trouvées pour PHYSAT (Alvain <i>et al.</i> , 2005) . . . . .	43
<b>Chapitre 3 : Utilisation des caractéristiques saisonnières des anomalies de luminance comme outil de classification complémentaire</b>	<b>55</b>
3.1 Exemple de nombre d'expansion de la fréquence des classes de $Ra(\lambda)$ pour le capteur SeaWiFS sur la période 1997-2010. . . . .	66
3.2 Détails sur les classes de $Ra(\lambda)$ SeaWiFS regroupées par les nœuds 1, 15 et 95 pendant la classification sur les indices phénologiques associées à la fréquence d'apparition de ces classes de $Ra(\lambda)$ . . . . .	75
<b>Chapitre 4 : Détection des assemblages phytoplanctoniques à l'échelle globale par PHYSAT</b>	<b>93</b>
4.1 Seuils sur les concentrations relatives en pigments biomarqueurs pour définir la dominance <i>in-situ</i> (Alvain <i>et al.</i> , 2005). . . . .	97

4.2	Information croisée sur les matchups montrant le nombre de mesures coïncidentes disponibles pour le jeu de données <i>in-situ</i> croisé avec les données du capteur de couleur SeaWiFS, et ce, pour plusieurs niveaux d'information décrits ci-dessus. . . . .	106
4.3	Concentrations relatives moyennes des cinq pigments biomarqueurs utilisés dans cette étude au sein des 25 classes pigmentaires définies par la carte auto-organisatrice. . . . .	114
4.4	Informations de labellisation concernant les nouvelles labellisation de classes de $Ra(\lambda)$ par rapport aux labellisations de Alvain <i>et al.</i> (2008) et Ben Mustapha <i>et al.</i> (2014) . . . . .	122
4.5	Informations de labellisation concernant les confirmations des labellisations existantes de classes de $Ra(\lambda)$ par rapport aux labellisations de Alvain <i>et al.</i> (2008) et Ben Mustapha <i>et al.</i> (2014) . . . . .	123
4.6	Informations de labellisation concernant les changements des labellisations existantes de classes de $Ra(\lambda)$ par rapport aux labellisations de Alvain <i>et al.</i> (2008) et Ben Mustapha <i>et al.</i> (2014) . . . . .	125
4.7	Nombre de $Ra(\lambda)$ associés à chaque groupe phytoplanctonique déterminés empiriquement <i>in-situ</i> . . . . .	130
4.8	Scores de validation des labels par groupe phytoplanctonique défini à partir des données <i>in-situ</i> . . . . .	133
4.9	Pourcentage de pixels associés mensuellement à un groupe phytoplanctonique de la version Mixte, par rapport aux nombres total de pixels pour lesquels un groupe fréquent a pu être identifiés. . . . .	148

**Chapitre 5 : Potentiel d'information porté par les anomalies de lumiance : étude en Mer du Nord** **167**

5.1	Table des corrélations cophénétiques entre les dissimilarités calculées sur les données CPR et celles issues d'algorithmes de regroupement. . . . .	174
5.2	Résultats de l'IndVal pour plusieurs niveaux de regroupements des échantillons CPR. . . . .	185
5.3	Répartition des valeurs de spécificité, de fidélité et d'IndVal du taxon <i>Thalassiosira spp.</i> au niveau de regroupement L5, distinguant six assemblages (colonnes). Les valeurs sont données en pourcentage et les valeurs maximales sont signalées par un astérisque. . . . .	186
5.4	Assemblages de diatomées issus de la méthode Indicator Value de Dufréne et Legendre (1997) . . . . .	187
5.5	Nombre de mesures coïncidentes (calibration et validation) par assemblages <i>in-situ</i> . . . . .	188

5.6	Exemple de validation intermédiaire concernant l'association entre classes de $Ra(\lambda)$ et assemblages CPR. . . . .	194
5.7	Exemple de validation <i>a posteriori</i> concernant l'association entre classes de $Ra(\lambda)$ et assemblages CPR après sélection des associations robustes. . . . .	195
5.8	Scores de validation totaux et moyens des procédures de calibration des anomalies de radiance en Mer du Nord. . . . .	198

**Chapitre 6 : Discussion générale et perspectives** **223**

6.1	Moyennes des paramètres environnementaux estimés de chaque groupe PHY-SAT, version 2008 et Mixte. . . . .	232
-----	---	-----

*Une liste de définitions est disponible afin de faciliter la lecture de la thèse dans son aspect multidisciplinaire. Dans le corps du manuscrit, les mots définis dans cette liste sont **mis en évidence en italique et emphase** lors de leur première apparition dans chaque chapitre, et sont listés ci-après dans l'ordre alphabétique.*

# Liste des définitions

---

**AOT - *Aerosol optical thickness*** *Aerosol optical thickness* ou épaisseur optique associée aux aérosols. C'est une grandeur sans unité qui définit la transmissibilité de la lumière au travers d'une couche d'atmosphère comprenant les aérosols pour une longueur d'onde donnée. De manière schématique, plus l'épaisseur optique des aérosols est grande, moins la lumière traverse l'atmosphère. 32

**Assemblage** Un assemblage phytoplanctonique est un regroupement d'espèces, qui co-existent sur une même fenêtre temporelle dans une même zone géographique (celles-ci pouvant être en mouvement) ; car elles possèdent une physiologie, morphologie et / ou d'autres facteurs qui induisent une même réponse face aux variations des conditions environnementales. 24, 47, 94, 168, 223

**Autotrophie** L'autotrophie est la production, par un être vivant (souvent végétal), de matières organiques à partir d'éléments inorganiques et minéraux. *A contrario*, les organismes produisant leurs matières organiques à partir d'autres matières organiques sont appelés hétérotrophes. Notez qu'il existe également des organismes mixotrophes (autotrophes ou hétérotrophes non-strict). 1

**Biomasse phytoplanctonique** En écologie, la biomasse désigne la masse totale d'organismes vivants dans un biotope déterminé à un moment donné. La biomasse phytoplanctonique est donc le poids des cellules phytoplanctoniques pour un volume d'eau. Elle est souvent exprimée en masse de carbone par volume. 1, 160

**Corrélations cophénétiqes** Les corrélations cophénétiqes servent à comparer des algorithmes de classifications en réalisant une corrélation entre les distances (ou dissimilarités) initiales (calculées sur les données brutes), et les distances (ou dissimilarités) des données groupées, dont les valeurs dépendent de l'algorithme de classification utilisé (e.g., moyenne, maximum). Ainsi, la corrélation cophénétiqes permet d'estimer à quel point la classification respecte l'agencement initial des données brutes. 67, 173, 174

**Cytométrie en flux - application au phytoplancton** La cytométrie en flux est une technique permettant d'analyser des échantillons cellules par cellules. Initialement développée pour le milieu médical, des développements du cytomètre (instrument de mesure), ont permis l'applicabilité à l'analyse des échantillons d'eau. Les cellules, entraînées par un flux, passent une à une devant un faisceau laser et la lumière réémise est mesurée par différents capteurs, permettant de classer les cellules suivant plusieurs critères (e.g., fluorescences, tailles). 41, 46, 155, 180, 224



**Erreur topologique** L'erreur topologique quantifie le respect de la topologie des observations. Pour une nouvelle donnée présentée à la carte auto-organisatrice, les deux références de classes les plus proches sont évaluées : si ces deux références sont dans des cellules adjacentes de la carte, l'erreur topologie sera faible. Si au contraire, les deux plus proches références de classes sont éloignées dans les cellules de la carte auto-organisatrice, alors l'erreur topologique augmente. 37, 102

**Erreurs de quantification vectorielle** Dans la réalisation d'un algorithme de classification de type carte auto-organisatrice, l'erreur de quantification vectorielle est mesurée par le calcul de la distance euclidienne entre la référence de la classe (ou vecteur référeny) et une observation quelconque qui a été assignée à cette classe. 37, 102

**Eurytopique - espèce eurytopique** Une espèce est dite "eurytopique" lorsqu'elle a une grande capacité de dispersion, qui est liée à de grandes gammes d'optimum de croissance et de développement pour de nombreux facteurs environnementaux. Elles sont aussi appelées espèces "cosmopolites" ou "ubiquistes". 179, 229

**Fidélité d'une espèce / d'un taxon au sens Dufrêne et Legendre (1997)** Une espèce, ou un taxon, est considéré(e) comme fidèle à un ensemble d'échantillons si elle/il apparaît dans, et uniquement dans ce groupe d'échantillons. 175

**HNLC - zone océanique HNLC** HNLC signifie High Nutrient, Low Chlorophyll. Ce qualificatif est utilisé pour définir des zones de l'océan où la plupart des conditions sont réunies pour voir fleurir du phytoplancton, mais pour lesquelles un élément trace manque : le fer. Cet élément est en effet utilisé en tant que co-facteur (ou "molécule d'assistance") des réactions enzymatiques permettant la photosynthèse (réduction des nitrates, incorporation du phosphate). Une zone HNLC connue est le gyre Pacifique Nord (zone sub-antarctique) e.g., Martin et Fitzwater (1988). 159

**HPLC** Accronyme de l'anglais *High Pressure Liquid Chromatography*, ou chromatographie en phase liquide à haute pression, est une technique de séparation analytique de molécules présentes dans un échantillon à analyser. 9, 41, 95, 96, 224

**Indicatrice - espèce indicatrice au sens Dufrêne et Legendre (1997)** Une espèce est indicatrice si elle est *fidèle* et *spécifique*. 170, 175

**Inégalité triangulaire** Dans le cas de l'utilisation de distances entre deux points, en mathématiques, l'inégalité triangulaire reflète le fait que la distance la plus courte entre deux points est une ligne droite. Aussi, par extension, pour trois points dans un espace mathématique, la plus grande distance est toujours inférieure ou égale à la somme des deux plus petites distance. Une métrique de distance est considérée comme vraie si elle respecte cette inégalité triangulaire. Si cette règle n'est pas respectée, le terme de "dissimilarité" est alors utilisé. 171

**IOPs** IOPs pour *inherent optical properties* ou propriétés optiques inhérentes. Ce sont les propriétés (i.e., paramètres assez stables pour être considérés comme caractéristiques) qui sont issues de l'interaction de la lumière avec les composants du milieu, mais qui ne dépendent que de la composition du milieu, et ne dépendent pas de la source de lumière. Lorsqu'un faisceau lumineux interagit avec un milieu (ici, l'eau), une partie de la lumière est transmise (pas d'interaction avec des composants), une partie est absorbée par les composants, et une partie est diffusée par les composants. L'absorption et la diffusion peuvent prendre plusieurs formes en fonction de la nature du composant (par exemple : absorption ou diffusion due à l'eau, et aux particules). Les IOPs sont additives, c'est-à-dire que la valeur totale de l'IOP considérée est la somme des valeurs de cette IOP pour tous les composants de l'eau. La décomposition se fait souvent ainsi : eau pure, particulaire (phytoplanctonique et non phytoplanctonique), dissoutes (colorée ou non). 11, 48, 90, 93, 221, 224

**Labellisation** La labellisation, dans le cadre de PHYSAT est le fait de "donner un nom" à un type d'anomalie de radiance à partir d'observations *in-situ*. Cette labellisation a été réalisée le plus souvent grâce à la calibration empirique, même si il est aussi possible de réaliser cette labellisation grâce à une expertise de terrain. 40, 91, 120, 168, 233

**Lagrangienne - Approche *in-situ* Lagrangienne** L'approche d'étude lagrangienne *in-situ* consiste à suivre l'évolution d'une masses d'eau, en utilisant le plus souvent une moyen de "locomotion inerte" pour l'instrumentation de mesures (bouée ou bateau dérivants), qui suit les trajectoires du fluide; par opposition à l'approche eulérienne, qui consiste à étudier un point fixe, et à reconstruire l'origine du point dans le temps d'étude grâce aux champs de vitesse. 221, 224

**Niche écologique** La niche écologique est l'ensemble des paramètres biotique et abiotique qui conditionne la répartition d'une espèce (??). Elle est représentée comme un hypervolume où chaque dimension est un paramètre déterminant pour l'espèce. 211, 228

**Néritique - eaux néritiques** Les eaux néritiques définissent le volume d'eau situé entre la surface de l'eau et la plateforme continentale, et à ce qu'elle contient (compartiments biotiques et abiotiques). 214

**Paramètre de distorsion** La distorsion est un indice calculé à partir des erreurs de quantification et des erreurs de topologie en prenant compte également la notion de voisinage. 37, 102

**Phénologie** La phénologie est l'étude de la temporalité des cycles saisonniers des événements périodiques. 58, 95, 190, 223

**Pigments biomarqueurs** Les pigments biomarqueurs sont les pigments spécifiques à un type phytoplanctonique. Il existe des pigments biomarqueurs exclusifs, c'est-à-dire spécifique à un seul type phytoplanctonique; et des pigments biomarqueurs

ambigus, qui seront présents chez plusieurs types phytoplanctoniques. 21, 41, 94, 96, 223

**Pompe biologique de carbone** Le phytoplancton est le principal acteur de la "pompe biologique" océanique. La photosynthèse, qui permet au phytoplancton de produire sa *biomasse*, consomme du CO<sub>2</sub>, dont le carbone est ainsi fixé dans la biomasse des organismes. C'est ce qu'on appelle la "pompe biologique". Ce carbone peut être transféré vers les niveaux trophiques supérieurs (consommation des autotrophes), où s'accumuler dans l'océan profond par sédimentation du phytoplancton. 2, 160

**Population** Une population désigne l'ensemble des individus d'une même espèce qui occupe simultanément le même milieu. 17, 94, 215

**Production primaire phytoplanctonique** La production primaire est la quantité de matière organique formée à partir de matière minérale et d'un apport d'énergie (la photosynthèse). C'est un flux exprimé en masse de carbone assimilé par unité de biomasse. 1

**Saturation du radiomètre** Une saturation temporaire du capteur peut survenir lorsque le milieu observé réfléchit fortement les rayons lumineux (e.g., *sun-glint*, glace ou milieu blanc, coccolithes). 32

**SOM** *Self Organizing Map*, ou carte auto-organisatrice, est un algorithme de classification de type réseau de neurone (Kohonen, 2013). Les cartes auto-organisatrices permettent le classement d'objets sur lesquels aucune information relative à leur classification n'est disponible *a priori*. 34, 56, 95, 181, 223

**Spécificité d'une espèce / d'un taxon au sens Dufrêne et Legendre (1997)**  
Une espèce, ou un taxon, est considéré(e) comme spécifique à un ensemble d'échantillons si elle/il apparaît dans tous les échantillons de ce groupe. 175

**Sténotopique - espèce sténotopique** Une espèce est dite "sténotopique" lorsqu'elle a une faible, ou moyenne capacité de dispersion, qui est liée à une ou plusieurs petites gammes d'optimum de croissance et de développement pour un ou plusieurs facteurs environnementaux. 211, 229

**Taxon (sg.) - Taxa (pl.)** En taxinomie, un taxon est une entité regroupant tous les organismes vivants possédant en commun certaines caractéristiques bien définies. Le terme taxon est utilisé dans la classification phylogénétique pour regrouper des êtres vivants en fonction de divers critères génétiques.. 3, 45, 161, 170

**UPGMA** L'UPGMA - *Unweighted Pair Group Method with Arithmetic mean* - est un algorithme de regroupement qui réduit progressivement la taille du jeu de données en groupant successivement deux éléments dont la distance est la plus faible. Le terme "Arithmetic mean" signifie que, une fois que deux éléments ont été groupés, l'algorithme considère la moyenne des distances de ces deux points au reste des données à grouper. 69, 102, 170





# Introduction générale

---

L'objectif de cette thèse est d'identifier des groupes fonctionnels phytoplanctoniques grâce à une méthode dite de "couleur de l'océan". Dans ce premier Chapitre, nous allons donc introduire le sujet d'étude, le phytoplancton, en insistant sur sa diversité fonctionnelle et ses rôles dans le système océanique. Nous introduirons ensuite la thématique de la couleur de l'océan, pour en venir à une problématique centrée sur une méthode spécifique qui permet l'identification du phytoplancton depuis l'espace (PHYSAT, Alvain *et al.* (2005)).

## 1.1 Le phytoplancton : définition, rôles et diversité

Le phytoplancton représente l'ensemble des végétaux ("phyto") en suspension dans l'eau ("plancton"). Dans le milieu marin, le phytoplancton a un rôle prépondérant dans les écosystèmes et la biogéochimie du milieu. Il est présent sous la forme d'algues microscopiques photosynthétiques, avec des tailles variant de 0,2 à 200  $\mu\text{m}$  (Sieburth *et al.*, 1978).

### 1.1.1 Le phytoplancton à la base du réseau trophique marin

Le phytoplancton est à la base du réseau trophique marin en tant qu'organisme photosynthétique *autotrophe*. Ainsi, ce compartiment produit la *biomasse* disponible pour les niveaux trophiques supérieurs (e.g., brouteurs, planctonophages) à partir de l'énergie lumineuse et d'éléments nutritifs (e.g., Edwards et Richardson (2004), Richardson et Schoeman (2004)). L'énergie potentiellement assimilable par les niveaux trophiques supérieurs dépend directement de l'énergie disponible dans le compartiment phytoplanctonique, et de son accessibilité (Hairston et Hairston (1993), Frederiksen *et al.* (2006)).

### 1.1.2 Le phytoplancton comme acteur de la régulation climatique

Puisqu'il utilise les éléments nutritifs inorganiques de son milieu, le phytoplancton joue un rôle clef dans les cycles des éléments biogéochimiques, et dans le système climatique. Bien que représentant 1% de la biomasse végétale terrestre totale, il contribue à hauteur d'environ 50% à la *production primaire* mondiale (Field *et al.* (1998)), soulignant son importance. Son rôle de producteur primaire lui confère une influence importante dans de

nombreux cycles biogéochimiques marins (Falkowski (1994), Beaugrand (2015)), comme ceux du carbone, de l'azote, de la silice ou encore du phosphore; qu'il intègre *via* les processus de photosynthèse, pour former sa propre biomasse.

Le rôle de "*pompe biologique*" imputé au phytoplancton lui permet d'être un élément clef dans l'équilibre climatique de la Terre, en participant notamment à la régulation de la quantité de carbone inorganique atmosphérique, qu'il soit d'origine anthropique ou naturelle. En effet, le phytoplancton utilise le dioxyde de carbone ( $\text{CO}_2$ ) atmosphérique, dissous dans l'eau sous forme de carbonates, pour fabriquer sa biomasse. Cela a deux conséquences majeures sur le cycle du carbone :

- i Lorsque le phytoplancton sédimente vers les fonds marins, il "piège" le carbone d'origine atmosphérique à long terme puisque ce carbone a été intégré à sa biomasse : c'est ce qu'on appelle la "pompe biologique" (Figure 1.1).
- ii En utilisant le carbone dissous dans l'eau pour ses processus de photosynthèse, le phytoplancton crée un léger déséquilibre entre la pression de  $\text{CO}_2$  atmosphérique et la pression de  $\text{CO}_2$  océanique de surface. Puisque cet équilibre doit être maintenu (lois physiques), du  $\text{CO}_2$  passera du compartiment atmosphérique vers la sub-surface de l'eau.

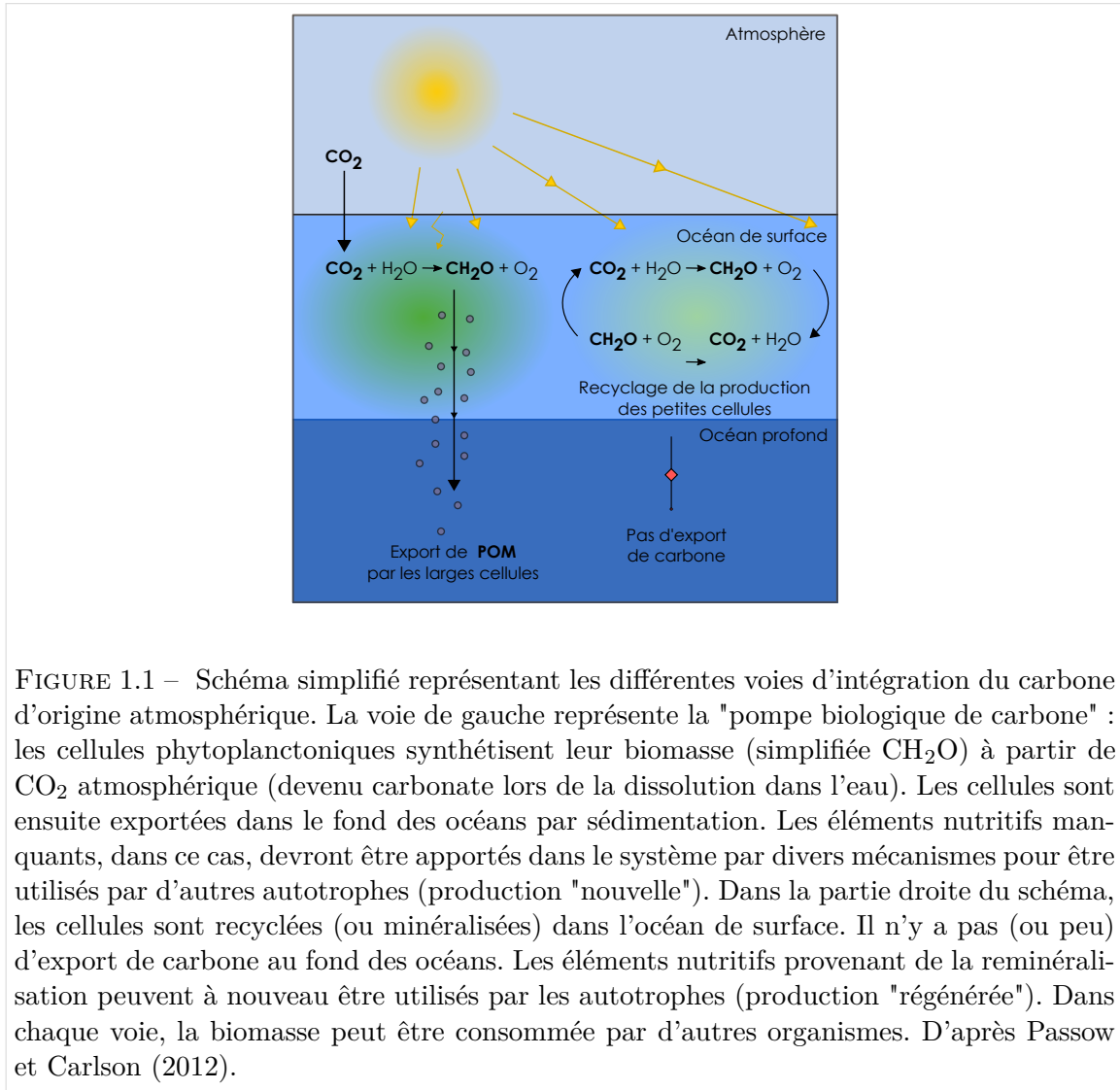


FIGURE 1.1 – Schéma simplifié représentant les différentes voies d'intégration du carbone d'origine atmosphérique. La voie de gauche représente la "pompe biologique de carbone" : les cellules phytoplanctoniques synthétisent leur biomasse (simplifiée  $CH_2O$ ) à partir de  $CO_2$  atmosphérique (devenu carbonate lors de la dissolution dans l'eau). Les cellules sont ensuite exportées dans le fond des océans par sédimentation. Les éléments nutritifs manquants, dans ce cas, devront être apportés dans le système par divers mécanismes pour être utilisés par d'autres autotrophes (production "nouvelle"). Dans la partie droite du schéma, les cellules sont recyclées (ou minéralisées) dans l'océan de surface. Il n'y a pas (ou peu) d'export de carbone au fond des océans. Les éléments nutritifs provenant de la reminéralisation peuvent à nouveau être utilisés par les autotrophes (production "régénérée"). Dans chaque voie, la biomasse peut être consommée par d'autres organismes. D'après Passow et Carlson (2012).

### 1.1.3 Diversité du phytoplancton et classifications fonctionnelles

Le phytoplancton présente une grande diversité de classes taxonomiques et de *taxa* (Figure 1.2), dont les distributions dans le temps et dans l'espace, et les interactions biogéochimiques avec le milieu, sont très variables. Chaque taxon possède en effet sa propre efficacité dans l'incorporation des éléments de son milieu (Le Quéré *et al.*, 2005).

La diversité phytoplanctonique est liée à la variabilité des distributions des nutriments (quantité, qualité et accessibilité), de la température, de la lumière (quantité, qualité et photopériode), mais aussi d'autres facteurs physiques (e.g., la stratification, Sverdrup (1953)). La variabilité environnementale et climatique influence donc, *via* le phytoplancton (mais pas uniquement), la structure et le fonctionnement des écosystèmes pélagiques dans leur intégralité. Par exemple, les diatomées (microphytoplancton) sont capables de grands changements dans leur cycle de vie en fonction des conditions climatiques, à l'échelle locale (Schloss *et al.*, 2012), et à plus grande échelle (Alvain *et al.*, 2013). Elles sont connues pour



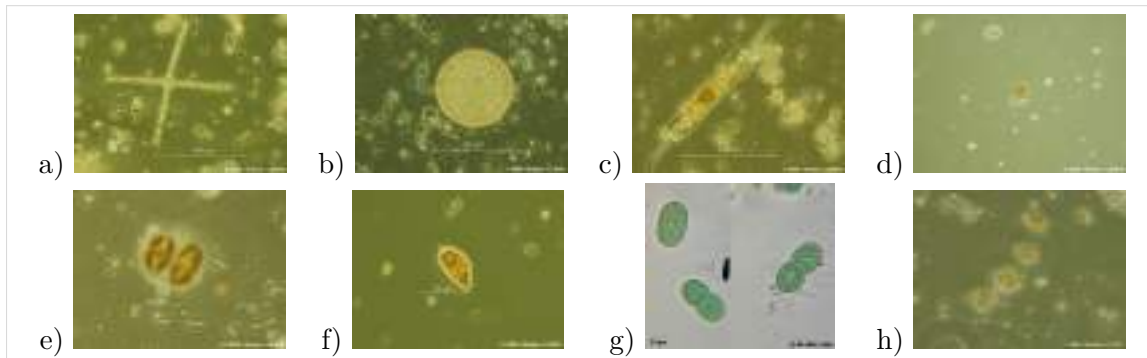


FIGURE 1.2 – Images de cellules phytoplanctoniques en microscopie illustrant la diversité en formes et en taille des cellules : a) *Asterionella* spp., b) *Coscinodiscus* spp., c) *Ditylum* spp., d) *Karenia* spp., e) *Podosira* spp., f) *Prorocentrum* spp., g) *Synechococcus* spp., h) *Thalassiosira* spp.. Source : Ifremer (LERPC), archive disponible en ligne

être plutôt présentes dans des milieux riches en nutriments (Sarthou *et al.*, 2005; Armbrust, 2009).

En fonction de l’affinité d’un type de phytoplancton pour son milieu, il est possible de définir des classes fonctionnelles (PFTs, pour *Phytoplankton Functional Types*) regroupant les taxa ou classes taxonomiques phytoplanctoniques en fonction de leur rôle dans le système biogéochimique. Principalement, il existe deux grands types de classifications fonctionnelles : en fonction de la taille des organismes et en fonction de leurs processus biologiques.

### 1.1.3.1 Classification par taille cellulaire

La classification du phytoplancton par taille est une première approche pour définir les classes fonctionnelles. En effet, la taille cellulaire est considérée comme un trait majeur qui structure les communautés (Marañón, 2015). Les trois classes communément utilisées sont (Sieburth *et al.*, 1978) :

- le picophytoplancton ( $< 2\mu m$ )
- le nanophytoplancton ( $2 - 20\mu m$ )
- le microphytoplancton ( $> 20\mu m$ )

Les processus biogéochimiques, tels que l’incorporation des nutriments et l’utilisation de l’énergie lumineuse, dépendent du rapport entre la surface de l’organisme et son volume (Margalef, 1978) ; rapport qui définit la capacité d’absorption des nutriments.

Par exemple, le picophytoplancton possède un ratio surface/volume élevé, le rendant capable d’incorporer ses nutriments sous des conditions limitantes, i.e., avec de faibles concentrations en nutriments. Ainsi, cette classe de taille est très compétitive pour se dé-

velopper dans les eaux oligotrophes, qu'elle domine (e.g., Kostadinov *et al.* (2010)). Les petites cellules induisent un export plus faible de carbone vers l'océan profond, car la reminéralisation dans la colonne d'eau est prépondérante (Waterbury *et al.*, 1986).

Le microphytoplancton a un ratio surface/volume faible, ce qui le rend peu compétitif pour les faibles concentrations en sels nutritifs, mais très compétitif pour les fortes concentrations. Ainsi, le microphytoplancton se développe et domine les eaux eutrophes, i.e., riches en nutriments. Les grandes cellules induisent un export de carbone vers l'océan profond plus important que les petites cellules (Buesseler, 1998).

Le nanophytoplancton représente un cas intermédiaire. Cette classe de taille est la plupart du temps opportuniste et se développe souvent en cas de non-développement des deux autres classes de taille de cellules phytoplanctoniques (Margalef, 1978).

Bien que la classification du phytoplancton par sa taille aide fortement dans la compréhension du milieu marin (Nair *et al.*, 2008), il existe des cas où cette classification n'est pas assez rigoureuse, notamment pour des études biogéochimiques. En effet, deux types phytoplanctoniques différents dans leurs fonctions biogéochimiques peuvent se retrouver dans la même classe de taille (Nair *et al.*, 2008). L'exemple le plus couramment cité est le cas des producteurs de diméthyle-sulphide (DMS) et des calcifieurs, qui font tous les deux partie du nanophytoplancton. Ces deux PFTs ont un impact très différent sur le système climatique : le DMS produit par le phytoplancton peut s'oxyder et former des précurseurs de condensation pour les nuages ; et induire indirectement un refroidissement de l'atmosphère (Malin *et al.*, 1992). Les calcifieurs, quant à eux, peuvent diminuer l'alcalinité des eaux de surface lorsqu'ils synthétisent leur squelette à partir de carbonates, et peuvent donc entraîner une augmentation de la pression partielle en CO<sub>2</sub>, et un relargage de CO<sub>2</sub> dans l'atmosphère (c'est la "contre-pompe des carbonates")<sup>1</sup>. Cet exemple montre qu'il faut aller plus loin que les classes de taille phytoplanctonique pour comprendre les systèmes biogéochimiques, en définissant les PFTs grâce aux fonctions biogéochimiques ou aux traits physiologiques (Hood *et al.*, 2006).

### 1.1.3.2 Classification par fonctions biogéochimiques

De manière classique, quatre classes fonctionnelles de phytoplancton sont définies : les diazotrophes, les silicifieurs, les calcifieurs et les producteurs de DMS (e.g., Nair *et al.* (2008), IOCCG *report* (2014)). Les processus principaux de ces groupes fonctionnels sont détaillés ensuite. Ces processus généraux sont à nuancer, car ils peuvent varier en fonc-

---

1. En réalité, le rôle des calcifieurs, et en particulier des coccolithophoridés, est plus complexe. En effet, ils participent également à la "pompe biologique" de carbone en réalisant la photosynthèse (Hoppe *et al.*, 2011) mais modifient également l'équilibre Calcaire/carbonates/dioxyde de carbone en fabriquant leur thèque ("contre pompe des carbonates") (Beaufort *et al.*, 2011). La balance entre la "pompe biologique de carbone" et la "contre pompe des carbonates" penche cependant plutôt en faveur de la pompe biologique de carbone actuellement (e.g., Beaufort *et al.* (2011)). Dans un contexte de changement climatique (et d'acidification des océans), ils seraient également capables d'une grande résilience (Lohbeck *et al.*, 2012).

tion des conditions environnementales (e.g., température, disponibilité et accessibilité des nutriments, pression en  $\text{CO}_2$ ).

#### — Diazotrophes

La forme de l'azote  $\text{N}_2$  dissous est la forme la plus abondante de l'azote dans l'océan. Cependant, la plupart des organismes photosynthétiques sont incapables de l'utiliser sous cette forme et préfèrent les formes nitrates ( $\text{NO}_3^-$ ), nitrites ( $\text{NO}_2^-$ ) et ammonium ( $\text{NH}_4^+$ ). Les diazotrophes ont la particularité de réduire les molécules de  $\text{N}_2$  dissous dans l'eau de surface pour les transformer en ammoniac ( $\text{NH}_3$ ), utilisable dans le processus de synthèse de matière organique. Ce sont souvent des petites cellules phytoplanctoniques (e.g., cyanobactéries), ou d'autres organismes parasites qui utilisent ce processus de diazotrophie. Les diazotrophes ont donc une influence directe sur le cycle de l'azote, en injectant de l'azote nouveau dans le milieu lorsqu'ils sont reminéralisés (Capone *et al.*, 2005). La diazotrophie entraîne un gain net d'azote ; et une quantité d'azote accrue dans l'océan permet de mobiliser, sous forme organique, une plus grande quantité de carbone. La diazotrophie, quoique mettant actuellement en jeu des quantités d'azote modestes, serait donc de nature à augmenter l'efficacité de la pompe biologique de carbone, et a donc une influence non négligeable sur l'équilibre climatique (Falkowski, 1997; Karl *et al.*, 2012).

#### — Silicifieurs

Les silicifieurs sont les organismes utilisant la silice ( $\text{SiO}_2$ ) comme nutriment ou élément structural. Il existe quatre groupes taxonomiques de silicifieurs : les Chrysophytes, les Silicoflagellés, les Xanthophytes et les Bacillariophytes (Brownlee et Taylor, 2002). Les diatomées, du groupe taxonomique des Bacillariophytes sont les silicifieurs dominants dans le système océanique. Elles ont une contribution d'environ 40% à la production primaire marine totale (Sarhou *et al.*, 2005). Elles utilisent la silice de forme biogénique pour former un squelette externe. Ces grandes cellules ont tendance à sédimenter (à cause d'une grande densité), participant à la pompe biologique de carbone, de nitrate et de phosphates.

#### — Calcifieurs

Le phytoplancton calcifieur produit des coquilles (ou des coccolithes) en carbonates de calcium. Ils sont souvent nommés "coccolithophores". Comme expliqué plus haut, le processus de calcification consomme les carbonates dissous pour former du carbonate de calcium, accompagné d'eau et de  $\text{CO}_2$ . Ainsi, la pression partielle en  $\text{CO}_2$  de l'eau de surface augmente par rapport à celle de l'atmosphère (induisant une diminution de l'alcalinité de l'eau) et il se produit un relargage de  $\text{CO}_2$  de l'eau vers l'atmosphère (Brown et Yoder, 1994).

#### — Producteurs de DMS

Il existe un nombre important de types de phytoplancton qui produisent du diméthyle sulfoniopropionate (DMSP), qui est le précurseur du diméthyle sulfure (DMS, composé

volatile). Ces composés (DMSP) ont un rôle d'anti-oxydant face à de nombreux stress environnementaux (e.g., contre les radiations solaires ultra-violettes, la limitation en CO<sub>2</sub>, la limitation en fer et en nitrates et la présence de peroxyde d'hydrogène (Sunda *et al.*, 2002, 2007)). Lorsque la cellule meurt ou est prise pour proie, il peut alors y avoir un relargage de ces composés DMSP dans l'eau, puis dans l'atmosphère (où ils deviennent du DMS). Les groupes taxonomiques producteurs de DMS sont principalement les Dinoflagellés et les Haptophytes, mais certaines Chrysophycées, Pelagophycées et Prasinophycées sont aussi connues pour en produire (Sunda *et al.*, 2002). Le DMSP produit par le phytoplancton peut s'oxyder en aérosols sulfatés, étant des précurseurs de condensation pour les nuages, et induire indirectement un refroidissement de l'atmosphère par augmentation de l'albédo atmosphérique (Malin *et al.*, 1992).

La définition de ces quatre grands types phytoplanctoniques montre que des différences notables existent pour leurs impacts sur le système océanique, notamment en ce qui concerne la biogéochimie. De plus, ils ont aussi des influences sur le réseau alimentaire, chaque groupe étant plus ou moins accessible aux niveaux trophiques supérieurs, et induisent des réseaux trophiques différents (e.g., Sailley *et al.* (2013)).

 **Résumé : Les enjeux de l'étude du phytoplancton**

Dans cette introduction, nous avons abordé l'importance du phytoplancton dans les écosystèmes marins (producteur primaire) et en tant que régulateur de mécanismes (ou cycles) ayant un impact sur le système climatique (direct ou indirect). Les interactions du phytoplancton sur ces systèmes sont modulés par les spécificités de chaque type de phytoplancton (ou PFTs). Le suivi spatial et temporel du phytoplancton, que ce soit par l'intermédiaire de son abondance, de sa biomasse, ou de sa diversité, fait l'objet d'un consensus scientifique (e.g., Boyce *et al.* (2010)) et politique (e.g., directives européennes DCSMM).

## 1.2 Suivi du phytoplancton

D'après les éléments précédemment introduits, le phytoplancton est d'une importance majeure dans les écosystèmes marins, et dans la régulation du climat. Il existe de nombreuses approches pour étudier sa biomasse, son abondance ou sa diversité ; et le choix de la méthode employée dépend du questionnement scientifique et de son échelle d'étude, i.e., de la micro-échelle (e.g., processus physiologiques) à la macro-échelle (e.g., répartition globale et climat). Ces approches peuvent être groupées dans trois grandes sous-catégories : *in-situ*, en laboratoire (i.e., expérimentale) et les études synoptiques. Hormis quelques techniques telle que la microscopie, ces approches sont souvent basées sur les propriétés optiques des cellules et notamment sur les propriétés optiques de la Chlorophylle-*a*.

En effet, la concentration en Chlorophylle-*a* est probablement le paramètre le plus fréquemment estimé en océanographie pour approximer l'abondance du phytoplancton. La Chlorophylle-*a* est un pigment photosynthétique qui est présent chez tous les taxa phytoplanctoniques<sup>2</sup>. Bien que la quantité de Chlorophylle-*a* ne soit pas le reflet exact de la biomasse des cellules, ce paramètre est considéré, à ce jour, comme la meilleure approche pour l'évaluer.

La Chlorophylle-*a* est une molécule chimique colorée essentielle à la photosynthèse. Dans les cellules de phytoplancton, la Chlorophylle-*a*, couplée à d'autres pigments (formant l'antenne collectrice), capte les rayons solaires pour en transmettre l'énergie dans le processus de photosynthèse. Plus précisément, la Chlorophylle-*a* absorbe majoritairement le signal dans les longueurs d'onde bleues (~440nm) et rouges (~675nm) du spectre de lumière, et

2. excepté chez les *Prochlorophytes* du genre *Prochlorococcus spp.* où il est remplacé par la Divinyl-Chlorophylle-*a*, la forme oxydée de la Chlorophylle-*a*.

diffuse les longueurs d'onde vertes. C'est pour cela que les eaux chargées en phytoplancton paraissent de couleur verte à l'œil nu.

*In-situ*, il est possible de quantifier la Chlorophylle-*a* soit par fluorescence *in-vivo*, ou en amenant des échantillons au laboratoire pour une analyse de fluorescence *in-vitro*, mais aussi par **HPLC** (*High Performance Liquid Chromatography*), qui permet une analyse du contenu pigmentaire (Chlorophylle-*a* et autres pigments accessoires)<sup>3</sup>.

Il est également possible d'estimer cette concentration en Chlorophylle à plus grande échelle spatio-temporelle grâce aux capteurs de couleur de l'océan embarqués sur des satellites.

Si les observations *in-situ* sont absolument indispensables à l'étude du phytoplancton, et notamment de sa concentration et de sa diversité, les observations satellites apportent depuis près de 40 ans une vision complémentaire. En effet, les satellites d'observation de la couleur de l'océan sont un outil unique pour les océanographes, car ils permettent une vue synoptique des propriétés optiques de la surface de l'eau à haute résolution spatiale et temporelle (de l'ordre du kilomètre voire moins et quasiment journalier). Cette résolution permet d'effectuer des études à l'échelle globale et sur plusieurs décennies, ce qui est nécessaire pour comprendre les systèmes biogéochimiques dans leur ensemble.

### 1.2.1 Introduction à la couleur de l'océan

La mesure dans le domaine appelé communément "couleur de l'océan" est basée sur un constat qui peut être schématisé comme suit : l'eau pure apparaît de couleur bleue (forte absorption dans les longueurs d'onde rouge et diffusion dans les longueurs d'onde bleues des molécules d'eau de mer pure) ; et des cellules phytoplanctoniques qui se développent dans l'eau tendent à la verdir. Ce phénomène est visible depuis l'espace et observable depuis les années 1970<sup>4</sup>, avec l'ajout de radiomètres (aussi appelés capteurs de couleur de l'océan) sur les satellites en orbite autour de la Terre. Une liste des principaux capteurs de couleur de l'océan à l'échelle globale est disponible en Annexe 1 (page 257).

La couleur de l'océan est définie comme les variations spectrales de la réflectance diffuse des couches océaniques supérieures (réflectance marine  $R(\lambda)$ ). La réflectance est le rapport entre l'éclairement ascendant émergeant juste sous la surface de l'eau (et ayant réagi avec ses composants) et l'éclairement descendant qui y pénètre (et provenant du soleil dans ce cas), Équation 1.1. Par définition, la réflectance diffuse  $R(\lambda)$  est propagée dans toutes les directions. En revanche, le satellite, équipé du capteur, n'observe la surface de l'eau

3. Il existe d'autres techniques d'estimation de la concentration en Chlorophylle-*a*, qui ne seront pas détaillées ici.

4. En 1978, le premier radiomètre dédié à l'observation de l'océan, CZCS (*Coastal Zone Colour Scanner*), a été disposé sur le satellite Nimbus-7 (appartenant à la NASA).

que dans une direction (appelée "angle d'observation") et capte donc une partie de la réflectance diffuse. Le signal reçu dans cette direction spécifique s'appelle alors luminance marine ( $L_w(\lambda)$ ). Le rapport entre l'éclairement descendant provenant du soleil et la luminance marine dans la direction du satellite peut alors être défini : c'est la réflectance de télédétection ou  $Rrs$ , Équation 1.2. La dernière étape est de normaliser les réflectances de télédétection afin de simuler une mesure satellite au nadir avec le soleil au zénith : c'est la **luminance marine normalisée**  $nLw(\lambda)$ , Équation 1.3.

**Réflectance  $R(\lambda)$  :**

$$R(\lambda) = \frac{E_u(\lambda)}{E_d(\lambda)} \quad (1.1)$$

**Réflectance de télédétection**

**$Rrs(\lambda)$  :**

$$Rrs(\lambda) = \frac{L_w(\lambda)}{E_d(\lambda)} \quad (1.2)$$

**Luminance normalisée  $nLw(\lambda)$  :**

$$nLw(\lambda) = Rrs(\lambda) * F_0(\lambda) \quad (1.3)$$

Où  $R(\lambda)$  est la réflectance totale,  $E_u(\lambda)$  est l'éclairement ascendant (intégrale des  $L_w(\lambda)$ ) et  $E_d(\lambda)$  l'éclairement descendant (la source);  $Rrs(\lambda)$  est la réflectance de télédétection,  $L_w(\lambda)$  est la luminance mesurée dans la direction du capteur;  $nLw(\lambda)$  est la luminance marine normalisée, qui est fonction de  $F_0(\lambda)$ , l'éclairement solaire incident au zénith, dont les valeurs sont connues.

À partir des luminances marines mesurées, il est possible d'estimer un grand nombre de paramètres utiles à l'étude et à la compréhension du milieu marin. Les algorithmes partant de ces mesures sont nombreux, et parfois basés sur des relations empiriques ou des relations déterminées par d'autres moyens, e.g., par modélisations, et permettent d'avoir accès à des estimations (liste et références associées non exhaustives) :

- La lumière disponible pour la photosynthèse (e.g., Frouin *et al.* (2012)),
- Les propriétés optiques inhérentes : absorption et rétrodiffusion particulière (e.g., Franz et Werdell (2010); Werdell *et al.* (2013))
- La concentration en Chlorophylle-*a* (e.g., O'Reilly *et al.* (2000); Hu *et al.* (2012))
- La concentration en carbone organique particulière (e.g., Stramski *et al.* (2008))
- La concentration en carbone inorganique particulière (e.g., Gordon *et al.* (2001); Balch *et al.* (2005))

### 1.2.2 Le signal marin

#### Propriétés optiques

La réflectance marine  $R(\lambda)$ , définie en Équation 1.1, peut également être exprimée en termes de propriétés optiques des eaux et de ses constituants (Équation 1.4).

$$R(\lambda) = F * \frac{b_b(\lambda)}{a(\lambda) + b_b(\lambda)} \quad (1.4)$$

Où  $R(\lambda)$  est la réflectance marine,  $b_b(\lambda)$  est le coefficient de rétrodiffusion et  $a(\lambda)$  est le coefficient d'absorption et  $F$  est un facteur de proportionnalité.

Les coefficients de rétrodiffusion et d'absorption sont des propriétés optiques inhérentes (**IOPs**), ce qui signifie que leur valeur dépend uniquement des constituants du milieu. De plus, les IOPs de l'eau de mer résultent de la somme des contributions respectives des composants de l'eau optiquement actifs, i.e., l'eau de mer pure, les particules (le phytoplancton et autres particules non algales), et les matières dissoutes optiquement actives (Sathyendranath et Morel, 1983).

En ce qui concerne le phytoplancton, voici les paramètres ayant une influence sur les IOPs, avec des degrés d'influence différents sur l'absorption et/ou la rétrodiffusion :

- La composition taxonomique du phytoplancton (e.g., Bricaud et Morel (1986); Bricaud *et al.* (1988); Hoepffner et Sathyendranath (1991); Ciotti *et al.* (2002));
- La morphologie (forme et taille) et la structure interne des cellules phytoplanctoniques (Bricaud et Morel, 1986; Dubelaar *et al.*, 1987; Stramski et Kiefer, 1991; Clavano *et al.*, 2007; Whitmire *et al.*, 2010);
- Les pigments (Chlorophylle-*a* et pigments accessoires) dans leur composition, leur concentration et leur modification face aux conditions de l'environnement (Bricaud et Morel, 1986; Johnsen *et al.*, 1994; Whitmire *et al.*, 2010);
- Les matières dissoutes associées optiquement actives, telle que la concentration en matière dissoute colorées (CDOM) qui sont produites pendant le déclin d'une communauté phytoplanctonique, et de manière générale le carbone organique dissous (Bricaud *et al.*, 1981; Siegel *et al.*, 2005);
- Enfin, le stade de croissance (Moutier *et al.*, 2017), induisant une variation de la taille des cellules, de la morphologie, et de l'organisation des cellules entre elles (Boss *et al.*, 2009) (cellules libres, en colonie<sup>5</sup>, cellules agrégées).

---

5. Une colonie est ensemble de cellule qui sont liées entre-elles soit par un mucus exogène, ou par des ponts entre cellules. Il existe plusieurs formes de colonies : cellules emprisonnées dans un mucus, cellules formant une chaîne linéaire, en zigzag ou étoile.



### Les types d'eaux

Nous avons vu dans le paragraphe précédent que les causes de variations de la couleur de l'océan dépendent des constituants de l'eau qui sont optiquement actifs. Dans ce cadre, une classification des types d'eaux est possible en considérant la contribution des différents éléments au spectre mesuré. Ces eaux peuvent être classées en de nombreux types (par exemple, 14 types pour Jerlov (1976)), mais la classification la plus usitée est celle définie par Morel et Prieur (1977) qui comprend deux classes :

**Les eaux du "Cas I"** Les eaux dites du "Cas I" sont définies comme les eaux dont les propriétés optiques sont influencées par l'eau de mer pure et le phytoplancton (le vivant, le détritique et les matières dissoutes associées à son activité physiologique et à sa dégradation). Elles représentent en général environ 95% de la surface de l'océan global.

**Les eaux du "Cas II"** Les eaux du "Cas II" peuvent être définies de manière grossière comme les eaux côtières ou sous influence terrestre. Les propriétés optiques de ces eaux sont alors influencées, comme pour les eaux du cas I, par l'eau de mer et le phytoplancton, ce à quoi on ajoute les autres composés optiquement actifs comme les sédiments et d'autres particules terrigènes (remis en suspension du fond marin, ou apportés par lessivage ou l'érosion des terres et des glaciers), et les matières dissoutes issues de l'érosion ou du lessivage des sols, ainsi que d'autres apports anthropogéniques possibles et ayant un impact optique.

Il existe également d'autres classifications et sous-classifications des types d'eau (e.g., Prieur et Sathyendranath (1981); Vantrepotte *et al.* (2012)), mais celle-ci (Cas I *vs.* II) est la plus répandue et communément admise. Les algorithmes pour estimer les différents paramètres précités sont souvent différents selon les types d'eaux. La communauté scientifique a récemment mis en place des algorithmes plus performants dédiés aux eaux du Cas II, qui montrent un fort intérêt économique (par exemple les algorithmes développés dans le cadre du projet GlobCoast, e.g., Loisel *et al.* (2016); Dessailly *et al.* (2016)).

Le type d'eau est un aspect à prendre en compte dans l'étude de la couleur de l'océan. Il y a également d'autres aspects à prendre en compte avant de pouvoir utiliser les données de luminance de l'océan, comme la correction atmosphérique.

### Corrections atmosphériques

Il est important de garder à l'esprit que le signal de l'océan qui atteint le capteur satellite est soumis à de nombreuses interactions sur son trajet vers le satellite. En effet, le satellite observe la planète depuis son orbite autour de la Terre, et donc à travers l'atmosphère. Celle-ci interagit avec le rayonnement ascendant de manière importante. Il est estimé que moins de 10% du signal reçu par le satellite correspond effectivement au signal marin

(Mobley, 1994). Ainsi, lorsque le signal total est reçu par le satellite, il faut estimer la part qui est due à l'atmosphère afin de corriger les  $nLw(\lambda)$ . Ceci constitue un domaine de recherche qui consiste à développer des méthodes dites de "corrections atmosphériques".

La part du signal due aux molécules de l'atmosphère est assez bien connue et peut être estimée en fonction de la pression atmosphérique. En revanche, la part des aérosols atmosphériques est plus difficile à estimer, notamment à cause de la variété des natures, et des variations de concentrations des aérosols. Afin de corriger le signal des aérosols, l'hypothèse de "l'océan noir" est souvent utilisée. En effet, pour des longueurs d'onde du proche infrarouge (à partir de 750nm, Gordon et Wang (1994)), le signal marin est faible voire nul (au moins valable pour les eaux des océans ouverts). Ainsi, le signal mesuré dans le proche infrarouge est identifiable aux aérosols, que l'on identifie en utilisant deux bandes spectrales dans le proche infrarouge, que l'on extrapole pour avoir leur contribution sur le signal optique dans les autres longueurs d'onde.

Les données peuvent aussi être masquées, c'est-à-dire qu'une partie de l'image est cachée. Il existe plusieurs type de masquage dont les standards sont les terres, les nuages, la glace, et le sun-glint (réverbération) dans le cadre de l'utilisation des données marines.

Une fois ces corrections et masques appliqués, les données de luminance marines sont utilisables à des fins algorithmiques ou d'applications.

### **1.2.3 La couleur de l'océan observée par satellite et son apport à l'étude du phytoplancton**

Afin d'estimer la concentration en Chlorophylle-*a* à partir de la couleur de l'océan, des modèles bio-optiques ont été développés depuis plus de 30 ans. Ils ont consisté, dans un premier temps, mais pas seulement, à établir des relations empiriques entre les observations *in-situ* (i.e., des estimations de concentrations en Chlorophylle-*a in-situ*) et les données radiométriques mesurées par le capteur. La Chlorophylle-*a* présente son pic d'absorption maximum à 440 nm, alors que son minimum d'absorption se situe à 550 nm. Ainsi, des approches basées sur le rapport entre ces deux bandes spectrales donnent une information sur le contenu en Chlorophylle-*a* (Morel, 1988). Les algorithmes d'estimation de la quantité de Chlorophylle-*a* sont dérivés du même principe : ils utilisent le rapport des longueurs d'onde bleues / vertes pour en estimer la concentration.

Actuellement, l'algorithme le plus couramment utilisé définit une relation polynomiale de degré 4 entre les rapports bleu/vert et la concentration en Chlorophylle-*a* (OCx, O'Reilly *et al.* (2000)), ce qui donne l'Équation :

$$\log_{10}([Chla - a]) = a_0 + \sum_{i=1}^4 a_i * (\log_{10} * (\frac{Rrs(\lambda_{blue})}{Rrs(\lambda_{green})}))^i \quad (1.5)$$

Où  $a_0, a_1, a_2, a_3$  et  $a_4$  sont des coefficients spécifiques à un jeu de longueurs d'onde. Les longueurs d'onde utilisées ( $Rrs(\lambda_{blue})$  et  $Rrs(\lambda_{green})$ ) varient également avec le capteur (voir Annexe 1, page 257) ; l'algorithme OCx pour l'estimation de la concentration en Chlorophylle-*a* est donc capteur-dépendant.

Il existe également d'autres approches, qu'elles soient empiriques ou semi-analytiques, mais l'algorithme OCx (O'Reilly *et al.*, 2000) est, à ce jour, le plus utilisé (cité 2600 fois selon l'éditeur). La Figure 1.3 est une illustration de l'application de cet algorithme aux données  $Rrs(\lambda)$  de trois capteurs (SeaWiFS, MODIS-Aqua et VIIRS).

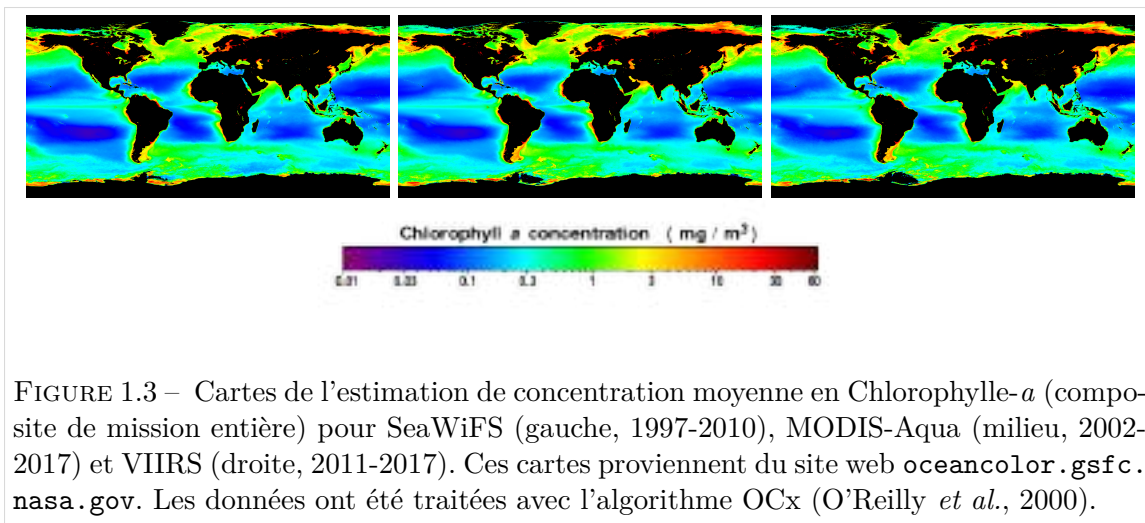


FIGURE 1.3 – Cartes de l'estimation de concentration moyenne en Chlorophylle-*a* (composite de mission entière) pour SeaWiFS (gauche, 1997-2010), MODIS-Aqua (milieu, 2002-2017) et VIIRS (droite, 2011-2017). Ces cartes proviennent du site web [oceancolor.gsfc.nasa.gov](http://oceancolor.gsfc.nasa.gov). Les données ont été traitées avec l'algorithme OCx (O'Reilly *et al.*, 2000).

L'estimation de la distribution de la concentration en Chlorophylle-*a* révèle des zones contrastées de l'océan, allant des zones oligotrophes des gyres océaniques, aux zones les plus riches en phytoplancton comme les zones côtières. La Figure 1.3, qui représente 6 ans (VIIRS) ou plus (SeaWiFS et MODIS) de données, illustre déjà ce qu'il est possible de réaliser grâce à la télédétection de la couleur de l'océan. Disposer de 10 ans de données est une grande opportunité pour étudier un processus à macro-échelle (e.g., la structuration des communautés phytoplanctoniques face aux changements climatiques globaux). Cependant, les images satellites ont une résolution spatio-temporelle qui permet également d'étudier des processus à plus fine échelle temporelle (journée, semaine) mais également spatiale (région particulière).

Comme nous l'avons vu au début de ce Chapitre, l'estimation de la concentration en Chlorophylle-*a* est un premier pas vers la compréhension du système océanique dans son ensemble (Behrenfeld et Falkowski, 1998). Cependant, la Chlorophylle-*a* est un pigment présent chez tous les organismes phytoplanctoniques (ou sous sa forme DivinylChlorophylle-*a*), et ne permet pas de distinguer les caractéristiques liées à des tailles ou des taxa particuliers. C'est pour apporter ces informations complémentaires que d'autres algorithmes ont été développés plus récemment. Ces données sont notamment attendues par les modélisateurs du climat et des cycles biogéochimiques de l'océan (e.g., Le Quéré *et al.* (2005); Bopp *et al.* (2005); Gehlen *et al.* (2006); Le Clainche *et al.* (2010); Vogt *et al.* (2013)), qui couplent des données d'estimations de distributions des communautés avec des modèles climatiques et environnementaux.

Des algorithmes permettent d'aller plus loin dans la compréhension de la répartition du phytoplancton et de la biogéochimie en détectant les classes de taille phytoplanctoniques, mais aussi la répartition des types fonctionnels, grâce à des algorithmes souvent empiriques.

### 1.3 Détection des communautés phytoplanctoniques depuis l'espace grâce à la couleur de l'océan

*Nota bene* Le terme de "types fonctionnels phytoplanctoniques" (PFTs), définit en Section 1.1.3.2, est utilisé pour caractériser des ensembles très particuliers de fonctions biogéochimiques (mais néanmoins diversifiés en termes de taxonomie). Bien que la plupart des études tendent à se rapprocher de cette définition, l'ensemble de la biodiversité liée à ces types phytoplanctoniques n'est jamais totalement évaluée. C'est pourquoi, dans la suite du manuscrit, nous préférons utiliser le terme de "groupe<sup>6</sup> phytoplanctonique".

Le phytoplancton est composé de divers groupes taxonomiques, qui se manifestent par différentes tailles, morphologies et compositions pigmentaires. Ces caractéristiques influencent l'absorption et la rétrodiffusion de la lumière (Bricaud *et al.*, 1988, 1995; Babin *et al.*, 2003), ce qui peut potentiellement engendrer des variations des signaux mesurés depuis l'espace. Des algorithmes existent pour détecter les groupes de taille phytoplanctonique, ainsi que les groupes fonctionnels. Dans cette Section introductive, nous allons aborder quelques-uns des algorithmes disponibles à ce jour.

---

6. Le terme groupe est utilisé dans le sens "sous-ensemble", et n'est pas en lien avec le terme utilisé en classification taxonomique.

### 1.3.1 État des lieux des algorithmes de détection du phytoplancton

#### Détection des classes de tailles

La taille des cellules phytoplanctoniques est considérée comme un facteur important pour les études écologiques et biogéochimiques (Sieburth *et al.*, 1978), c'est pourquoi de nombreux algorithmes ont été développés pour estimer ce paramètre. La communauté scientifique s'accorde sur les gammes de taille qui sont définies en 3 classes : le picophytoplancton (0,2 - 2 $\mu$ m), le nanophytoplancton (2 - 20 $\mu$ m) et le microphytoplancton (20 - 200 $\mu$ m), ce qui est en accord avec Sieburth *et al.* (1978).

Dans les méthodes de détection des classes de taille par satellite, une distinction peut être faite entre celles qui sont entièrement dédiées aux cellules phytoplanctoniques ("PSC" pour *Phytoplankton Size Classes*), de celles qui sont dédiées à l'ensemble des particules (algales ou non, "PSD" pour *Particule Size Distribution*). De nombreuses méthodes existent, en voilà une liste succincte, illustrant quelques caractéristiques :

- Kostadinov *et al.* (2009) et Kostadinov *et al.* (2010) ont développé un algorithme dédié à la détection des classes de taille particulaire. Cet algorithme est développé à partir de données satellites uniquement, et il est validé grâce à des données *in-situ* (HPLC). À partir de la luminance marine normalisée, Kostadinov *et al.* (2009) estiment le coefficient de rétrodiffusion particulaire ( $b_{bp}$ ), qu'ils relient à une table de référence  $b_{bp}$ /concentrations différentielles en particules, issues de simulations théoriques, pour estimer une contribution de chaque classe de taille particulaire à la valeur du  $b_{bp}$  observé.
- Uitz *et al.* (2006) et Brewin *et al.* (2010) estiment la contribution des classes de taille phytoplanctonique à la Chlorophylle-*a* totale à partir de données *in-situ* de contenus pigmentaires.

Mouw *et al.* (2017) ont réalisé un compte rendu détaillé des données nécessaires au développement de quelques unes de ces méthodes, des validations et des avantages et limites des méthodes d'estimation des *PSC* et des *PSD*. D'autres méthodes existent, qui peuvent être régionales ou globales, et développées à partir de diverses variables (e.g., couplage *in-situ*-satellite, approches spectrales), et décrivant la fraction chlorophyllienne de tout ou d'une partie des classes de taille, par exemple Ciotti et Bricaud (2006), Mouw et Yoder (2010), Fujiwara *et al.* (2011), Hirata *et al.* (2011), Bricaud *et al.* (2012), Li *et al.* (2013).

Bien que ces algorithmes soient très utiles, et peuvent suffire à répondre à certaines questions écologiques, d'autres algorithmes ont été développés pour tenter cette fois, de détecter les groupes fonctionnels de façon à faire un lien direct avec les rôles biogéochimiques de chacun.

### Détection par satellite des groupes fonctionnels

Il existe donc des algorithmes qui permettent de détecter un ou plusieurs groupe(s) fonctionnel(s) de phytoplancton. Ces algorithmes peuvent être basés sur la variation spectrale de la luminance induite par la présence d'un groupe de phytoplancton particulier, par exemple en termes d'absorption spécifique, comme c'est le cas de l'algorithme développé par Sathyendranath *et al.* (2004), permettant d'estimer les fréquences d'occurrence des diatomées par rapport à d'autres *populations* mixtes, en étudiant deux modèles d'absorption phytoplanctonique en fonction de la concentration en Chlorophylle-*a* (un modèle mixte et un modèle diatomée). Actuellement, de nombreuses études de ce type ont permis de détecter par satellite un groupe phytoplanctonique, voire un taxon spécifique par exemple la détection de *Trichodesmium spp.*, e.g., Subramaniam *et al.* (1999); Westberry *et al.* (2005), des coccolithophores (e.g., Smyth *et al.* (2002); Sadeghi *et al.* (2012)), des diatomées (Sathyendranath *et al.*, 2004).

Ces approches empiriques ou semi-empiriques sont initialement basées sur des paramètres et des postulats différents, dont les plus répandus sont :

- Des algorithmes sont basés sur l'abondance, ou la concentration en Chlorophylle-*a*, qui, en comparaison avec d'autres paramètres (e.g., absorption, HPLC), permettent de mettre en évidence la présence, la fréquence ou l'abondance spécifique d'un groupe phytoplanctonique ou plus (e.g., Sathyendranath *et al.* (2001)). En effet, l'utilisation de l'abondance en tant qu'indicateur est possible grâce au postulat suivant : une relation existe entre la taille des cellules et des gammes de concentration en Chlorophylle-*a*, ce qui permet de distinguer des classes de taille, voire des taxa spécifiques (diatomées et cyanobactéries par exemple).
- D'autres algorithmes sont basés sur des paramètres écologiques, en utilisant les paramètres qui font varier les communautés phytoplanctoniques (température, lumière, nutriments, couche de mélange etc.), afin de trouver des relations entre la présence spécifique d'un groupe et son environnement (e.g., Raitsos *et al.* (2008)), ce qui provient du postulat que l'environnement chimique et physique induit des communautés phytoplanctoniques différentes (Margalef, 1978).
- Enfin, d'autres algorithmes sont basés sur des paramètres spectraux, en utilisant différents types de réponses spectrales observables spécifiquement pour des groupes phytoplanctoniques (e.g., Brown et Yoder (1994); Sathyendranath *et al.* (2004); Alvain *et al.* (2005); Ciotti et Bricaud (2006); Bracher *et al.* (2009); Sadeghi *et al.* (2012); Organelli *et al.* (2017)). Ces réponses spectrales peuvent être observées sur les IOPs (principalement sur l'absorption), le signal spectral brut, le signal spectral normalisé par la concentration en Chlorophylle-*a* estimée. Ces algorithmes sont basés sur tout ou partie des hypothèses suivantes : la variabilité de l'absorption est induite par la composition phytoplanctonique (Sathyendranath *et al.*, 1987;

Bricaud *et al.*, 1988; Hoepffner et Sathyendranath, 1991; Bricaud *et al.*, 1995; Ciotti *et al.*, 2002; Babin *et al.*, 2003), par la modulation des pigments dans les complexes photosynthétiques (Johnsen *et al.*, 1994), par l'effet d'empaquetage des pigments (Kirk, 1994); la couleur de l'océan est due au premier ordre à la concentration en Chlorophylle-*a*, et au second ordre (après normalisation par la concentration en Chlorophylle-*a*) à la composition phytoplanctonique (en incluant les nuances citées juste avant pour l'absorption).

Bien qu'il soit possible de compiler des algorithmes détectant un seul groupe phytoplanctonique afin d'avoir une idée des communautés présentes, il existe également quelques algorithmes qui, sur la base d'une même analyse, permettent de discriminer plusieurs groupes phytoplanctoniques. Principalement, il existe trois approches en couleur de l'océan pour détecter plusieurs groupes phytoplanctoniques en même temps : OC-PFT (*Ocean Color Phytoplankton Functional Type*, Hirata *et al.* (2011)), PhytoDOAS (*Phytoplankton retrieved by Differential Optical Absorption Spectroscopy*, Bracher *et al.* (2009); Sadeghi *et al.* (2012)) et PHYSAT (*Phytoplankton by satellite*, Alvain *et al.* (2005, 2008)). Les caractéristiques de ces trois approches sont présentées dans la Table 1.1.

TABLE 1.1 – Caractéristiques des trois principales méthodes permettant d'identifier simultanément plusieurs types phytoplanctoniques : OC-PFT, PhytoDOAS et PHYSAT, décrivant leurs caractéristiques, les points forts et points faibles d'après les auteurs respectifs, le IOCCG report (2014) et Mouw *et al.* (2017).

<b>Algorithme</b> Référence(s)	<b>OC-PFT</b> Hirata <i>et al.</i> (2011)	<b>PhytoDOAS</b> Bracher <i>et al.</i> (2009) Sadeghi <i>et al.</i> (2012)	<b>PHYSAT</b> Alvain <i>et al.</i> (2005) Alvain <i>et al.</i> (2008)
Type d'approche	Empirique	Spectroscopie d'absorption optique différentielle	Empirique
Paramètre(s) déterminé(s)	Contribution à la chlorophylle-a des pico-, nano- et micro-phytoplancton, Picoeucaryotes, haptophytes, diatomées, pro-caryotes, algues vertes, dinoflagellés, et <i>Prochlorococcus spp.</i>	Estimation de la chlorophylle-a attribuée aux Cyanobactéries, diatomées, coccolithophores et dinoflagellés	Présence, fréquence et contribution à la chlorophylle-a des nanoeucaryotes, diatomées, <i>Synechococcus spp.-like</i> , <i>Prochlorococcus spp.</i> et <i>Phaeocystis spp.-like</i>
Paramètre(s) satellite + <i>in-situ</i> lors du développement	[Chl-a] + pigments HPLC	Luminance, absorption hyperspectrale	Luminance (4 ou 5 longueurs d'ondes), [Chl-a] + pigments HPLC
Validation	Pigments HPLC	Pigments HPLC et modélisation	Pigments HPLC



Table 1.1 (suite)

<b>Algorithme</b> Référence(s)	<b>OC-PFT</b> Hirata <i>et al.</i> (2011)	<b>PhytoDOAS</b> Bracher <i>et al.</i> (2009) Sadeghi <i>et al.</i> (2012)	<b>PHYSAT</b> Alvain <i>et al.</i> (2005) Alvain <i>et al.</i> (2008)
Point(s) fort(s)	Groupes détectés nombreux (10 au total); méthode applicable à plusieurs capteurs si les relations sont retrouvées; les erreurs de paramétrisation sont déterminées	Estimation de l'abondance; pas directement dépendant de l'estimation de la concentration en chlorophylle- <i>a</i> ; pas de relation empirique, bien que des données <i>in-situ</i> soient utilisées pour la validation; potentiel de développement important avec l'augmentation des capteurs hyperspectraux (e.g., OMI, GOME, Sentinel- <i>x</i> )	N'utilise pas de produits de couleur de l'océan dérivés, ce qui induit moins de propagation des erreurs; utilise la réflectance et l'estimation de la concentration en chlorophylle- <i>a</i> , ce qui est applicable à tous les capteurs de couleur de l'océan
Point(s) faible(s)	Impossible de distinguer des populations avec des distributions de chlorophylle- <i>a</i> associées similaires; difficilement paramétrisable à l'échelle régionale (car nécessite beaucoup de données HPLC pour déterminer les distributions spécifiques)	Nécessite des données hyperspectrales (capteurs moins nombreux que les multispectraux, et plus récents); actuellement, faible couverture spatio-temporelle des données (capteur SCIAMACHY, pixels de 30*30km et 6 jours pour une couverture globale); spectre d'absorption spécifique sensible aux variations de taille de population (changement des espèces et conditions environnementales)	Détection de groupe dominant uniquement; nécessite une recalibration pour différents capteurs (voir Chapitre 2); n'est pas applicable aux bases de données de couleur de l'océan fusionnée actuellement (e.g., OCCCI, GlobColour)

Les trois méthodes décrites en Table 1.1 sont dédiées à la détection simultanée de groupes phytoplanctoniques, ce qui est utile à la modélisation et la compréhension des océans d'un point de vue synoptique. Ces trois méthodes ont d'ailleurs toutes fait l'objet d'utilisations dans diverses études écologiques ou climatiques.

La plupart des algorithmes de détection des PSC, PSD ou PFTs sont empiriques et nécessitent donc des bases de données *in-situ* exhaustives (souvent basées sur les pigments, mais aussi sur des comptages cellulaires, comme pour Raitsos *et al.* (2008)), qui sont analysées différemment selon les méthodes et mises en relation avec différentes sources de données satellites par le biais de mesures coïncidentes.

Dans cette thèse, nous allons nous intéresser au développement de la méthode empirique PHYSAT (Alvain *et al.*, 2005, 2008; Ben Mustapha *et al.*, 2014), qui, à ce jour, détecte plusieurs groupes phytoplanctoniques lorsqu'ils sont dominants dans les eaux de surface, à l'échelle de l'océan global, et à partir de l'anomalie de luminance (Table 1.1, colonne 3).

#### 1.3.2 Principe de la méthode PHYSAT

*Nota bene* Cette Section est destinée à introduire le principe de la méthode PHYSAT telle qu'elle a été établie par le passé. Les travaux de développements et d'extension de cette méthode sont présentés par la suite dans le Chapitre 2 qui fera donc référence à cette base.

PHYSAT est un algorithme qui permet la détection de cinq groupes phytoplanctoniques à l'échelle globale lorsqu'ils sont dominants dans les océans ouverts (Alvain *et al.*, 2005). À l'heure actuelle (2017), la méthode PHYSAT permet de détecter les groupes dominants suivants : les nanoeucaryotes, les diatomées, les *Synechococcus spp.-like*, les *Prochlorococcus spp.* et les *Phaeocystis spp.-like*. Cette méthode est empirique, i.e., elle a été développée à partir de relations observées entre anomalies de la couleur de l'océan et la présence d'un groupe dominant particulier ; qui lui, a été déterminé au préalable à partir de l'analyse des données *in-situ* de concentrations en **pigments biomarqueurs** (i.e., les pigments qui sont des marqueurs spécifiques d'un groupe ou taxon).

Dans la première version de PHYSAT (Alvain *et al.*, 2005), les anomalies de luminance spécifiques à chaque groupe ont été définies par des enveloppes (ou "tubes") qui permettent de caractériser des valeurs minimales et maximales d'anomalies de luminances acceptables pour chaque groupe dominant à chaque longueur d'onde (Figure 1.4).

À l'échelle globale, et selon les archives satellites disponibles, il est possible d'obtenir des cartes journalières de présence de ces 5 groupes, mais aussi des cartes de fréquences de détection globales des groupes par pixel (de 8 jours, jusqu'à la durée de vie du capteur) et

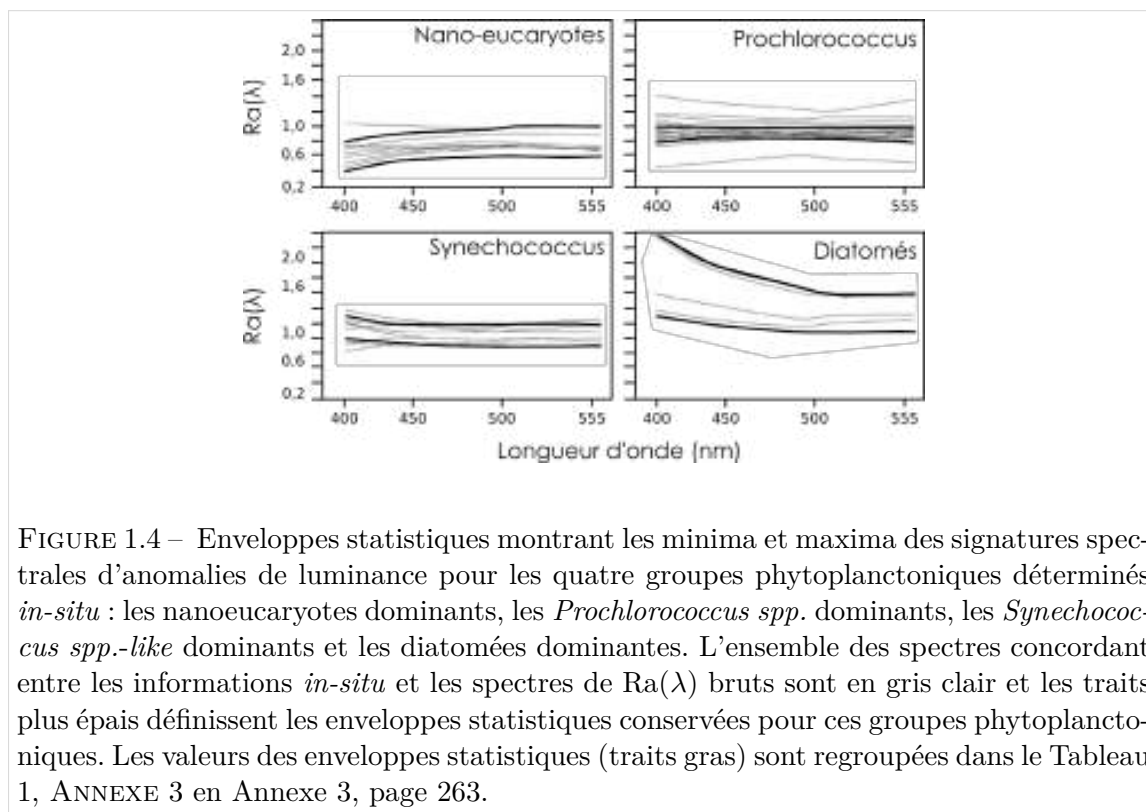
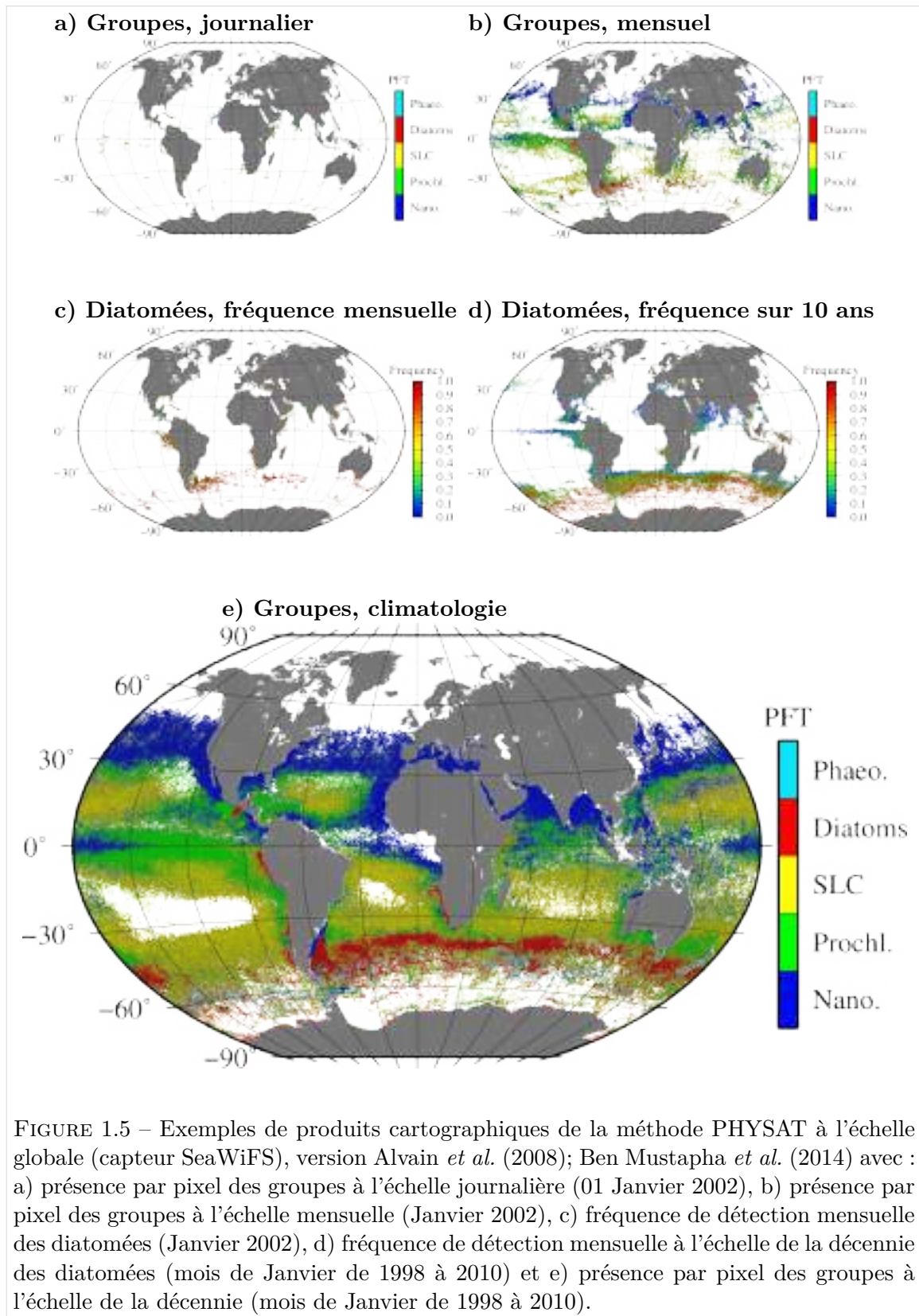


FIGURE 1.4 – Enveloppes statistiques montrant les minima et maxima des signatures spectrales d’anomalies de luminance pour les quatre groupes phytoplanctoniques déterminés *in-situ* : les nanoeucaryotes dominants, les *Prochlorococcus spp.* dominants, les *Synechococcus spp.-like* dominants et les diatomées dominantes. L’ensemble des spectres concordant entre les informations *in-situ* et les spectres de  $Ra(\lambda)$  bruts sont en gris clair et les traits plus épais définissent les enveloppes statistiques conservées pour ces groupes phytoplanctoniques. Les valeurs des enveloppes statistiques (traits gras) sont regroupées dans le Tableau 1, ANNEXE 3 en Annexe 3, page 263.

par région, ainsi que des cartes de répartition des groupes en considérant le groupe le plus fréquent sur une période de temps. Quelques illustrations sont montrées en Figure 1.5. Cette archive est disponible sur demande sur le site : <http://log.cnrs.fr/Physat-2>.

Cette méthode exploite une relation empirique qui a été déterminée entre la présence *in-situ* d’un groupe dominant et une anomalie de couleur de l’océan. Cette anomalie de luminance, ou  $Ra(\lambda)$  (pour *Radiance anomaly*, en anglais), est définie par la différence entre la mesure de luminance observée, pour un pixel donné, et la mesure moyenne "attendue", selon l’archive du capteur et pour la valeur de Chlorophylle-*a* du pixel observé. En effet, la première variable influençant la luminance est la concentration en Chlorophylle-*a* (Morel, 1988), mais d’autres variations des valeurs de luminances, plus faibles, sont exploitables.

Ces anomalies de luminance sont caractérisées par des formes et des amplitudes spectrales, et ce sont ces caractéristiques qui varient d’un groupe phytoplanctonique à un autre (Figure 1.4 et Alvain *et al.* (2005)). Depuis la première publication (Alvain *et al.*, 2005), cette méthode a connu plusieurs développements, empiriques (Alvain *et al.*, 2008), mathématiques avec une nouvelle classification spectrales des données d’anomalies de luminance (Ben Mustapha *et al.*, 2014) et théoriques (Alvain *et al.*, 2012), qui seront repris au Chapitre 2. Cependant, la méthode reste encore limitée par quelques points qui sont détaillés dans la Section suivante.



## 1.4 Problématique et objectifs de la thèse

La méthode PHYSAT (version 2008 et 2014) est utilisée dans de nombreuses applications (e.g., Bopp *et al.* (2005), Volpe *et al.* (2007), Brown *et al.* (2008), D’Ovidio *et al.* (2010)). Bien que PHYSAT ait mis en évidence un lien empirique entre les caractéristiques d’anomalies de luminance et la présence d’un groupe phytoplanctonique spécifique, l’algorithme reste limité pour diverses applications (détails au Chapitre 2) :

- 1 Seuls les cas de dominance phytoplanctonique sont détectés pour 5 groupes à l’échelle globale, ce qui est potentiellement une sous-utilisation du potentiel d’information porté par les anomalies de luminance selon Alvain *et al.* (2012)
- 2 La mise en évidence des relations empiriques dépend de la définition de seuils sur les anomalies de luminance, formant des enveloppes statistiques qui ne doivent pas se recouper, excluant de nombreux signaux potentiellement porteurs d’informations (Figure 1.4, Alvain *et al.* (2005, 2008)) ;
- 3 La nouvelle classification mathématique des anomalies de luminance (Ben Mustapha, 2013; Ben Mustapha *et al.*, 2014) entraîne potentiellement une sur-estimation du nombre de types d’anomalies de luminance, surtout si l’on considère la diversité phytoplanctonique caractérisée<sup>7</sup>.

L’objectif de ma thèse a donc été d’améliorer le potentiel de la méthode PHYSAT, en apportant des outils permettant de définir de nouvelles relations empiriques entre : une classification d’anomalies de luminance améliorée et une analyse différente des données *in-situ* afin d’accéder à l’information "**assemblages**". Ceci a été réalisé en recherchant des relations empiriques entre les anomalies de luminance et la présence d’assemblages de phytoplancton. Ce manuscrit commence par décrire la méthode PHYSAT et ses évolutions depuis sa première publication en Chapitre 2 (à partir de la page 27).

Afin d’atteindre l’objectif annoncé, il faut en priorité dépasser le point n°3 des limitations, qui sera le point de départ des autres études. Ce point est abordé au Chapitre 3 (à partir de la page 55), dans lequel nous avons utilisé une nouvelle classification pour définir les types d’anomalies en fonction de leurs caractéristiques spectrales, mais aussi spatiales et temporelles.

Au Chapitre 4 (à partir de la page 93), nous avons évalué le potentiel de PHYSAT dans la détection d’assemblages phytoplanctoniques à l’échelle globale, en utilisant les mêmes données *in-situ* qu’en 2005 (Alvain *et al.*, 2005), c’est-à-dire des données de concentrations

---

7. La classification de Ben Mustapha *et al.* (2014) permet de définir 100 types d’anomalies de luminance, alors que l’on recherche à identifier cinq groupes phytoplanctoniques (dans l’étude citée). Ceci est expliqué plus en détails au Chapitre 2.

en pigments biomarqueurs, mais analysées de manière innovante, afin d'en extraire une information plus nuancée que les cas de dominance uniquement.

Au Chapitre 5 (à partir de la page 167), ce potentiel de développement a été évalué à l'échelle régionale, en utilisant cette fois de nouvelles données *in-situ* pour la calibration (*Continuous Plankton Recorder*), qui sont décrites avec grande précision (à l'échelle de l'espèce ou du taxon).

*Nota bene* Bien que chaque Chapitre comporte une partie discussion pour les points qui lui sont propres, un dernier Chapitre (à partir de la page 223) est également dédié à cet exercice concernant le potentiel de développement général de la méthode PHYSAT à l'avenir.



# Approche détaillée et historique de la méthode PHYSAT

---

## Introduction

La méthode PHYSAT (PHYtoplancton par SATellite) a été développée depuis 2005 (Alvain *et al.*, 2005). Elle met empiriquement en relation des informations de couleur de l'océan avec la présence d'un groupe phytoplanctonique en surface des océans. C'est une méthode empirique, c'est-à-dire qu'elle a été développée grâce aux liens observés entre les données satellites et les données *in-situ*, chacune de ces données ayant été analysée séparément au préalable.

Depuis les années 1970, il est possible d'estimer la concentration en Chlorophylle-*a* depuis l'espace grâce à la couleur de l'océan et au lancement du capteur CZCS (*Coastal Zone Color Scanner*, en fonctionnement de 1978 à 1986). Bien que l'estimation de la concentration en Chlorophylle-*a* depuis l'espace ait représentée une grande avancée, permettant, entre autres, une compréhension globale et une modélisation plus précise du système océanique, il subsistait un grand intérêt de la communauté dans la détection de groupes particuliers de phytoplancton, afin d'affiner les connaissances à grande échelle en termes de cycles biogéochimiques et de contribution des groupes phytoplanctoniques à la concentration en Chlorophylle-*a*.

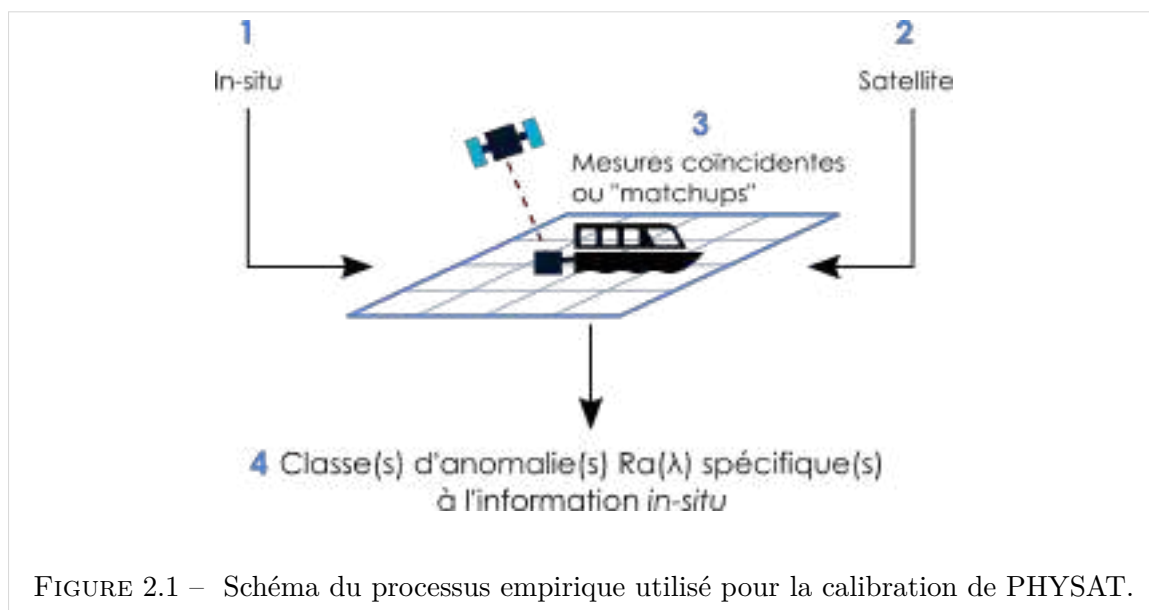
L'idée de la méthode PHYSAT a donc émergée, en alliant la disponibilité d'une grande base de données *in-situ* (donnant accès à une information sur les groupes phytoplanctoniques) et de données satellites plus performantes, ce qui a augmenté les chances d'obtenir des mesures colocalisées, indispensables à l'établissement d'algorithmes empiriques.

Plus précisément, la méthode PHYSAT utilise les "anomalies de la luminance". En effet, la "couleur", ou spectre de luminance de l'océan, dépend principalement de la concentration en Chlorophylle-*a*. Plus il y a de Chlorophylle-*a*, plus l'eau est verte; moins il y en a,



plus la couleur de l'eau s'approche de celle de l'eau pure, c'est-à-dire bleue. L'idée de la méthode PHYSAT est de s'affranchir de l'effet de la Chlorophylle-*a* sur le spectre de luminance de l'océan afin d'observer la forme et l'amplitude du signal restant, appelé anomalie de luminance. Ainsi, pour une concentration en Chlorophylle-*a* donnée, il existe un spectre de luminance moyen, obtenu grâce à un grand nombre de données; et une nuance est observée dans la mesure autour de cette moyenne (l'anomalie de luminance) : c'est cette nuance par rapport au spectre moyen qui a permis de caractériser des signaux spécifiques pour des groupes phytoplanctoniques donnés.

Comme toutes les méthodes empiriques, PHYSAT est basé sur l'observation simultanée des informations mesurées par satellites et observées sur le terrain (Figure 2.1).



Dans la première partie de ce Chapitre, nous allons aborder l'anomalie de luminance (calcul et pré-traitements de l'information). Nous détaillerons la manière de calibrer PHYSAT ainsi que les données utilisables pour la calibration en deuxième partie. En troisième partie, nous détaillerons le lien théorique qui existe entre anomalies de luminance et présence d'un groupe phytoplanctonique ; ce qui nous amènera à discuter des limites actuelles de la méthode et de son potentiel de développement.

## 2.1 Les anomalies de luminance

### 2.1.1 Calcul de l'anomalie de luminance

Comme expliqué en Introduction (Chapitre 1), la principale source de variations de la couleur de l'océan est la concentration en Chlorophylle-*a* (Morel, 1988). La méthode PHYSAT s'est intéressée à l'information dite de "second ordre"<sup>1</sup> de la couleur de l'océan, que l'on appellera par la suite "anomalie"<sup>2</sup> de luminance" (ou  $Ra(\lambda)$  pour *Radiance anomaly* en anglais).

Une "anomalie" est, par définition, établie par rapport à une moyenne qui tient lieu de référence. Comme l'on souhaite supprimer l'effet dû à la concentration en Chlorophylle-*a* sur le spectre de luminance, cette référence se doit d'être une fonction de ce paramètre. Ainsi, pour des petits pas de concentrations en Chlorophylle-*a*, des valeurs de luminance moyennes ont été déterminées. Cela représente la table de référence de PHYSAT, présentant les spectres de luminance moyens pour des concentrations en Chlorophylle-*a* croissantes (Figure 2.2). Cette table de référence a été construite empiriquement à partir de nombreuses mesures satellitaires<sup>3</sup> pour recenser les luminances moyennes mesurées pour une concentration en Chlorophylle-*a* donnée (Alvain *et al.* (2005)). La Figure 2.2 illustre la table de référence du capteur de couleur SeaWiFS. L'encadré dans la Figure 2.2 présente un exemple d'anomalie de luminance, à une concentration estimée en Chlorophylle-*a* de  $0,1 \text{ mg.m}^{-3}$ . À cette concentration, on s'attend à observer une certaine forme du spectre de luminance (spectre bleu dans l'encadré de la Figure 2.2) ; or le spectre observé pour une mesure prise au hasard sur le globe est légèrement différent (spectre orange dans l'encadré de la Figure 2.2), tantôt supérieur à la moyenne (e.g.,  $Ra(412\text{nm})$  sur la Figure 2.2), tantôt inférieur (e.g.,  $Ra(442\text{nm})$  sur la Figure 2.2).

L'anomalie de luminance  $Ra(\lambda)$  est ensuite calculée selon l'Équation suivante, répétée pour toutes les longueurs d'onde du capteur<sup>4</sup> :

$$Ra(\lambda) = \frac{nLw(\lambda)_{\text{obs}}}{nLw_{\text{ref}}(\lambda, \text{Chl-}a)} \quad (2.1)$$

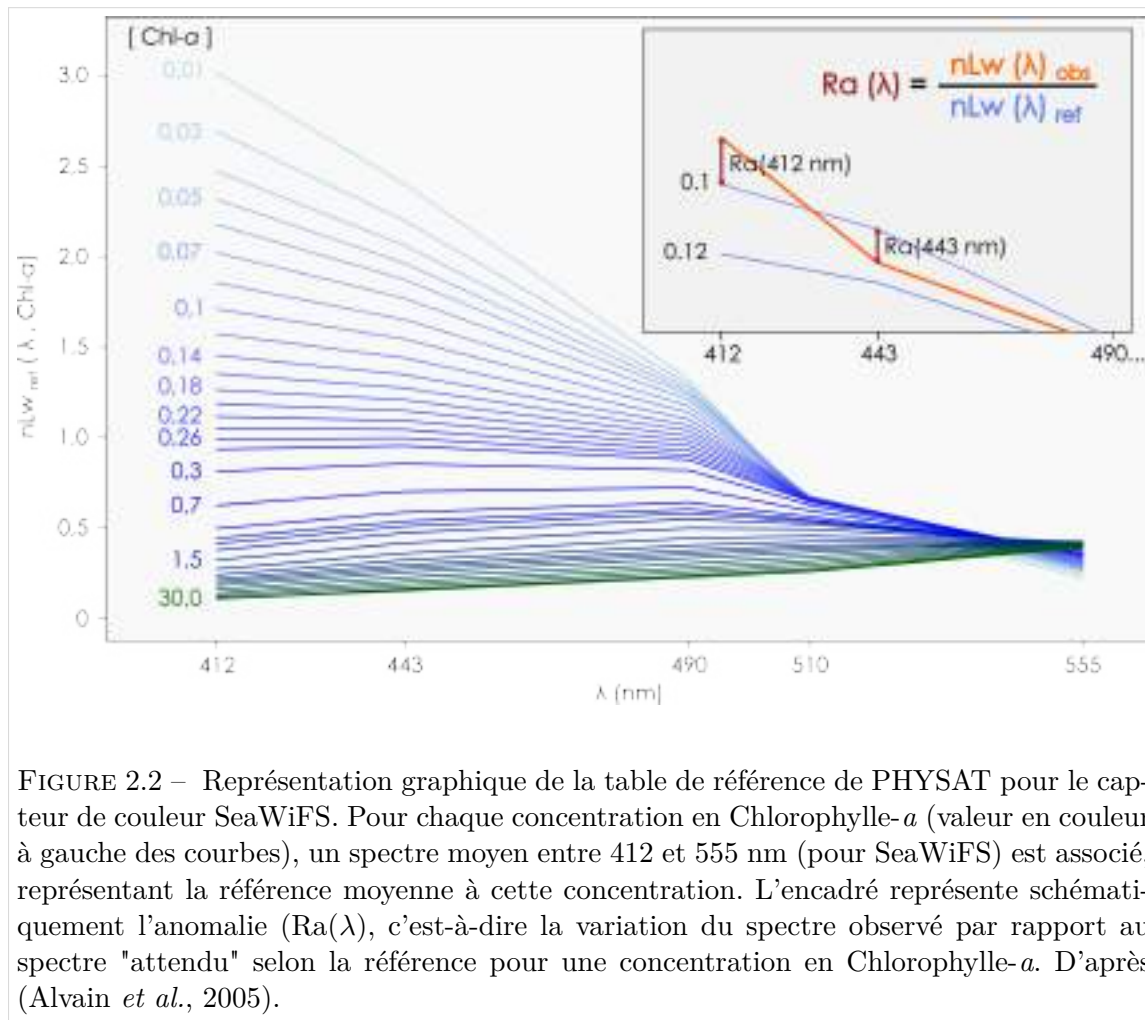
où  $Ra(\lambda)$  est l'anomalie de luminance pour chaque longueur d'onde, tenant compte du ratio entre le spectre  $nLw(\lambda)$  observé sur un pixel par rapport au spectre de référence moyen  $nLw_{\text{ref}}(\lambda, \text{Chl-}a)$  donné pour la concentration en Chlorophylle-*a* estimée sur ce pixel.

1. "Second ordre" de variation de la couleur de l'océan  $nLw(\lambda)$ , considérant que la concentration en Chlorophylle-*a* est la source de premier ordre de variation de  $nLw(\lambda)$

2. Définition de "Anomalie" : Écart positif ou négatif des valeurs d'un élément par rapport à sa moyenne (Dictionnaire Larousse, sens n°4)

3. En pratique, les estimations de concentrations en Chlorophylle-*a* de toute l'archive satellite sont considérées pour calculer le spectre de luminance moyen, en utilisant les sélections détaillées en Section 2.1.2.

4. Les longueurs d'onde diffèrent selon les capteurs, se référer à l'Annexe 1 (page 257) pour la radiométrie de quelques capteurs à l'échelle globale.



Il apparaît donc que  $Ra(\lambda)$  n'a pas de grandeur physique. Sa valeur varie généralement entre 0 et 3 autour du signal moyen. Lorsque  $Ra(\lambda)$  est égal à 1, cela signifie que le signal observé est égal à la moyenne des signaux observés sur le globe, ou une région considérée (en fonction de l'étendue de la table de référence). Au dessus de 1,  $Ra(\lambda)$  est légèrement supérieur à la moyenne observée à large échelle et en dessous de 1, le signal  $Ra(\lambda)$  observé est inférieur à la moyenne observée à large échelle. Artificiellement, en mettant bout-à-bout les  $Ra(\lambda)$  calculés à chaque longueur d'onde du capteur, un spectre de  $Ra(\lambda)$  est obtenu et peut être utilisé pour des études empiriques ou théoriques.

**Exemple de calcul de  $Ra(\lambda)$  :** Un pixel pris au hasard sur le globe et observé pour capteur SeaWiFS, a une concentration en Chlorophylle-*a* estimée de  $0,3978 \text{ mg.m}^{-3}$  (arrondi à  $0,4 \text{ mg.m}^{-3}$ ), et présente les valeurs de luminance  $nLw(\lambda)$  mesurées suivantes :

$\lambda$ (nm)	412	443	490	510	555
$nLw(\lambda)$ mesuré	0,5548	0,6159	0,6150	0,4892	0,2907

À cette concentration en Chlorophylle-*a*, les valeurs  $nLw_{\text{ref}}(\lambda, \text{Chl-}a = 0,4)$  sont :

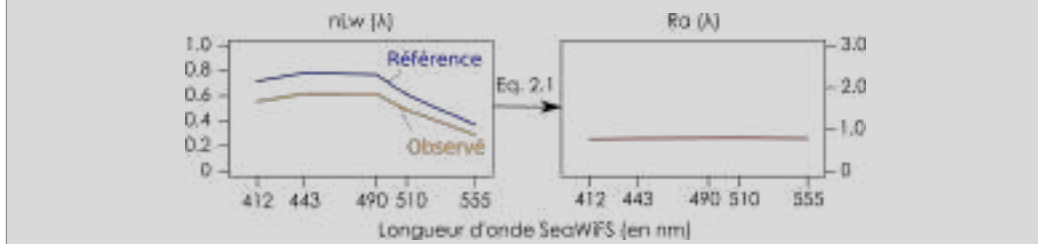
$\lambda$ (nm)	412	443	490	510	555
$nLw_{\text{ref}}(\lambda, \text{Chl-}a)$	0,7181	0,7828	0,7700	0,6097	0,3686

Ces valeurs proviennent de la moyenne d'environ  $8.10^{07}$  pixels qui avaient une concentration en Chlorophylle-*a* estimée à  $0,4 \text{ mg.m}^{-3}$ .

En divisant les valeurs du premier tableau par celles du deuxième tableau (colonne par colonne), comme indiqué dans l'Équation 2.1, les valeurs d'anomalie de luminance par longueur d'onde sont les suivantes :

$\lambda$ (nm)	412	443	490	510	555
$Ra(\lambda)$	0,7726	0,7868	0,7987	0,8024	0,7887

Ce qui permet, à partir du spectre observé (orange sur la Figure suivante) et du spectre moyen (bleu sur la Figure suivante), d'obtenir le spectre "artificiel" de  $Ra(\lambda)$ , en rouge sur la Figure suivante :



### 2.1.2 Critères de qualité permettant le calcul de l'anomalie de luminance

Dans la pratique, le calcul de l'anomalie de luminance est soumis à un certain nombre de critères de qualité, qui permettent d'assurer une fiabilité maximale du calcul de l'Équation 2.1, ainsi que du calcul de la table de référence :

- Les eaux concernées sont du Cas I; ou assimilées (voir définition en Chapitre 1). Dans ces eaux, la couleur de l'océan est influencée principalement par le phytoplancton et les matières dissoutes et particulaires associées; pour des concentrations en Chlorophylle-*a* estimée de  $0,04$  à  $4 \text{ mg.m}^{-3}$

- L'épaisseur optique des aérosols (**AOT** pour *Aerosol Optical Thickness*) (e.g., mesure classique à 865nm pour SeaWiFS) ne doit pas excéder 0,15 (sans unité), ce qui assure une bonne qualité de la correction atmosphérique des luminances  $nLw(\lambda)$ , et doit être homogène pour plusieurs pixels proches dans l'espace (une forte variabilité pour un pavé de 3\*3 pixels indique un risque d'être en bord de nuage).
- Les seuils classiques (ou "*flags*") sur les luminances sont appliqués pour supprimer en amont du processus de calcul tous les spectres qui pourraient avoir entraîné une **saturation du capteur** à une ou plusieurs longueurs d'onde.
- La table de référence  $nLw_{ref}(\lambda, Chl-a)$ , définissant les spectres moyens pour une concentration en Chlorophylle-*a* donnée, doit être définie pour chaque capteur de couleur de l'océan, et pour une zone donnée en cas d'étude particulière (par exemple, l'étude régionale du Chapitre 5). Cela permet de prendre en compte les spécificités des radiomètres, des algorithmes de traitement d'images, et de la zone d'étude.

*Nota bene* Les trois premiers critères sont aussi valables pour la sélection des spectres qui servent à établir les tables de valeurs moyennes de référence pour PHYSAT (exemple en Figure 2.2 pour SeaWiFS).

### 2.1.3 Caractéristiques des anomalies de luminance

Les anomalies de luminance peuvent être représentées comme un spectre sur plusieurs longueurs d'onde du visible. Ce spectre adimensionnel a une amplitude qui dépend, pour les longueurs d'onde prises une à une, du ratio entre la luminance de référence  $nLw_{ref}(\lambda, Chl-a)$  et la luminance observée à une concentration en Chlorophylle-*a* donnée. L'amplitude du spectre d'anomalie de luminance varie généralement entre 0,5 et 2,5.

Ce spectre peut également avoir des formes variées qui peuvent être classées en trois principales formes : plate, concave et convexe. Ces formes sont dues aux différents écarts à la moyenne qui peuvent exister pour les longueurs d'onde. En effet, les écarts à la moyenne ne sont pas linéaires pour toutes les longueurs d'ondes : les plus grands écarts à cette moyenne sont souvent observés pour les longueurs d'ondes en marge (faibles et fortes longueurs d'onde, e.g., 412 et 555nm pour SeaWiFS), donnant les formes de spectres concaves et convexes.

**La forme et l'amplitude du spectre représentent les principales caractéristiques de ce spectre d'anomalie de luminance** (Alvain *et al.*, 2005, 2012). Les principales formes sont illustrées en Figure 2.3. C'est la combinaison forme-amplitude qui pourra être

associée avec un groupe spécifique de phytoplancton (Alvain *et al.*, 2005, 2008, 2012).

Les spectres d’anomalies à eux-seuls représentent un premier niveau d’information concernant les groupes phytoplanctoniques (Alvain *et al.*, 2012), même lorsqu’il est impossible de caractériser précisément cette situation (i.e., d’identifier un nom de groupe phytoplanctonique pour un spectre d’anomalie particulier).

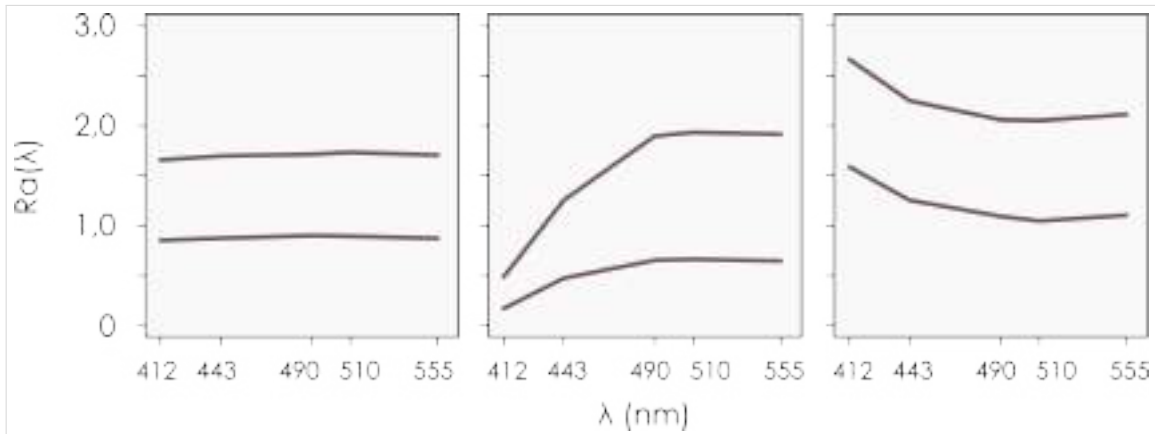


FIGURE 2.3 – Exemple de spectres d’anomalies de luminance  $Ra(\lambda)$  illustrant les 3 principales formes de spectres (de droite à gauche) : plat, concave et convexe; et différentes amplitudes. Extrait du calcul des anomalies de luminance pour le capteur SeaWiFS.

#### 2.1.4 Classification des anomalies de luminance

*Nota bene* Ce paragraphe est consacré aux caractéristiques des  $Ra(\lambda)$ . Le processus empirique qui a permis d’associer ces enveloppes de  $Ra(\lambda)$  à des groupes spécifiques de phytoplancton est détaillé ensuite, dans la Section 2.2.

Lors de la première utilisation des anomalies de luminance (Alvain *et al.*, 2005), ces données ont été utilisées de manière brute, c’est-à-dire que, pour chaque information *in-situ*, les spectres d’anomalies de luminance, tels que calculés par l’Équation 2.1, ont été collectés. De cette manière, chaque couple d’information *in-situ*-satellite était unique. Pour établir un lien empirique, il a donc été nécessaire de définir, *a posteriori*, quels types de spectres (en forme et amplitude), correspondaient à des situations phytoplanctoniques déterminées. Dans cette version de PHYSAT (Alvain *et al.*, 2005, 2008), des enveloppes ont donc été définies pour caractériser les situations phytoplanctoniques, excluant les recouvrements entre les différentes enveloppes de  $Ra(\lambda)$  pour des groupes phytoplanctoniques différents. Ces enveloppes sont présentées en Figure 2.4 (les valeurs définissant chaque enveloppe pouvant être trouvées dans Alvain *et al.* (2005) et en Annexe 3, page 263).

Le problème lié aux enveloppes d’anomalies de luminance minimales et maximales accep-

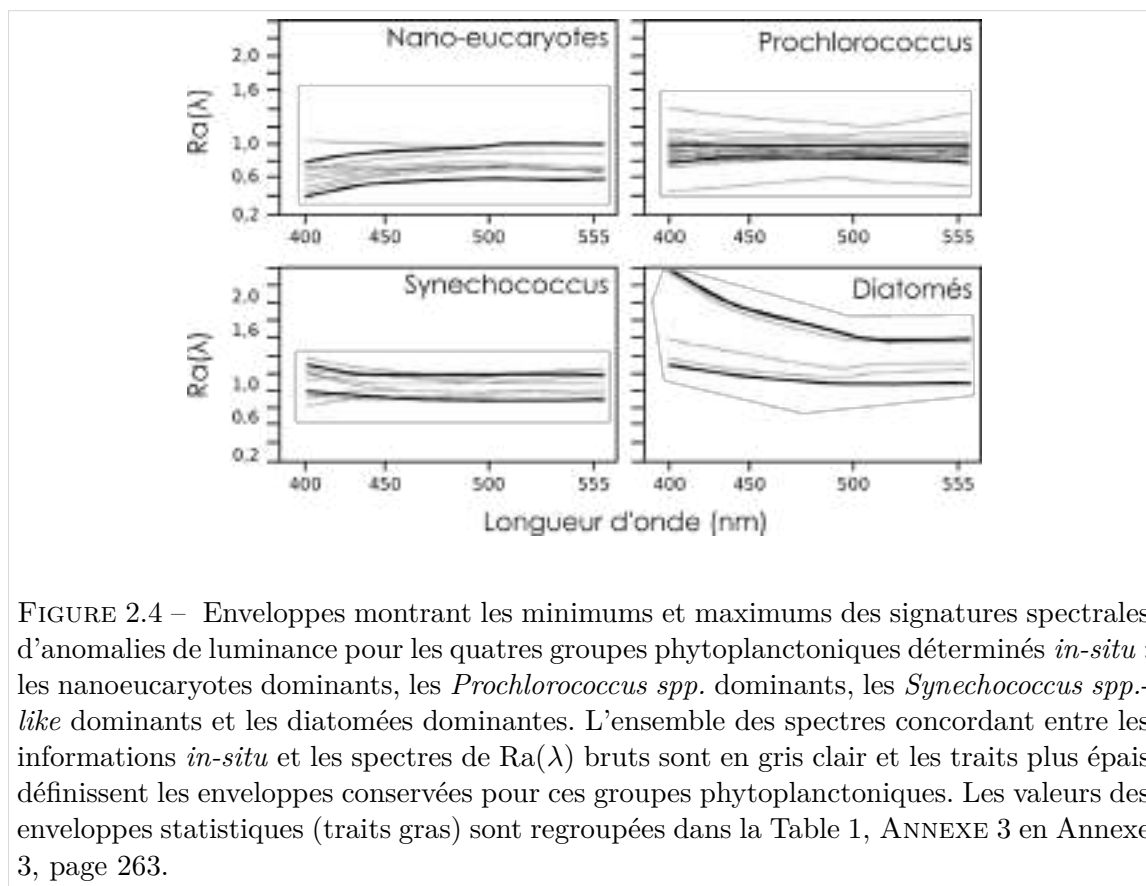


FIGURE 2.4 – Enveloppes montrant les minimums et maximums des signatures spectrales d’anomalies de luminance pour les quatre groupes phytoplanctoniques déterminés *in-situ* : les nanoeucaryotes dominants, les *Prochlorococcus spp.* dominants, les *Synechococcus spp.-like* dominants et les diatomées dominantes. L’ensemble des spectres concordant entre les informations *in-situ* et les spectres de  $Ra(\lambda)$  bruts sont en gris clair et les traits plus épais définissent les enveloppes conservées pour ces groupes phytoplanctoniques. Les valeurs des enveloppes statistiques (traits gras) sont regroupées dans la Table 1, ANNEXE 3 en Annexe 3, page 263.

tables pour les groupes phytoplanctoniques est que certains spectres ne peuvent pas être caractérisés comme reflétant la présence d’un groupe, alors que l’information *in-situ* le montre. Ces spectres sont les spectres gris clairs qui "sortent" des enveloppes (traits épais) sur la Figure 2.4. De plus, ces enveloppes ont été définies de manière "intuitive", et sont donc développeur-dépendantes.

Ainsi, il a fallu trouver une nouvelle approche afin de mieux caractériser l’ensemble de la diversité des  $Ra(\lambda)$  à partir de leur forme et amplitude, et de manière plus condensée. Pour cela, une classification est apparue comme étant la meilleure méthode : elle permet de grouper les  $Ra(\lambda)$  similaires sous forme de classes de  $Ra(\lambda)$ , puis de définir les caractéristiques de chaque classe de  $Ra(\lambda)$ . Un réseau de neurone de type carte auto-organisatrice (ou **SOM** pour *Self Organizing Maps*, Kohonen (2013)) a été utilisé (Ben Mustapha, 2013; Ben Mustapha *et al.*, 2014).

L’utilisation du réseau de neurones de type SOM non supervisé dans le cadre de la classification des anomalies de luminance présente plusieurs avantages :

- Ce type d’algorithme de regroupement est capable de traiter un grand volume de données non linéaires et bruitées.

- Le nombre de classes final peut être choisi, ce qui permet d'affiner la classification à souhait, en fonction de ce que l'on cherche à obtenir (minimisation des erreurs de classification, pertinence des classes selon l'objectif, variabilité acceptée).
- Le réseau de neurone conserve la typologie initiale du jeu de données; il est peu affecté par les données aberrantes ou bruitées; et ne nécessite pas de connaissances préliminaires sur les données (e.g., hypothèses sur la distribution).

La classification des  $Ra(\lambda)$  par l'algorithme *SOM* se déroule en deux étapes :

- 1 Une carte est construite, sur la base d'un jeu de données de  $Ra(\lambda)$  aléatoire : c'est la phase d'apprentissage;
- 2 L'ensemble des  $Ra(\lambda)$  sont projetés sur cette carte : c'est la phase d'exploitation.

Dans les paragraphes suivants, nous allons rappeler le principe de la carte auto-organisatrice, puis détailler les caractéristiques de la carte auto-organisatrice dans le contexte de la méthode PHYSAT. Ce travail a été réalisé dans le cadre de la thèse de Z. Ben Mustapha (Ben Mustapha, 2013), dont les paragraphes ci-dessous sont un bref résumé.

### Les principes de base de l'algorithme de Kohonen (Kohonen, 2013)

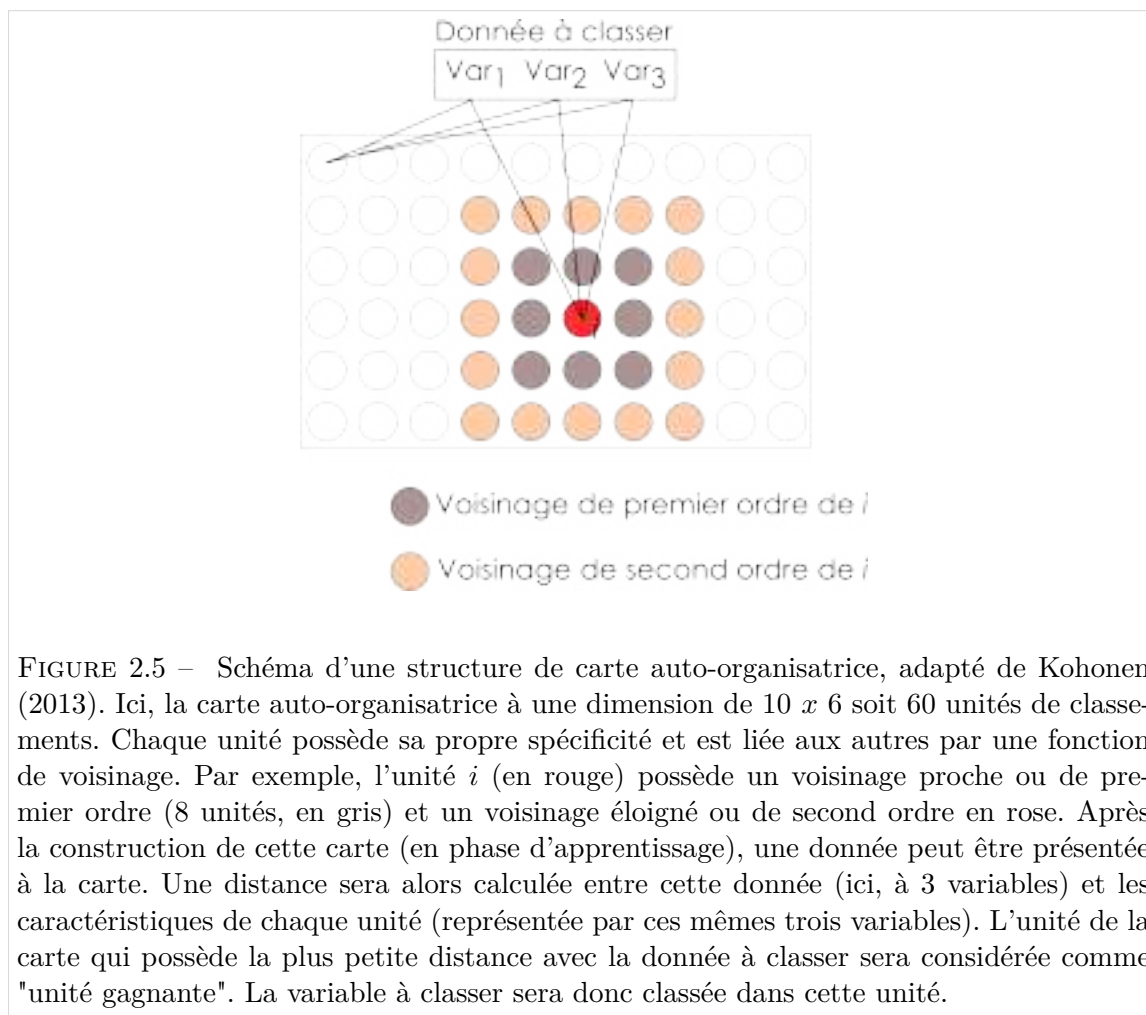
Les cartes auto-organisatrices sont utilisées pour cartographier un espace de données de grandes dimensions. Dans la pratique, ces cartes auto-organisatrices sont donc utilisées pour réaliser de la classification de données. Elles font partie des méthodes de classification de type "réseau de neurones", qui rappelle leur capacité à reproduire le fonctionnement des cartes neuronales du cerveau. En effet, chaque unité de la carte est spécialisée dans la représentation d'un groupe de données qui possède ses spécificités, tout en étant plus ou moins liée aux unités adjacentes. Schématiquement, les cartes auto-organisatrices à deux dimensions<sup>5</sup> sont représentées comme en Figure 2.5. C'est une surface plane ( $l$  lignes et  $c$  colonnes) où chaque *neurone*, c'est-à-dire chaque unité de la carte, est connecté avec ses voisins par une fonction de voisinage. La fonction de voisinage est particulièrement importante en phase d'apprentissage (construction de la carte), qui se déroule en plusieurs étapes :

- 1 **La phase d'initiation** : Après avoir choisi une dimension et une structuration de carte, des données (souvent aléatoires, mais qui doivent être représentatives du problème à traiter) sont présentées à l'algorithme de Kohonen. Cela permet d'établir une "pré-carte" où chaque unité possède déjà les spécificités des données aléatoires d'entrées, que l'on appelle des "références". Les références sont une représentation

---

5. Les cartes auto-organisatrices peuvent avoir plus de deux dimensions pour des problématiques très complexes.





condensée de l'information des données initiales. Par exemple, si les données à classer possèdent cinq variables (comme c'est le cas des  $Ra(\lambda)$ ), la référence d'une unité présentera les informations condensées de ces cinq variables. Chacune des unités possèdera des caractéristiques différentes de ces cinq variables (si la structuration de la carte est adéquate).

**2 La phase d'apprentissage (itérations) :** dans cette étape, des observations aléatoires vont être présentées à la carte. Chaque observation va être comparée (le plus souvent au moyen d'une distance euclidienne) à l'ensemble des références de la carte. Cela permet de déterminer une "unité gagnante", c'est-à-dire la classe de la carte pour laquelle la distance entre la nouvelle observation et la référence de cette classe est la plus petite. Les valeurs de références de cette classe vont alors être modifiées pour correspondre à l'ensemble des données dans cette classe (i.e., les données initiales de la phase d'initiation + cette nouvelle donnée). De même, les valeurs de références des classes alentours (voisinage, détaillé en Figure 2.5) vont être modifiées selon une fonction de voisinage (cette modification étant pondérée en fonction de la distance à l'unité gagnante). Cette étape est répétée plusieurs fois,

au cours desquelles la fonction de voisinage devient de moins en moins "puissante", c'est-à-dire qu'elle affecte un voisinage de plus en plus restreint.

**3 Fin de la construction de la carte auto-organisatrice :** la configuration optimale de la carte auto-organisatrice est obtenue lorsque l'ajout de nouvelles données ne modifient plus significativement les valeurs de référence des unités de la carte. De plus, des paramètres statistiques permettent d'évaluer la qualité de cette carte par rapport aux données initiales : les *erreurs de quantification vectorielle*, l'*erreur topologique*, et le *paramètre de distorsion*.

*Nota bene* Les fonctions de voisinage, ainsi que leur modification au cours des itérations de la carte, n'ont pas été abordées dans le détail dans cette partie focalisée sur le principe de la méthode PHYSAT ; ces détails peuvent être trouvés dans Kohonen (2013) (principe général) et dans Ben Mustapha (2013) (application aux  $Ra(\lambda)$ ).

**Projection de nouvelles données.** Une fois que la carte auto-organisatrice est considérée comme stable selon les indices statistiques ci-dessus, de nouvelles données (par exemple, d'autres valeurs de  $Ra(\lambda)$ , issues de cartes journalières) peuvent être projetées sur cette carte afin qu'elles aussi trouvent leur unité gagnante. La carte auto-organisatrice sert ainsi d'espace de référence à la classification de nouvelles données non intégrées en phase d'apprentissage. Lors de cette étape de projection, les valeurs des références des classes ne sont plus modifiées. En revanche, si la distance entre une nouvelle observation et l'unité gagnante la plus proche de cette observation est trop grande, ou si les plus petites distances entre la nouvelle donnée et plusieurs unités de la carte sont trop proches, cela signifie que la carte auto-organisatrice ne prend pas assez en compte la diversité des cas présents dans un grand jeu de données, ou que sa structuration n'est pas optimale. Cela peut donc suggérer (si ce cas est fréquent) que la carte n'a pas été initiée avec suffisamment de données, voire que le nombre de classes n'est pas adéquat (trop grand ou trop petit).

**Application à PHYSAT : Phase d'apprentissage de la classification des anomalies de luminance par une carte auto-organisatrice de Kohonen (Ben Mustapha et al., 2014)**

Dans cette première étape, il est question de définir la base de la carte auto-organisatrice des anomalies de luminance. Cette carte sera effectuée pour chaque table de référence de PHYSAT afin de prendre en compte le signal moyen, utilisé comme référence, correspondant à chaque capteur, ou chaque région d'étude.

Afin de définir la structuration de la carte auto-organisatrice dans le cadre de la classification des anomalies de luminance, plusieurs tests ont été effectués (Ben Mustapha, 2013). Cette étude a montré que, à l'échelle globale, et pour le capteur de couleur de l'océan

SeaWiFS, une centaine de classes (i.e., de types de  $Ra(\lambda)$  de référence) permettait de caractériser un maximum de type de spectres de  $Ra(\lambda)$  (combinaisons de forme et d'amplitude), tout en minimisant les erreurs de classification (voir ci-dessus). Cette carte est carrée (ses dimensions sont de 10 x 10), et la fonction de voisinage est gaussienne.

La carte a été initiée avec un grand nombre de spectres qui ont été choisis au hasard mais de manière à échantillonner toutes les régions, et en toutes saisons.

**Application à PHYSAT : résultats de la classification des  $Ra(\lambda)$  sur une carte auto-organisatrice (Ben Mustapha *et al.*, 2014), exemple pour le capteur SeaWiFS, à l'échelle globale**

La Figure 2.6 illustre les spectres référents (i.e., l'information condensée des spectres de  $Ra(\lambda)$  classées dans ces unités) des 100 unités de la carte auto-organisatrice créée pour classer les anomalies de couleur de l'océan du capteur SeaWiFS, à l'échelle globale. Les résultats des cartes auto-organisatrices pour les autres capteurs de couleur de l'océan sont disponibles en Annexe 2 (page 259). Sur ces cartes, l'ordre topologique est bien respecté en fin de phase d'apprentissage (les unités adjacentes de la carte ont des spectres de  $Ra(\lambda)$  similaires).

Lorsque la carte est construite (Figure 2.6 et Annexe 2, page 259), il est possible de projeter les  $Ra(\lambda)$  observés sur cette carte, afin que chaque pixel soit associé à une unité de cette carte, à condition que la forme et l'amplitude du spectre respecte bien les informations du vecteur de référence. Ainsi, il est donc possible qu'un pixel particulier ne soit pas attribué à une classe, tant la forme et l'anomalie du spectre de  $Ra(\lambda)$  sont différentes des références<sup>6</sup>. Le but est d'éviter que les données classées en phase d'exploitation de la carte divergent trop (i.e., soient en dehors de la dispersion des écarts types) des vecteurs référents.

Lorsque les pixels ont été associés à un type de spectre en accord avec la classification de Kohonen, il est possible d'établir des cartes d'occurrences et de fréquences de tout ou partie des classes. Des exemples de cartes sont donnés en Figure 2.7. Dans la Figure 2.7 (gauche), on peut remarquer que des structures cohérentes apparaissent à large échelle spatio-temporelle, ce qui est également un critère de qualité dans le jugement de la carte, si l'on se base sur l'hypothèse que ces classes de  $Ra(\lambda)$  sont elles-mêmes porteuses d'information sur les groupes phytoplanctoniques.

---

6. En moyenne, environ  $3,9 \pm 0,4$  % des pixels contenant une information  $Ra(\lambda)$  ne peuvent pas être attribués à une classe de  $Ra(\lambda)$  *som*.

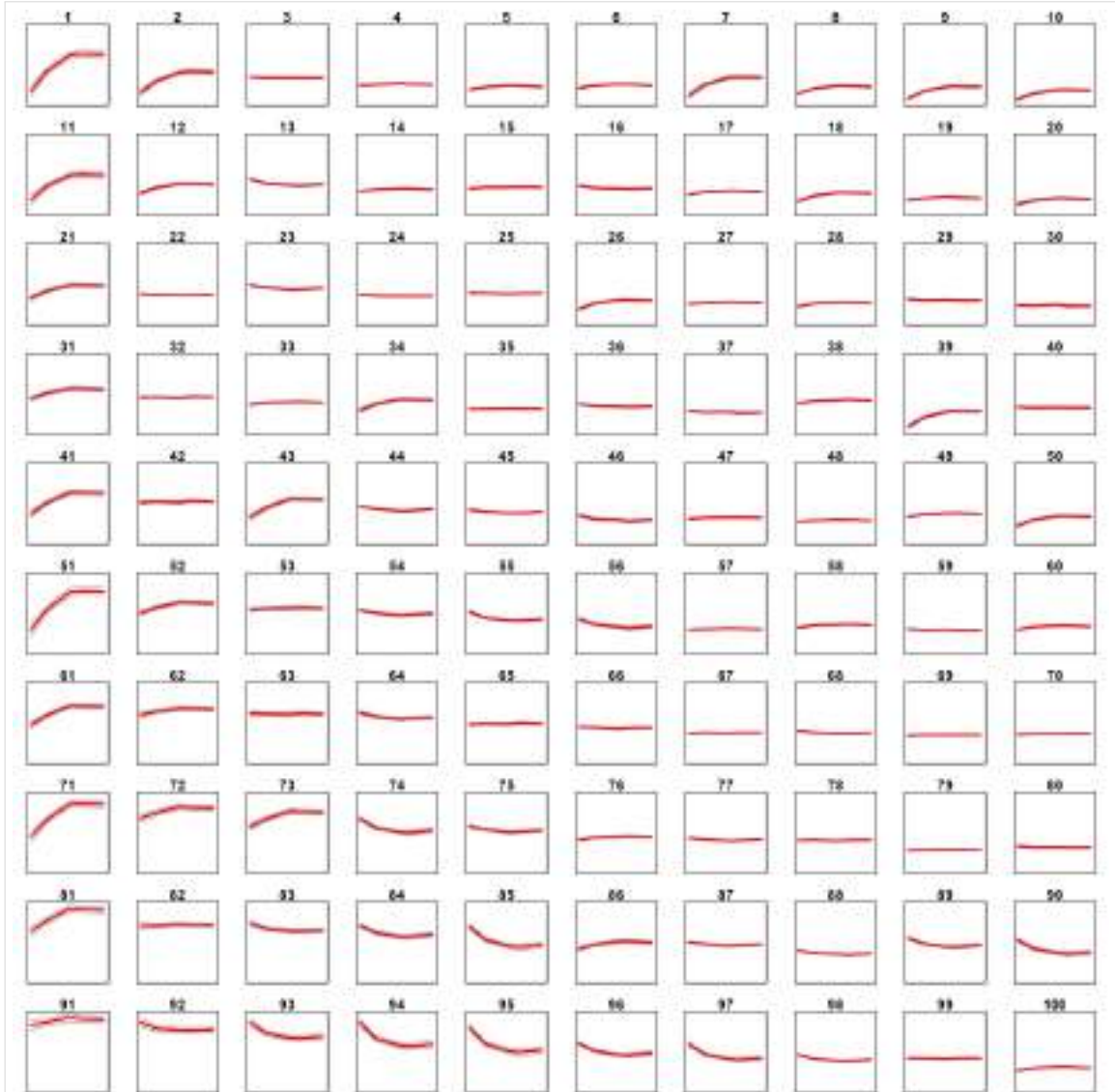


FIGURE 2.6 – Spectres de références (informations condensées et classées de toutes les entrées) des unités de la carte auto-organisatrice des anomalies de luminance du capteur de couleur de l’océan SeaWiFS, avec écarts types (lignes grises), à l’échelle globale. Pour chaque graphique, l’axe des abscisses représente la longueur d’onde (412,443,490,510 et 555 nm pour SeaWiFS), et l’axe des ordonnées représente l’anomalie de luminance (sans unité), réalisé avec l’algorithme de Kohonen (Kohonen, 2013), sur la base de Ben Mustapha *et al.* (2014).

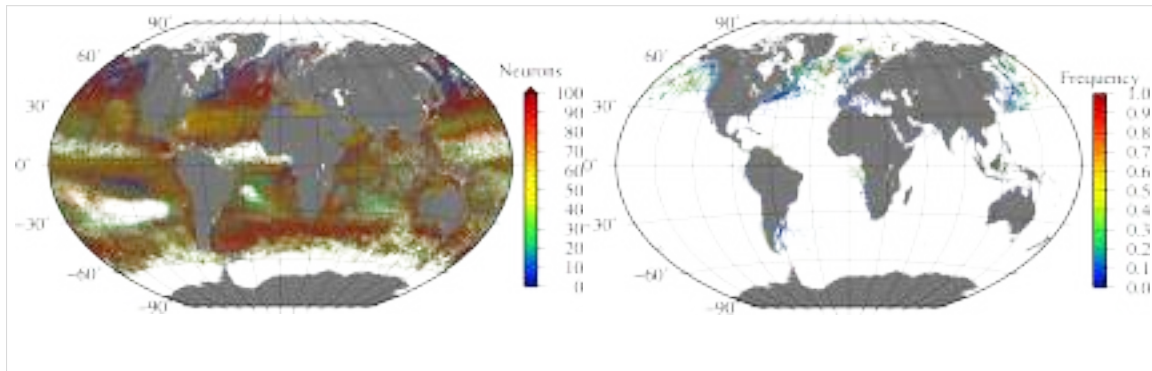


FIGURE 2.7 – Exemples de cartes d’occurrence des classes de  $Ra(\lambda)$ , à partir des données du capteur SeaWiFS en climatologie de 1998 à 2010 avec, à gauche, les classes de  $Ra(\lambda)$  dominantes au mois d’Avril (classe la plus fréquente pour chaque pixel des mois d’Avril 1998 à Avril 2010) et, à droite, la fréquence d’apparition de la classe n°20 au mois d’Avril (fréquence d’apparition de la classe par pixel pour les mois d’Avril 1998 à Avril 2010).

### Résumé : Les anomalies de luminance

Les anomalies de luminance sont calculées à partir d’une table de référence, construite pour chaque capteur ou chaque région spécifique. Elles sont caractérisées par une forme et une amplitude, qui est un premier niveau d’information sur la composante phytoplanctonique de l’océan. Ces anomalies peuvent être classées par une carte auto-organisatrice, qui permet d’obtenir 100 classes statistiquement différentes. Ces 100 classes représentent la diversité de formes et d’amplitudes des  $Ra(\lambda)$  que l’on peut obtenir. La prochaine étape du travail est d’essayer de caractériser les anomalies de luminance, ou les classes d’anomalies de luminance en termes de communautés phytoplanctoniques : c’est la *labellisation*.

## 2.2 Objectifs de la labellisation empirique : donner un nom aux anomalies de luminance observées

La *labellisation* est le fait de donner un nom à une anomalie de luminance, en fonction de l’analyse des observations *in-situ* disponibles. Une labellisation n’est donc jamais définitive et peut être remise en cause par l’ajout de nouvelles données *in-situ*.

Des labellisations des anomalies de luminance PHYSAT ont déjà été effectuées par le passé de deux manières :

- Les données *in-situ* sont associées à des anomalies de luminance spécifiques lors d’une procédure de calibration empirique (basée sur l’analyse de mesures coïncidentes). En pratique, les paires récurrentes groupes phytoplanctoniques / type(s) d’anomalies de luminance spécifique(s) sont recherchées.
- Une expertise terrain permet de savoir qu’un groupe de phytoplancton particulier est présent sur une zone et à un moment spécifique (informations basées sur des observations *in-situ*). Un type d’anomalie de luminance particulier est également observé au même endroit et pour les mêmes périodes. Bien que les mesures coïncidentes (même latitude, longitude et date) n’existent pas au moment de l’analyse, l’expertise de terrain permet de labelliser ces anomalies de luminance : c’est la labellisation par expertise. Dans ce cas, une validation des observations n’est pas possible dans l’immédiat, et il faudra collecter des échantillons *in-situ* pour valider les labellisations.

Pour le premier point, les données *in-situ* utilisées pour la calibration de PHYSAT, et donc la labellisation de certains types d’anomalies, ont été de diverses natures à ce jour : concentration en **pigments biomarqueurs** (Alvain *et al.*, 2005, 2008), données de **cytométrie en flux** (Thyssen *et al.*, 2015). Ces deux points, ainsi que la calibration par expertise (Alvain *et al.*, 2008), sont détaillés ci-dessous.

Nous avons vu dans la Section 2.1 que les anomalies se présentaient sous deux formes : leur forme brute (anomalie de luminance), ou leur forme classée (classe d’anomalies de luminance). La labellisation peut s’effectuer pour l’un ou l’autre de ces formats. Les paragraphes suivants détaillent les labellisations qui ont conduit à la problématique de cette thèse.

### 2.2.1 À partir d’une base de données de concentration en pigments biomarqueurs publiée par le passé (Alvain *et al.*, 2005)

Lors de la première calibration de PHYSAT, les données pigmentaires, et plus spécifiquement les **pigments biomarqueurs**, estimés par **HPLC** (*High Pressure Liquid Chromatography*) ont été utilisés. En effet, ce sont les données relatives au phytoplancton qui sont le plus communes et diversifiées spatialement et temporellement, avec une grande distribution en océan ouvert. Elles sont souvent estimées lors de campagnes océanographiques à travers le monde (Jeffrey *et al.*, 1997).

Alvain *et al.* (2005) ont utilisé les concentrations de pigments biomarqueurs suivants :

Péridinine (abrégié *Peri*, biomarqueur des dinoflagellés), Fucoxanthine (abrégié *Fuco*, biomarqueur des diatomées), Zéaxanthine (abrégié *Zea*, biomarqueur des *Synechococcus*-like), 19'-HexanoyloxyFucoxanthine (abrégié *HF19*, biomarqueur des Nanoeucaryotes), Divinyl-Chlorophylle-*a* (abrégié *Div-chla*, biomarqueur des *Prochlorococcus*). Ces concentrations ont été estimées lors des campagnes GEP&CO (Dandonneau *et al.*, 2004), qui ralliaient Le Havre à la Nouvelle-Calédonie en passant par le canal de Panama.

La concentration en pigments n'étant pas le reflet d'une abondance, Alvain *et al.* (2005) ont pallié ce problème en utilisant les concentrations en pigments relatives aux concentrations en Chlorophylle-*a* (Équation 2.2). Des seuils statistiques ont ensuite été évalués à partir de la littérature pour permettre la distinction des cas de dominances, c'est-à-dire les cas pour lesquels le groupe phytoplanctonique représente plus de 50% de la concentration en Chlorophylle-*a*<sup>7</sup> (Alvain *et al.*, 2006, 2008). Ces seuils sont présentés en Table 2.1.

$$[\text{Pigment}]_{\text{Rel}} = \frac{[\text{Pigment}]}{[\text{Chl-}a] + [\text{Div. Chl-}a]} \quad (2.2)$$

TABLE 2.1 – Seuils sur les concentrations relatives en pigments biomarqueurs utilisés pour associer empiriquement un inventaire de pigments à un groupe de phytoplancton dominant spécifique. Les valeurs en gras correspondent aux biomarqueurs de chaque groupe nommé en première colonne. Extrait de Alvain *et al.* (2005).

	Div.a	Périd.	Fuco.	HF'19	Zéa.
Diatomées	<0,40	<0,10	> <b>0,18</b>	-	<0,20
Prochlorococcus	> <b>0,40</b>	<0,10	-	-	>0,35
Nanoeucaryotes	<0,40	<0,10	-	> <b>0,14</b>	<0,20
<i>Synechococcus-Like</i>	<0,40	<0,10	-	-	> <b>0,20</b>
Dinoflagellés	<0,40	> <b>0,10</b>	-	-	<0,20

Dans cette première étape qui concerne les données *in-situ* uniquement, la méthode a permis de mettre en évidence, au sein des données *in-situ* GEP&CO, qu'il existait des échantillons *in-situ* dominés par un groupe particulier de phytoplancton (soit des diatomées, des nanoeucaryotes, des *Prochlorococcus spp.* ou des *Synechococcus spp.-like*).

La seconde étape a été la réalisation des mesures coïncidentes entre les données *in-situ* et les anomalies de luminance. Comme il a été expliqué auparavant, à ce moment, les anomalies de luminance n'étaient pas groupées en classes, mais étaient utilisées telles quelles (les  $R_a(\lambda)$  issues de l'Équation 2.1). Ainsi, en face de chaque groupe phytoplanctonique dominant, composé de plusieurs échantillons *in-situ*, il y avait une collection d'anomalies

7. Dans ce contexte, un seuil de "100%" représentant les ratios sur les concentrations pigmentaires en cas de présence exclusive du groupe phytoplanctonique considéré.

TABLE 2.2 – Scores de validation des associations trouvées pour PHYSAT, en pourcentage (Alvain *et al.*, 2005). Les bonnes identifications sont mises en évidence sur fond bleu. Les mauvaises identifications (autres cellules du tableau) sont séparées en fonction des groupes. Exemple de lecture (ligne 1) : Les nanoeucaryotes identifiés par le signal PHYSAT correspondent effectivement à des nanoeucaryotes dans 82% des données *in-situ* de validations. Le signal est confondu à 10% avec des *Prochlorococcus spp. in-situ*, à 5 % avec des *Synechococcus spp.-like*, et à 3% avec des diatomées.

In-situ → PHYSAT ↓	Nano.	Proc.	Syne.	Diat.
Nanoeucaryotes	82	10	5	3
<i>Prochlorococcus spp.</i>	14	61	25	0
<i>Synechococcus spp.-like</i>	20	23	57	0
Diatomées	24	0	3	73

de luminance à analyser. Ces anomalies de luminance avaient des formes et des amplitudes qui étaient différentes en fonction du groupe dominant, ce qui signifiait qu'il existait une "signature spectrale" sur l'anomalie de luminance pour des groupes spécifiques de phytoplancton.

Afin de pouvoir cartographier ces anomalies spécifiques, des seuils ont été déterminés sur les valeurs d'anomalies de luminance pour chaque longueur d'onde, formant des enveloppes, qui caractérisent spécifiquement ces groupes phytoplanctoniques dominants déterminés *in-situ* (Figure 2.4, page 34).

Ces labellisations ont été validées avec succès si l'on considère le niveau de connaissance et les moyens de traitement des données utilisés à l'époque. Pour une validation indépendante, un jeu de données, non-utilisé pour la calibration, a été utilisé. Les anomalies de luminance labellisées, et cartographiées, ont été identifiées par mesures coïncidentes à des données de validation et les résultats ont été comparés. Les scores de validations (bonnes et mauvaises identifications) sont présentés en Table 2.2.

Une fois que les critères ont été établis sur les anomalies de luminance, et validés, il a été possible de rechercher, dans toute l'archive des données de couleur de l'océan (Figure 2.8, pour le capteur SeaWiFS), les spectres qui correspondent à ces critères. Cela permet de réaliser des cartes synthétiques de groupes dominants à l'échelle des océans du globe, et sur les périodes de temps pouvant aller de journée, jusqu'à la décennie (pour SeaWiFS).



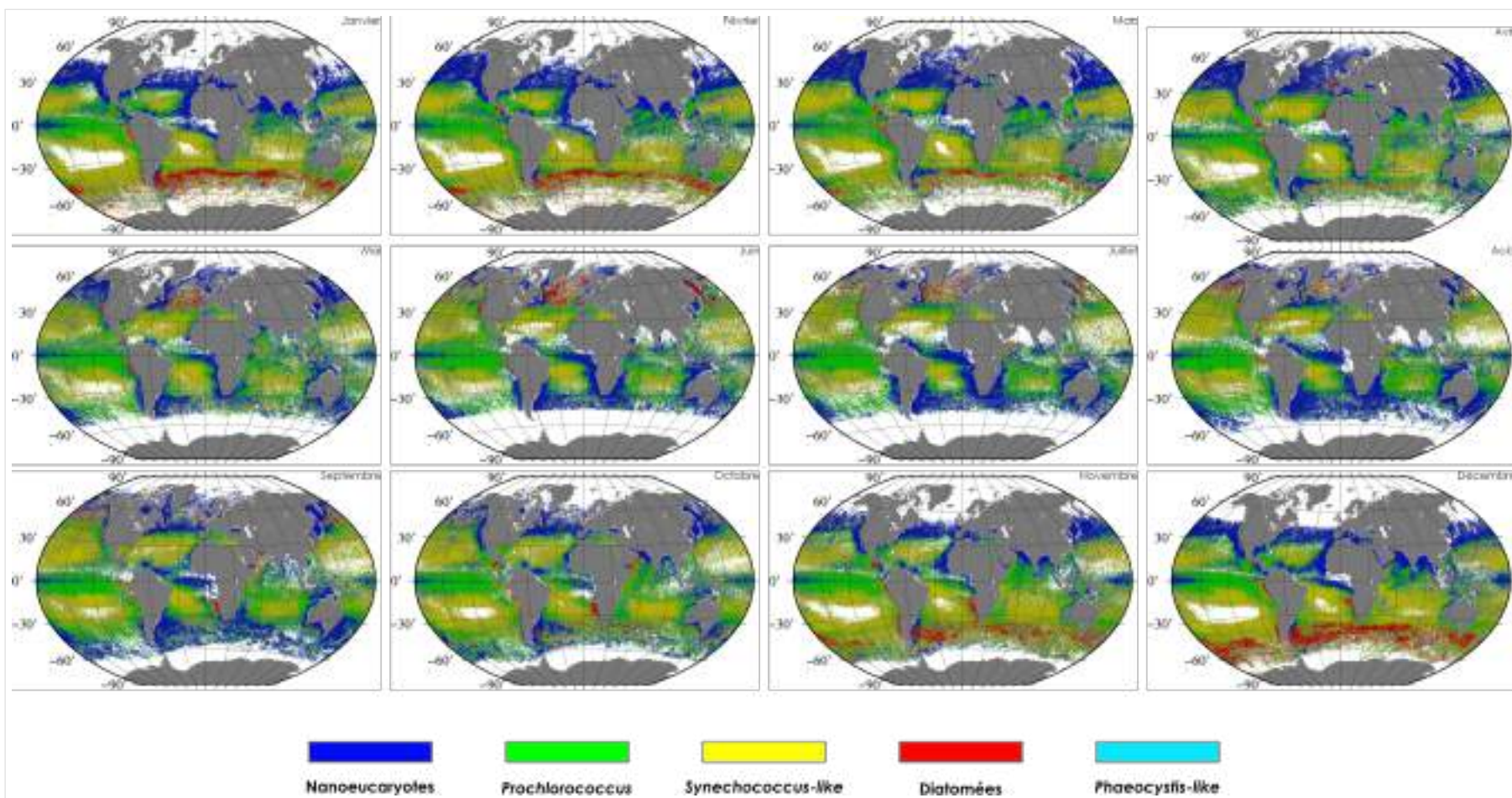


FIGURE 2.8 – Cartes de synthèses mensuelles de la distributions des cinq groupes dominants de PHYSAT (groupes de Alvain *et al.* (2005) projetés sur la carte-autoorganisatrice de Ben Mustapha *et al.* (2014)) issues du traitement et de la labellisation des anomalies de couleur de l’océan du capteur SeaWiFS entre 1998 et 2010.

### 2.2.2 À partir d'une expertise de terrain (Alvain *et al.*, 2008)

Après l'étude de Alvain *et al.* (2005), les auteurs de la méthode PHYSAT se sont étonnés du nombre de pixels "sans dominance" représentant des patches cohérents dans l'Océan Austral. En effet, au regard des structures spatiales observées pour les anomalies de luminance de l'Océan Austral, ainsi que de la récurrence de ces données dans le temps, ils ont fait l'hypothèse qu'il s'agissait d'un groupe phytoplanctonique particulier, ou au moins d'une structure ayant un sens écologique. Malheureusement, le manque de mesures *in-situ* dans cette zone ne permettait pas de caractériser précisément cette structure en suivant la même approche que Alvain *et al.* (2005) (i.e., une calibration et une validation indépendante des labels). Cependant, de nombreux auteurs ont reporté une efflorescence phytoplanctonique récurrente du **taxon** *Phaeocystis spp.* (Goffart *et al.*, 2000; Smith *et al.*, 2003; Croot *et al.*, 2004; Schoemann *et al.*, 2005). Ainsi, Alvain *et al.* (2008) ont fait l'hypothèse que les anomalies de luminance spécifiques identifiées dans l'Océan Austral, ayant des caractéristiques qui sont peu (voire pas) observées dans les autres océans, pouvaient être dues à ce taxon. En effet, *Phaeocystis spp.* est connue pour son impact optique particulier, qui est dû à la formation d'un mucus blanc excrété par les cellules en période de blooms pour les protéger du broutage. Alvain *et al.* (2008) ont donc défini des gammes de variations de valeurs des anomalies de luminance qui semblent correspondre à l'apparition du bloom de *Phaeocystis spp.* dans l'Océan Austral. La Figure 2.9 est un exemple de carte obtenue en analysant la répartition de ces anomalies particulières (couplé à l'analyse de Alvain *et al.* (2005)).

La labellisation de ce type de spectre pour *Phaeocystis spp.* dans l'Océan Austral n'a pas pu être validée de manière classique par manque de données *in-situ*. En revanche, lorsque les cartes de *Phaeocystis spp.* ont été établies par Alvain *et al.* (2008), le type de spectre "*Phaeocystis spp.-like*" apparaît également dans le Sud de la Mer du Nord, où des efflorescences de ce taxon sont en effet observées (Lancelot *et al.*, 1994; Rousseau *et al.*, 2002).

Dans l'océan Austral, entre Novembre et Janvier, une succession de diatomées et de *Phaeocystis spp.-like* dominants contribue à plus de 70% à la concentration en Chlorophylle-*a*, telle qu'estimée par les données de couleur de l'océan, avec une contribution plus importante des diatomées (50% de la concentration totale en Chlorophylle-*a*) (Alvain *et al.*, 2008).

L'ajout du groupe *Phaeocystis spp.-like* dans l'algorithme PHYSAT est un vrai avantage pour des premières études biogéochimiques si l'on considère la fonction de ce taxon, i.e., producteur de DMS, molécules dont les cartographies (e.g., Belviso *et al.* (2000)) coïncident avec la répartition de ce groupe dans PHYSAT. Cependant, nous notons, à travers cette approche relativement approximative (comparée à une calibration classique), qu'il est indispensable de disposer de bases de données d'observations sur le terrain concer-

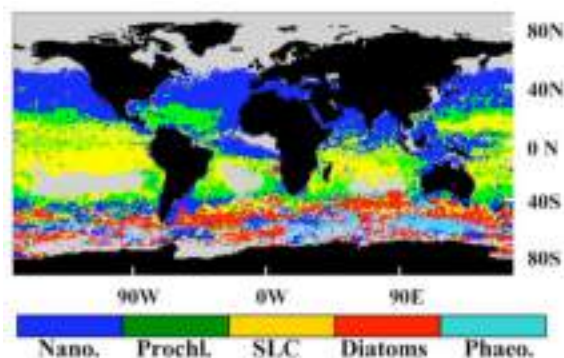


FIGURE 2.9 – Carte des groupes de phytoplancton dominants de Janvier 2002 issue du traitement des images du capteur de couleur SeaWiFS par la méthode PHYSAT Alvain *et al.* (2005), agrégée du groupe *Phaeocystis spp.* identifié par expertise. Extraite de Alvain *et al.* (2008).

nant le phytoplancton à large échelle. Notons également que l'étude d'un signal spécifique d'anomalie de luminance peut permettre, *a posteriori* d'avoir une idée des zones et des périodes où un échantillonnage *in-situ* serait intéressant.

### 2.2.3 À partir de données de cytométrie (Thyssen *et al.*, 2015)

La méthode PHYSAT a été dédiée à la détection des groupes dominants à l'échelle globale (Alvain *et al.*, 2005, 2008). Cependant, afin d'évaluer les limites de cette méthode, il était important d'évaluer le potentiel de PHYSAT pour des groupes phytoplanctoniques non dominants, et à l'échelle régionale (mais toujours dans les eaux du Cas I). L'étude de Thyssen *et al.* (2015) a été une étude préliminaire dans cette optique.

En effet, Thyssen *et al.* (2015) ont réalisé une première calibration empirique des anomalies de luminance PHYSAT à partir de données *in-situ* de **cytométrie en flux** (*Scanning Flow Cytometry* ou SFC). Pour cela, les mesures coïncidentes entre les données cytométriques de la campagne DYPHYMA (8-12 Mai 2011, sud-ouest de la Mer du Nord), et les données d'anomalies du capteur de couleur MODIS (résolution de 4km) ont été réalisées. Pour cette campagne, 15 mesures coïncidentes (avec les critères de qualité cités en Section 2.1.2) ont pu être trouvées entre les données *in-situ* issues de la cytométrie en flux et les données d'anomalies de luminance MODIS. Cette étude étant régionale, la table de référence de PHYSAT  $nLw_{ref}(\lambda, Chl-a)$  a été calculée spécifiquement pour cette zone, afin de mieux rendre compte des spécificités de la région. L'idée sous-jacente étant d'évaluer la disponibilité d'un signal associé à un groupe phytoplanctonique, sans être trop influencé par des caractéristiques régionales qui seront à étudier à l'avenir (Thyssen *et al.*, 2015). Les anomalies de luminance ont ensuite été calculées selon la procédure standard (Équation 2.1).

Deux types d'anomalies (formes et amplitudes) ont été détectées au niveau des mesures coïncidentes. Plus précisément, le premier type d'anomalie avait des valeurs supérieures à 1 (un signal plus grand que la moyenne) alors que le second type était inférieur à 1 (un signal plus petit que la moyenne).

La combinaison des mesures cytométriques et des observations PHYSAT a montré que le premier type d'anomalies (supérieur à la moyenne) pouvait être associé à une communauté de nano- et de micro-phytoplancton (des diatomées et des cryptophytes) et le second type d'anomalies (inférieur à la moyenne) était associé à du pico-phytoplancton. Ces deux communautés ont également été reliées à des environnements spécifiques en termes de salinité, température et profondeur de couche de mélange.

Cette étude montre qu'il est possible de calibrer empiriquement les anomalies de PHYSAT également au moyen de mesures *in-situ* de cytométrie. Ici, le but de l'étude n'était pas de réaliser une calibration exhaustive des données PHYSAT, mais d'observer si un signal spécifique existait pour les deux communautés phytoplanctoniques déterminées en analysant les données *in-situ*. Il a été montré qu'en effet, les signaux coïncidents sont différents pour les deux communautés. Ainsi, Thyssen *et al.* (2015) ont été les premiers à montrer que PHYSAT avait un potentiel de détection au-delà des cas de dominance, en associant des signaux spécifiques des *assemblages* phytoplanctoniques. Bien que cette étude n'ait pas fait l'objet d'une validation robuste (nombre de mesures coïncidentes trop faible), elle a été très importante pour la poursuite des investigations sur le potentiel de PHYSAT, en encourageant les auteurs à repousser les limites de la méthode en améliorant différentes composantes : traitement des données d'anomalies de luminance et analyses des mesures de terrain.

#### 2.2.4 État des lieux des données potentiellement utilisables pour la calibration de PHYSAT

Actuellement, il n'existe pas de restriction apparente dans l'utilisation des données *in-situ*. En effet, il est possible, comme nous l'avons vu dans les parties précédentes, d'analyser les données *in-situ* de manière à en extraire une information écologique pertinente (groupe dominant, assemblages, fonctions). Dans le cadre de l'évaluation des limitations de PHYSAT, il serait idéal d'établir des méthodes "clefs en main" pour chaque type de données qui concerne le phytoplancton (comptages cellulaires, cytométrie en flux, analyse pigmentaire, données ADNs, voire ARNs).

Cependant, avant d'en arriver à utiliser ce type de données, il conviendra de développer des outils adéquats et d'évaluer le potentiel d'utilisation de ces données dans le cadre d'une calibration ; afin de répondre à la question : existe-t-il des signaux d'anomalies de luminance spécifiques pouvant être associés de façon récurrente à des situations phytoplanctoniques au-delà de la dominance ?

## 2.3 Analyse théorique du lien empirique observé entre la présence d'un groupe phytoplanctonique dominant et d'une anomalie de luminance spécifique

La méthode PHYSAT a permis de montrer qu'il existe un lien empirique entre des anomalies de luminance de forme et d'amplitude spécifiques et la présence d'un groupe phytoplanctonique particulier (Alvain *et al.*, 2005, 2008; Thyssen *et al.*, 2015). Il y avait un intérêt particulier à comprendre pourquoi ce lien existait d'un point de vue "optique", quels paramètres et comment ces paramètres influençaient-ils la forme et l'amplitude des spectres d'anomalies de luminance en fonction de la longueur d'onde (Alvain *et al.*, 2012).

Afin de réaliser cette étude théorique, il a d'abord fallu définir les paramètres d'influence pour théoriser ces spécificités, c'est-à-dire les paramètres qui sont susceptibles d'avoir un impact sur l'anomalie de luminance (forme et amplitude). Puisque le signal moyen reçu par le satellite peut être exprimé en fonction des propriétés optiques inhérentes au milieu (*IOPs* pour *inherent optical properties*) (e.g., Morel et Gentili (1991)), ces propriétés optiques inhérentes ont un impact sur le signal détecté par PHYSAT (Alvain *et al.*, 2012).

Partant de ce constat, les IOPs considérées pour l'étude théorique ont été : la rétrodiffusion des particules (notée  $b_{bp}$  pour *particular backscattering*), l'absorption par les cellules phytoplanctoniques (notée  $a_{phy}$ ) et l'absorption par les matières dissoutes colorées (notées  $a_{CDOM}$  pour *Colored Dissolved Organic Matters*). Les valeurs de ces trois paramètres optiques dépendent de la nature du milieu, des constituants du milieu et de leurs structures (tailles, formes, agencement des constituants entre-eux). Dans les eaux du Cas I, définies en Introduction, ce sont en effet ces trois IOPs, dont les valeurs sont induites par la présence du phytoplancton (quantité et qualité), qui décrivent les variations du signal reçu par la satellite.

L'utilisation du programme *HydroLight* (Mobley, 1994) a permis de réaliser des simulations afin d'étudier la sensibilité des anomalies de luminance en fonction de différents sets de propriétés optiques inhérentes. Premièrement, une référence de luminance marine théorique ( $nLw_{ref, théorique}(\lambda, Chl-a)$ ) a été construite en considérant les relations moyennes entre les IOPs et la concentration en Chlorophylle-*a* pour différentes concentrations en Chlorophylle-*a* (de 0,02 à 3,0 mg.m<sup>-3</sup>, gamme d'étendue de valeur pour PHYSAT (Alvain *et al.*, 2005)), sur la base des travaux de Bricaud *et al.* (1998) (pour la relation  $Chl-a-a_{phy}$ ), Morel et Maritorena (2001); Gordon et Morel (2012); Loisel et Morel (1998); Morel *et al.* (2002); Loisel *et al.* (2006); Antoine *et al.* (2011) (pour la relation  $Chl-a-b_{bp}$ ) et Bricaud *et al.* (1981); Morel (2009); Siegel *et al.* (2002) (pour la relation  $Chl-a-a_{CDOM}$ ). La table de référence théorique ainsi calculée est assez proche de celle calculée pour le capteur de couleur de l'océan SeaWiFS dans l'étude de Alvain *et al.* (2005) (Alvain *et al.*, 2012).

La seconde étape de l'étude était donc de générer des valeurs de luminances théoriques "observées" qui tiennent compte de la variabilité naturelle des valeurs d'IOPs (Loisel *et al.*, 2010). Les valeurs moyennes d'IOPs, qui ont été définies dans la première étape de l'étude afin de construire la table de référence  $nLw_{\text{ref, théorique}}(\lambda, \text{Chl-}a)$ , ont été multipliées d'un facteur 0,5 à un facteur de 1,9 (incrément de 0,2), afin de calculer des luminances théoriques "observées" (*via HydroLight*) qui correspondent à des gammes de variations naturelles des IOPs correspondantes.

Alvain *et al.* (2012) disposaient donc d'un set de référence des luminances théoriques  $nLw_{\text{ref, théorique}}(\lambda, \text{Chl-}a)$  (valeurs d'IOPs moyennes), et d'un set de luminances théoriques observées (gammes de valeurs naturelles des IOPs). Grâce à ces deux sets théoriques, les anomalies de luminance théoriques ont pu être calculées, grâce à l'Équation 2.1 (page 29), pour différentes concentrations en Chlorophylle-*a*. En ayant l'intégralité de ces informations utilisées dans PHYSAT, simulées par le modèle *HydroLight*, les auteurs ont pu conclure sur l'impact de la variation des IOPs sur l'anomalie de luminance, et en expliquer les subtilités.

### Résultats majeurs de Alvain *et al.* (2012) ayant orienté les études menées dans cette thèse

Par cette étude, il a été démontré que l'absorption du phytoplancton  $a_{phy}$  avait une grande influence sur la forme du spectre théorique d'anomalie de luminance (Figure 2.10, gauche). En effet, un  $a_{phy}$  inférieur à la moyenne (moyenne définie dans l'étape 1 de l'étude) induit une forme de spectre théorique concave entre 443 et 510 nm avec une amplitude haute alors qu'une  $a_{phy}$  supérieure à la moyenne donne un spectre théorique d'anomalie de luminance convexe entre 443 et 510 nm et de faible amplitude. Cette observation est renforcée pour les fortes concentrations en Chlorophylle-*a* (le spectre est légèrement plus convexe ou concave pour une plus forte concentration en Chlorophylle-*a*), mais la variabilité des spectres de  $Ra(\lambda)$  est beaucoup plus marquée pour différentes valeurs de  $a_{phy}$  que pour différentes concentrations en Chlorophylle-*a*.

La variation de la rétrodiffusion induit quand-à elle une grande différence dans les amplitudes des spectres théoriques d'anomalies de luminance, et assez peu sur la forme du spectre (Figure 2.10, milieu). Pour une valeur de  $b_{bp}$  inférieure à la moyenne, le spectre de  $Ra(\lambda)$  théorique a une amplitude faible et est très légèrement décroissant entre 412 et 555nm; alors que pour une valeur de  $b_{bp}$  supérieure à la moyenne, l'amplitude est forte et le spectre est légèrement croissant entre 412 et 555nm.

L'absorption des matières dissoutes colorées  $a_{CDOM}$  induit une grande variabilité dans la forme et l'amplitude du  $Ra(\lambda)$  théorique, en particulier dans la partie bleue du spectre (412-490 nm) (Figure 2.10, droite). Cela se traduit par un spectre décroissant entre 412 et

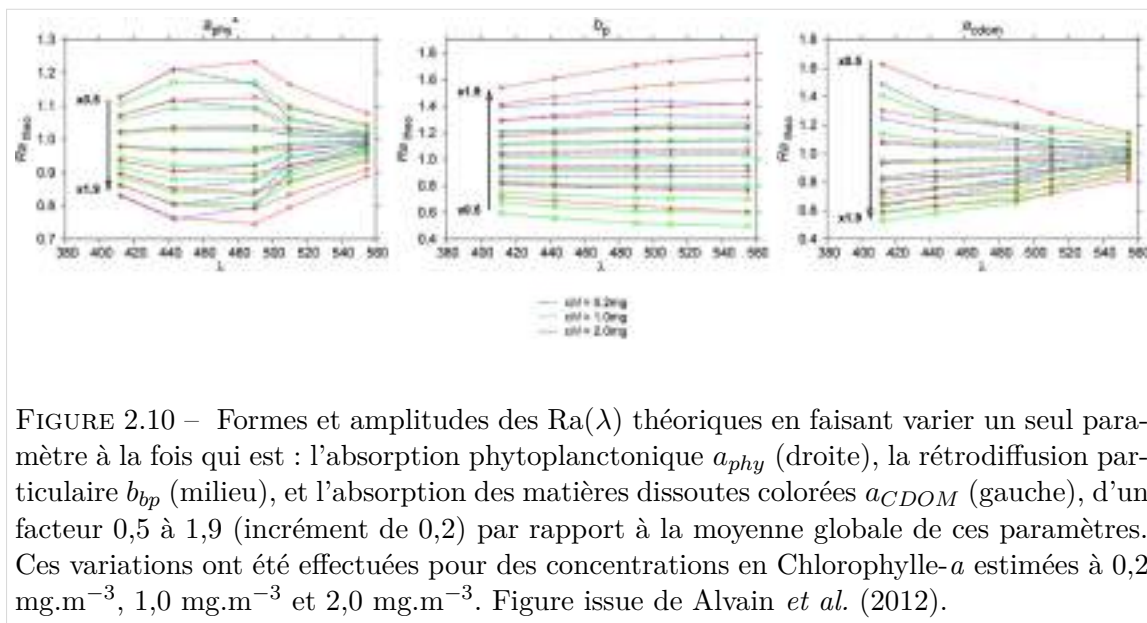


FIGURE 2.10 – Formes et amplitudes des  $Ra(\lambda)$  théoriques en faisant varier un seul paramètre à la fois qui est : l’absorption phytoplanctonique  $a_{phy}$  (droite), la rétrodiffusion particulaire  $b_{pp}$  (milieu), et l’absorption des matières dissoutes colorées  $a_{CDOM}$  (gauche), d’un facteur 0,5 à 1,9 (incrément de 0,2) par rapport à la moyenne globale de ces paramètres. Ces variations ont été effectuées pour des concentrations en Chlorophylle- $a$  estimées à 0,2  $mg \cdot m^{-3}$ , 1,0  $mg \cdot m^{-3}$  et 2,0  $mg \cdot m^{-3}$ . Figure issue de Alvain *et al.* (2012).

555 nm pour une valeur de  $a_{CDOM}$  inférieure à la moyenne et un spectre de  $Ra(\lambda)$  théorique croissant pour une valeur de  $a_{CDOM}$  supérieure à la moyenne.

En analysant une grande base de données *in-situ* de mesures d’IOPs couplés à des mesures pigmentaires (sur lesquelles il était possible de définir les groupes dominants *in-situ*), Alvain *et al.* (2012) ont défini des gammes de variations de ces paramètres bio-optiques pour les groupes phytoplanctoniques tels que détectés par PHYSAT. Cela a permis d’identifier des types de spectres de  $Ra(\lambda)$  théoriques, parmi les simulations effectuées, qui correspondent aux anomalies de luminance caractérisées dans PHYSAT (Figure 2.11). Les anomalies de luminance théoriques identifiées pour trois groupes phytoplanctoniques de PHYSAT sont relativement cohérentes avec les spectres d’anomalie de luminance observés, i.e., ils montrent le même ordre dans les amplitudes, bien que des différences soient observées pour la forme.

Alvain *et al.* (2012) soulignent que c’est la combinaison des valeurs de ces IOPs observées qui induit une forme de spectre d’anomalies de luminance observé. À l’heure actuelle, il n’est pas possible de connaître la proportion de variation du signal d’anomalie induite par telle ou telle IOP en milieu naturel. Cependant, cette étude a permis de mettre en évidence que les anomalies de luminance pouvaient être porteuses d’une information phytoplanctonique au-delà de la dominance, tant que la combinaison des valeurs des IOPs engendre une anomalie de luminance spécifique. Ces anomalies de luminance seront variables (formes et amplitudes) en fonction de la contribution des trois IOPs au signal total.

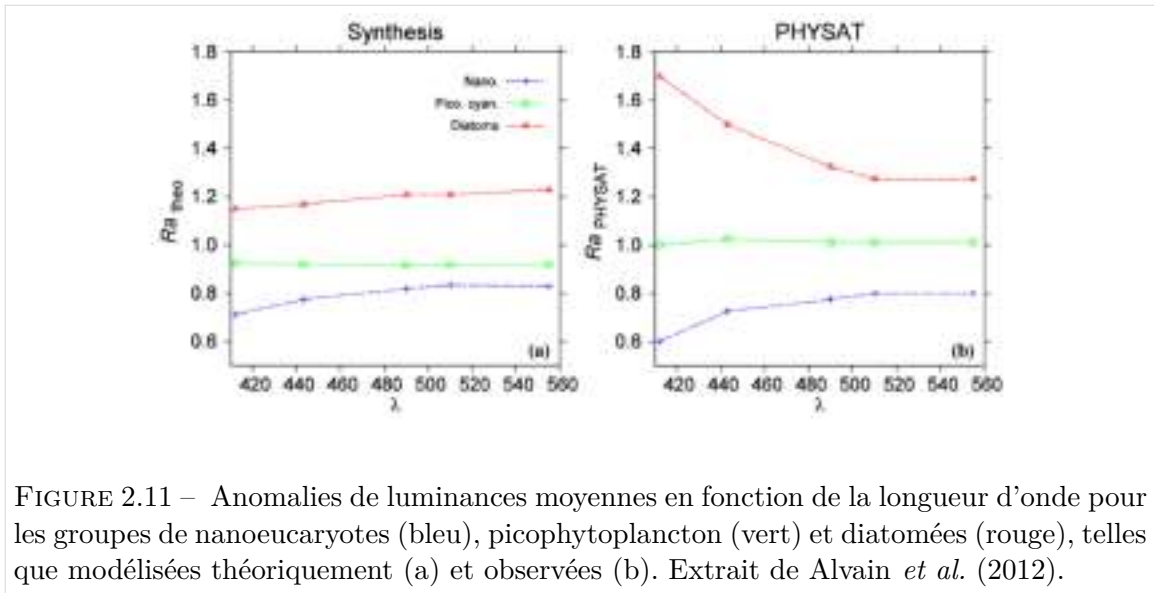


FIGURE 2.11 – Anomalies de luminances moyennes en fonction de la longueur d’onde pour les groupes de nanoeucaryotes (bleu), picophytoplancton (vert) et diatomées (rouge), telles que modélisées théoriquement (a) et observées (b). Extrait de Alvain *et al.* (2012).

## 2.4 Limites actuelles et questionnement sur le potentiel de développement de la méthode PHYSAT

Les études de Alvain *et al.* (2012) et Thyssen *et al.* (2015) ont été encourageantes pour développer davantage la méthode PHYSAT. En effet, en expliquant le lien théorique qui relie les anomalies de luminance à la présence d’un groupe phytoplanctonique, les auteurs ont suggéré qu’il était possible de détecter plus que des cas de dominance phytoplanctonique, en détectant par exemple des cas de composition mixte en phytoplancton, à partir du moment où le couplage des effets des trois IOPs mises en jeu induit un signal spécifique. L’étude de Thyssen *et al.* (2015) conforte cette hypothèse, puisque des signaux spécifiques ont été reliés à des compositions phytoplanctoniques mixtes. Cependant, cette dernière étude reste limitée pour parler de "preuve" de détection des cas mixtes si l’on considère le peu de données issues de la campagne *in-situ* (15 mesures coïncidentes).

La détection uniquement des cas de dominance par la méthode PHYSAT représente donc probablement une sous-utilisation du potentiel d’informations qui peut être déduit des anomalies de luminance, si l’on considère l’étude théorique (Alvain *et al.*, 2012) et l’étude menée à partir de la cytométrie en flux (Thyssen *et al.*, 2015). Ainsi, le principal objectif de cette thèse est de tester ce potentiel de détection des assemblages phytoplanctoniques. Pour cela, nous avons réalisé deux études qui répondent aux questions suivantes :



- ▶ Est-il possible, en se basant sur une analyse plus précise des données *in-situ* (ou tout au moins, plus nuancée), d'associer les anomalies de luminance PHYSAT à la présence de situations phytoplanctoniques plus diversifiées ? En d'autres mots, si l'on ne cherche pas à associer une "dominance" aux anomalies de luminance, mais que l'on essaie d'y associer une composition mixte, des relations satisfaisantes peuvent-elles être trouvées entre la présence d'un assemblage et un type spécifique d'anomalie de luminance ?

→ *Chapitre 4, Potentiel de détection des assemblages mixtes de groupes taxonomiques à l'échelle globale par la méthode PHYSAT.*

- ▶ Si l'information *in-situ* est de plus en plus détaillée, quel en sera l'impact sur les relations empiriques avec les classes de  $Ra(\lambda)$  déterminées grâce aux mesures coïncidentes ? Ou, en d'autres termes, une relation empirique est-elle décelable lorsque l'on considère des assemblages très précis (i.e., des assemblages de taxa) ?

→ *Chapitre 5, Potentiel de détection des assemblages d'espèce à l'échelle régionale par la méthode PHYSAT.*

Pour répondre à ces questions, la classification développée par Ben Mustapha (2013), qui permet d'analyser un grand nombre de types de spectres d'anomalie de luminance, est un avantage. En effet, au regard de l'analyse théorique de Alvain *et al.* (2012), avoir des classes de  $Ra(\lambda)$  dont les caractéristiques spectrales sont définies dans toute leur diversité (e.g., Figure 2.6) est une avancée pour une caractérisation plus précise des situations phytoplanctonique pouvant être rencontrées. Cependant, cette classification mathématique n'a pas permis jusqu'à présent une calibration empirique *via* des mesures coïncidentes *in-situ*. En effet, la formation, au cours de la création d'une carte auto-organisatrice, de 100 classes de types de spectres pose un problème lors de l'étape d'association de ces types de spectres à une composition phytoplanctonique (détaillé ensuite). À l'heure actuelle, ces 100 classes avaient été "labellisées" grâce à une projection de la distribution de la calibration de Alvain *et al.* (2008), pixel par pixel, sur la distribution des classes d'anomalies

de luminance<sup>8</sup>. Cette forme de labellisation n'est pas totalement satisfaisante, car elle équivaut à considérer des enveloppes du même type que celles définies intuitivement par Alvain *et al.* (2005) (Figure 2.4, page 34), dont le problème était la définition des seuils sur les  $Ra(\lambda)$  *a posteriori* pour chaque groupe.

Bien que cette classification aide à la compréhension de la distribution des types d'anomalies de luminance (e.g., Figure 2.7), le nombre de classes est potentiellement trop élevé, surtout si l'on considère la diversité phytoplanktonique caractérisée à l'heure actuelle. Le nombre de classes a été choisi volontairement important, dans le but de minimiser les regroupements d'anomalies de luminance statistiquement forcés (sur-apprentissage ou "*overfitting*"), et de laisser des possibilités d'extension de la méthode PHYSAT (distinguer un maximum de cas possible, sans restriction statistique *a priori*). Un problème se pose cependant au moment de la calibration, qui est dû au grand nombre de classes. En effet, avec l'augmentation du nombre de classes d'anomalies de luminance, il peut devenir difficile de trouver des associations robustes entre la présence des signaux *in-situ* et satellites. Les chances de trouver une association diminuent lorsque le nombre de classes augmente car l'information *in-situ* peut être répartie sur plusieurs classes de  $Ra(\lambda)$  censées lui correspondre (phénomène de *splitting*).

Avant de tester le potentiel de PHYSAT dans la détection des assemblages phytoplanktoniques, il faut donc résoudre ce problème de nombre de classes. Une méthodologie est proposée au Chapitre 3 afin de prendre en compte à la fois la diversité "spectrale" des  $Ra(\lambda)$ , mais aussi l'information que l'on veut en déduire *in fine*, c'est-à-dire, l'association avec des situations phytoplanktoniques spécifiques.

---

8. En pratique, les cartes journalières de classes d'anomalies de luminance et groupes dominants sont comparées. Si une classe d'anomalie de luminance est toujours associée aux pixels d'un groupe particulier, cette classe est labellisée par "projection".



# Utilisation des caractéristiques saisonnnières des anomalies de luminance comme outil de classification complémentaire

---

## Problématique de ce Chapitre

Comment améliorer les sorties de la carte auto-organisatrice pour perfectionner PHYSAT en vue des problématiques écologiques et simplifier le traitement des observations satellites ?

Quels sont les avantages et les limites de l'utilisation de la phénologie en tant qu'outil de regroupement des anomalies de luminance similaires ?

## Introduction

Les anomalies de luminance sont actuellement classées selon deux caractéristiques (amplitude et forme spectrale) au moyen d'une carte auto-organisatrice (une classification de type "réseau de neurones", aussi appelée "*Self-organising map*" ou "*SOM*", Kohonen (2013); Ben Mustapha *et al.* (2014)). Le travail effectué par Ben Mustapha *et al.* (2014) montre que les classes de  $Ra(\lambda)$  proches sur la carte auto-organisatrice (qui ont donc des caractéristiques spectrales proches) sont souvent associées à un même groupe phytoplanktonique (voir en exemple la Figure 3.1). Les signaux provenant de deux classes de  $Ra(\lambda)$  de la carte auto-organisatrice peuvent être très proches en termes de caractéristiques (forme et amplitude), alors que la carte auto-organisatrice indique qu'il existe statistiquement une différence significative dans ces caractéristiques.

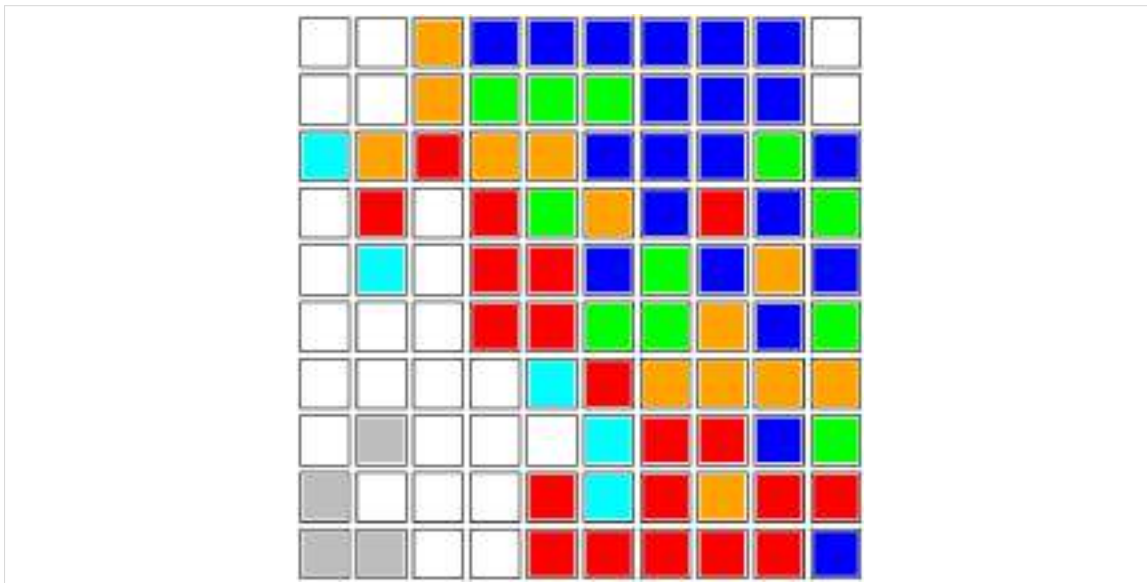


FIGURE 3.1 – Répartition des groupes phytoplanktoniques dominants définis par Alvain *et al.* (2005, 2008) sur la carte auto-organisatrice de Ben Mustapha *et al.* (2014) pour le capteur de couleur de l'océan SeaWiFS. Chaque case de cette Figure représente des classes de  $Ra(\lambda)$ , ayant des caractéristiques différentes en termes de formes et d'amplitudes. Ces classes sont associées empiriquement à la présence d'un groupe phytoplanktonique : les nanoeucaryotes (■), les *Prochlorococcus* (■), les *Synechococcus-like* (■), les diatomées (■), les *Phaeocystis-like* (■) et les coccolithes (■) selon la procédure définie par Alvain *et al.* (2005). Les classes de  $Ra(\lambda)$  proches (cellules de la carte adjacentes) sont la plupart du temps identifiées comme appartenant à des groupes identiques (Ben Mustapha *et al.*, 2014). Les classes blanches n'ont pas été associées empiriquement à la présence d'un groupe phytoplanktonique à ce jour (26% des classes de  $Ra(\lambda)$ ).

---

Cependant, comme expliqué dans le Chapitre 2 (Section 2.4), cette classification mathématique pose problème lors de l'étape de calibration de PHYSAT. La carte auto-organisatrice utilisée pour classer les anomalies de luminance a été configurée pour former mathématiquement 100 classes d'anomalies significativement différentes (Ben Mustapha *et al.*, 2014). Ce nombre a été choisi volontairement grand, afin de minimiser les regroupements d'anomalies de luminance statistiquement forcés (sur-apprentissage ou "*overfitting*"), et de laisser des possibilités d'extension de la méthode PHYSAT (distinguer un maximum de cas possible, sans restriction statistique *a priori*). En revanche, plus le nombre de classes de  $Ra(\lambda)$  est grand, plus il peut devenir complexe de trouver empiriquement des associations. En effet, les chances de trouver une association diminuent lorsque le nombre de classes augmente car l'information *in-situ*, associée à une même situation phytoplanktonique, peut être répartie sur plusieurs classes de  $Ra(\lambda)$  censées lui correspondre. Il était donc nécessaire de rechercher une solution permettant de garder un nombre de classes de  $Ra(\lambda)$  avec une grande diversité d'amplitudes et de formes de spectres, mais aussi un nombre de classes permettant de représenter au mieux les enjeux écologiques.

Ainsi, dans le but de conserver un nombre de classes de  $Ra(\lambda)$  représentatif à la fois de la diversité des caractéristiques spectrales des  $Ra(\lambda)$  mais aussi de la diversité phytoplanktonique, deux solutions ont été envisagées :

- (1) grouper les sorties de la carte auto-organisatrice en se servant des caractéristiques des  $Ra(\lambda)$  référents de chaque classe, c'est-à-dire former des super-classes à partir des mêmes caractéristiques (formes et amplitudes) que celles utilisées pour former la carte auto-organisatrice. C'est une technique courante dans l'utilisation des classifications de type "réseau de neurones" (e.g., Cottrell *et al.* (2003))
- (2) grouper les sorties de la carte auto-organisatrice grâce à une nouvelle caractéristique, non prise en compte pour former le réseau de neurone et indépendante des paramètres étudiés jusqu'ici, mais en lien avec l'objet de notre étude, le phytoplankton.

La première solution, consistant à former des "super-classes de  $Ra(\lambda)$ ", reviendrait à diminuer le nombre de classes de la carte auto-organisatrice *a posteriori*, et donc de réaliser une classification forcée à partir des mêmes caractéristiques. Bien que l'utilisation de super-classes formées *a posteriori* d'une classification de type carte auto-organisatrice soit conseillée pour de nombreuses applications (e.g., Cottrell *et al.* (2003)), le fait de classer les vecteurs référents de la carte auto-organisatrice peut poser problème. En effet, il a été observé que les super-classes ne regroupent que des classes contiguës (proches sur la carte auto-organisatrice), ce qui s'explique par la propriété de respect de la topologie de l'algorithme de Kohonen, puisqu'une notion de voisinage est définie entre les classes (Kohonen, 2013). La Figure 3.1 montre que les classes adjacentes sont en effet le plus sou-

vent associées à des groupes phytoplanctoniques similaires. En revanche, cette observation n'est pas toujours vraie pour des classes dont les caractéristiques des  $Ra(\lambda)$  référents sont proches, ce qui a d'ailleurs été conjecturé par l'étude théorique (Alvain *et al.*, 2012) : deux types phytoplanctoniques différents pourraient être associés avec le même type d'anomalie de luminance si la combinaison des facteurs optiques qui induisent cette anomalie engendre la même forme et amplitude. Bien que cette hypothèse pourrait théoriquement représenter des cas exceptionnels (Alvain *et al.*, 2012), nous avons souhaité garder la possibilité de les prendre en compte dans cette étude.

La seconde approche est alors apparue comme étant la meilleure solution, puisque l'on cherche à se détacher des caractéristiques classiques des anomalies de luminance calculées (i.e., amplitude et forme spectrales). La méthode employée permet de considérer une caractéristique qui soit liée à la classe de  $Ra(\lambda)$ , et non pas à l'anomalie de luminance elle-même (ou à l'anomalie de luminance référente de la classe de  $Ra(\lambda)$ ). L'information portée par chaque classe de  $Ra(\lambda)$  est notamment spatiale, ou temporelle, au-delà de la forme et de l'amplitude du spectre d'anomalie de luminance ; c'est-à-dire qu'il est possible de considérer son apparition dans le temps et dans l'espace. Pour regrouper les classes de  $Ra(\lambda)$ , nous avons choisi d'utiliser uniquement les informations temporelles, afin de conserver un outil de "validation" *a posteriori* : l'information spatiale<sup>1</sup>.

Dans la suite de ce Chapitre, nous allons donc définir l'étude des variations temporelles des classes de  $Ra(\lambda)$ , i.e., leur **phénologie**, puis la classification des classes de  $Ra(\lambda)$  en fonction de ces caractéristiques temporelles, et enfin nous étudierons une application qui montre le potentiel de l'utilisation de la phénologie pour le regroupement des classes de  $Ra(\lambda)$ .

En utilisant cette caractéristique, nous avons fait l'hypothèse que deux classes de  $Ra(\lambda)$ , provenant de la carte auto-organisatrice, portant une information temporelle similaire, et dont les caractéristiques des spectres sont également proches, devaient être deux classes d'anomalies qui avaient été considérées différentes "à tort"<sup>2</sup> par l'algorithme de regroupement de la carte auto-organisatrice. Alors, ces classes de spectres devraient avoir une répartition spatiale similaire également. Dans la suite du manuscrit, ces classes d'anomalies de la carte auto-organisatrice dont la distribution temporelle est similaire sont appelées classes de  $Ra(\lambda)$  "phénologie". L'hypothèse de ces classes "phénologie" est déduite du fait qu'il existe également des classes "non jumelles", c'est-à-dire des classes qui ont un spectre de  $Ra(\lambda)$  proche, mais pour lesquelles ni la fréquence d'apparition, ni la répartition spatiale n'est similaire. C'est le cas par exemple des classes *som* n°47 et 48. Ces deux classes ont été associées respectivement à des nanoeucaryotes et des *Synechococcus spp.-like* dominants lors d'une précédente calibration (Alvain *et al.*, 2008), les deux classes sont adjacentes sur

---

1. L'utilisation de l'information temporelle pour la classification demande moins de temps de calcul.

2. Évidemment, il existait des raisons statistiques valables à la séparation de ces pixels lors de l'étape de classification mathématique par la carte auto-organisatrice.

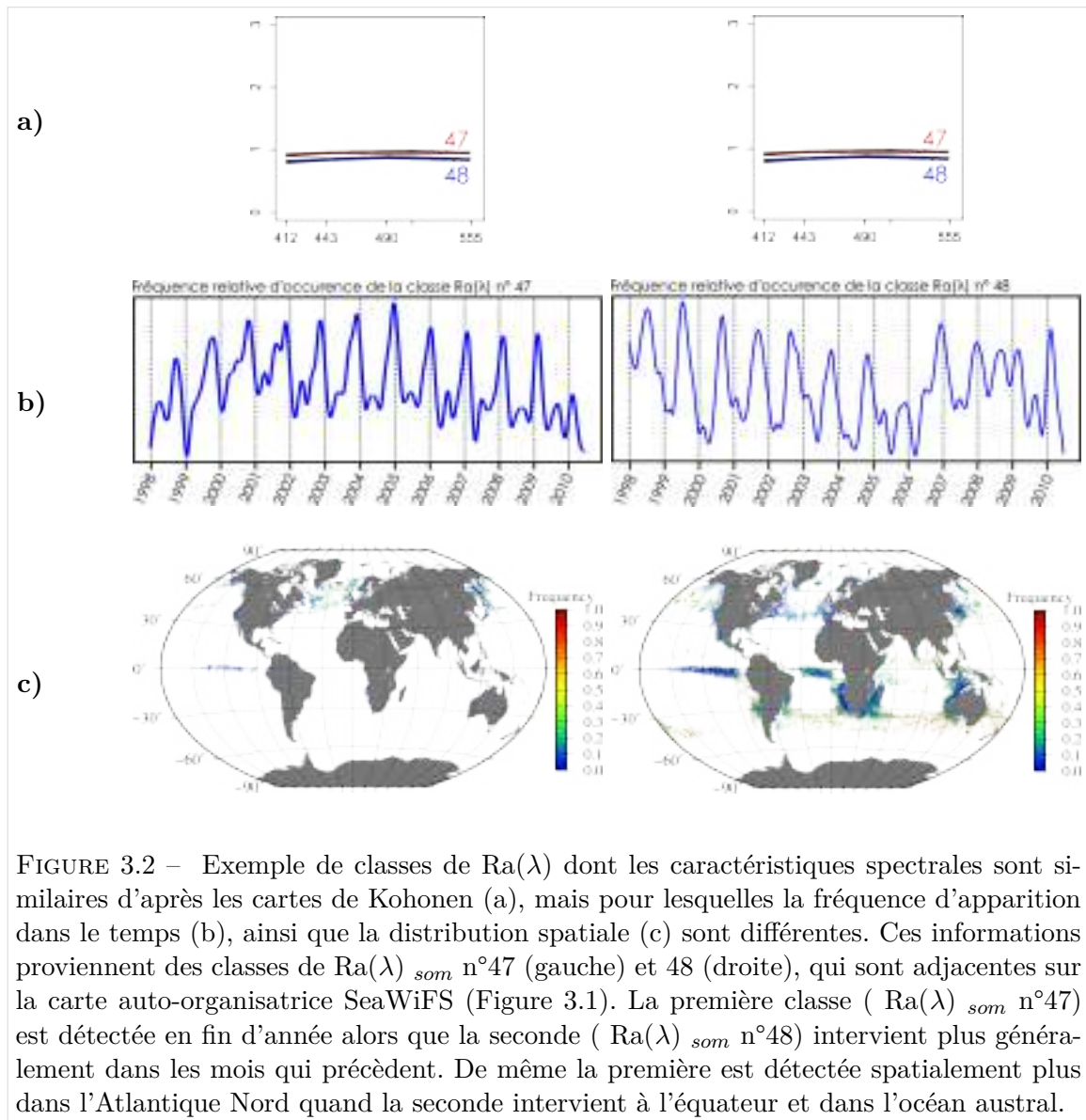


FIGURE 3.2 – Exemple de classes de  $Ra(\lambda)$  dont les caractéristiques spectrales sont similaires d’après les cartes de Kohonen (a), mais pour lesquelles la fréquence d’apparition dans le temps (b), ainsi que la distribution spatiale (c) sont différentes. Ces informations proviennent des classes de  $Ra(\lambda)_{som}$  n°47 (gauche) et 48 (droite), qui sont adjacentes sur la carte auto-organisatrice SeaWiFS (Figure 3.1). La première classe ( $Ra(\lambda)_{som}$  n°47) est détectée en fin d’année alors que la seconde ( $Ra(\lambda)_{som}$  n°48) intervient plus généralement dans les mois qui précèdent. De même la première est détectée spatialement plus dans l’Atlantique Nord quand la seconde intervient à l’équateur et dans l’océan austral.

la carte auto-organisatrice (Figure 3.1), et leurs spectres sont semblables (Figure 3.2, a)). En revanche, les fréquences d’apparitions et la distribution spatiale de ces deux classes de  $Ra(\lambda)$  sont tout à fait dissemblables (Figure 3.2, b) et c)).

### 3.1 Définition et calcul de la phénologie des anomalies de luminance

Afin d’étudier les caractéristiques dites "temporelles" des classes de  $Ra(\lambda)$ , nous avons utilisé des outils développés en *phénologie* (science des phénomènes temporels), qui sont utilisés dans de nombreuses applications, toutes sciences confondues e.g., Schwartz (2003), ou dans le domaine de l’océanographie e.g., Platt *et al.* (2009).



### 3.1.1 Définition de la phénologie et de ses indicateurs

La phénologie est l'étude de la temporalité des événements ayant une cyclicité (Schwartz, 2003). Considérant cette définition, les outils de la phénologie peuvent être appliqués à un grand nombre de domaines (liste non exhaustive dans Schwartz (2003)). Pour décrire l'aspect temporel d'un événement, les métriques phénologiques sont classiquement utilisées (Figure 3.3) :

- **L'initiation** : c'est le moment où débute l'événement biologique
- **Le moment du maximum (et sa valeur)** : c'est le moment où la variable mesurée atteint son maximum sur un cycle
- **La fin** : c'est le moment où l'événement est terminé (la variable revient à un niveau basal, ou un niveau de bruit)
- **La durée** : c'est le temps écoulé entre l'initiation et la fin de l'événement
- **La durée d'expansion et la durée de déclin** : c'est le temps écoulé entre l'initiation et le maximum et entre le maximum et la fin de l'événement, respectivement.

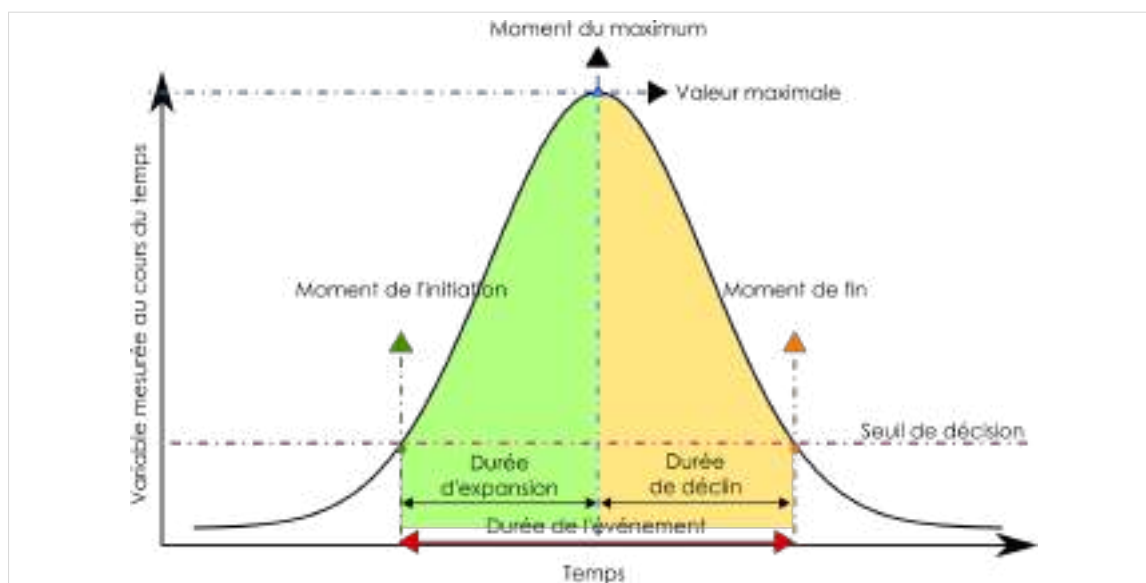


FIGURE 3.3 – Indicateurs décrivant une dynamique saisonnière (adapté de Platt et Sathyendranath (2008)). Une fonction gaussienne a été utilisée pour représenter une dynamique théorique d'un événement biologique. Le seuil correspond à un seuil de décision pour déterminer le début et la fin de la dynamique (voir au dessus). Veuillez noter qu'un événement naturel ne présente que rarement une structure temporelle gaussienne, et qu'il existe souvent une asymétrie entre la durée d'expansion et de déclin.

La variable étudiée peut être de différente nature, tant qu'elle reflète la saisonnalité de l'événement étudié : effectifs (ou abondance), effectifs relatifs, fréquence (ou occurrence), fréquence relative (Schwartz, 2003).

Le moment du maximum et sa valeur sont les seules métriques de la phénologie directement extraites des données temporelles brutes. Les autres indices sont estimés grâce à un seuil statistique (Figure 3.3). Ce seuil permet de définir un niveau de base (ou "bruit"), ne faisant pas partie du pic cyclique.

Il existe de nombreux types de seuils statistiques permettant d'estimer les métriques de la phénologie. La médiane est classiquement utilisée (e.g., Racault *et al.* (2012), Siegel *et al.* (2002)); le seuil le plus utilisé étant celui de la médiane +5% (Siegel *et al.*, 2002); mais d'autres seuils comme la moyenne ou le maximum peuvent être utilisés (incrémentés ou non). Le choix du seuil adéquat peut être solutionné par différentes approches mathématiques (e.g., méthode du point d'inflexion) et/ou statistiques (e.g., approche probabiliste).

### 3.1.2 L'information temporelle portée par les classes de $Ra(\lambda)$

L'utilisation des classes de  $Ra(\lambda)$  présente l'avantage de pouvoir être associée au calcul de paramètres tels que la fréquence relative d'apparition d'une classe de  $Ra(\lambda)$  dans le temps.

De manière classique, la fréquence de détection d'une classe de  $Ra(\lambda)$  peut être calculée pour un pixel donné ou moyennée pour une zone géographique plus large (régionale, voire globale). De même, la fréquence peut être moyennée pour une période de 8 jours à 1 mois ou plus selon la problématique scientifique et la zone à étudier (lié à la couverture nuageuse par exemple).

La fréquence est calculée selon l'Équation 3.1. Deux exemples de séries temporelles de fréquences calculées sont présentés dans la Figure 3.4.

$$F_x(Ra(\lambda))_{i,j} = \frac{O_x(Ra(\lambda))_{i,j}}{\Sigma(O_x(Ra(\lambda))_{.,j})} \quad (3.1)$$

Où  $F_x(Ra(\lambda))_{i,j}$  est la fréquence de la classe de  $Ra(\lambda)_i$  calculée sur la zone  $j$  en tenant compte d'une fenêtre temporelle de  $x$  jours (par exemple fréquence à 15 jours);  $O_x(Ra(\lambda))_{i,j}$  est l'occurrence de la classe de  $Ra(\lambda)_i$  détectée sur la zone  $j$  en tenant compte d'une fenêtre temporelle de  $x$  jours; et  $\Sigma(O_x(Ra(\lambda))_{.,j})$  est la somme des occurrences de toutes les classes de  $Ra(\lambda)$  détectées sur la zone  $j$  en tenant compte d'une fenêtre temporelle de  $x$  jours.

Les variations observées dans les fréquences de détection des signaux d'anomalies de luminance semblent cohérentes avec des événements biologiques (Figure 3.4). Une cyclicité est

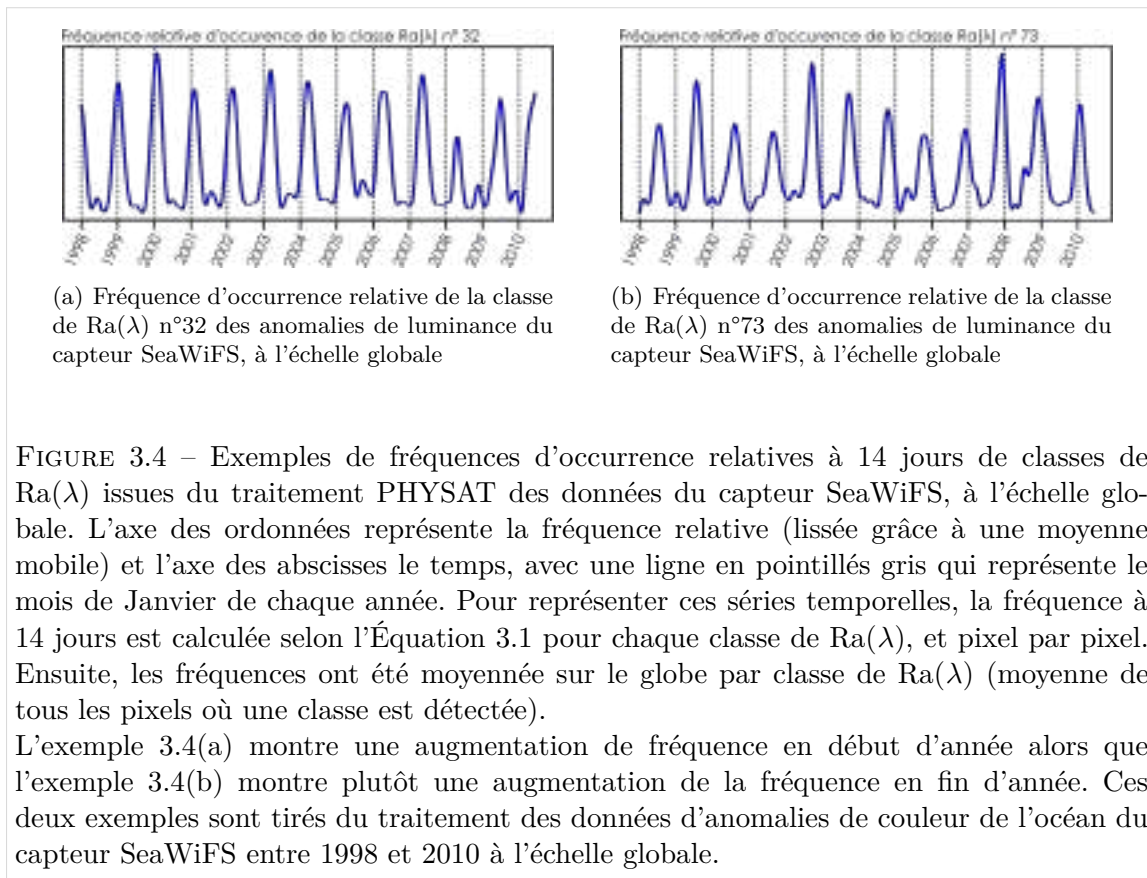


FIGURE 3.4 – Exemples de fréquences d'occurrence relatives à 14 jours de classes de  $Ra(\lambda)$  issues du traitement PHYSAT des données du capteur SeaWiFS, à l'échelle globale. L'axe des ordonnées représente la fréquence relative (lissée grâce à une moyenne mobile) et l'axe des abscisses le temps, avec une ligne en pointillés gris qui représente le mois de Janvier de chaque année. Pour représenter ces séries temporelles, la fréquence à 14 jours est calculée selon l'Équation 3.1 pour chaque classe de  $Ra(\lambda)$ , et pixel par pixel. Ensuite, les fréquences ont été moyennée sur le globe par classe de  $Ra(\lambda)$  (moyenne de tous les pixels où une classe est détectée).

L'exemple 3.4(a) montre une augmentation de fréquence en début d'année alors que l'exemple 3.4(b) montre plutôt une augmentation de la fréquence en fin d'année. Ces deux exemples sont tirés du traitement des données d'anomalies de couleur de l'océan du capteur SeaWiFS entre 1998 et 2010 à l'échelle globale.

observée, ainsi qu'une variabilité inter-annuelle. L'ensemble des fréquences de détection des classes de  $Ra(\lambda)$  pour le capteur de couleur de l'océan SeaWiFS est présenté en Annexe 4 (page 265). Il est important de garder à l'esprit que les valeurs de fréquences, bien que très utiles dans la description de la cyclicité des  $Ra(\lambda)$ , restent subjectives. En effet, le calcul de la valeur de la fréquence dépend de la représentativité de la classe de  $Ra(\lambda)$  (spatialement et temporellement) par rapport aux autres classes de  $Ra(\lambda)$  déterminées par la carte auto-organisatrice. Comme nous l'avons indiqué plus haut, la classification des  $Ra(\lambda)$  réalisées grâce à une carte auto-organisatrice peut engendrer une séparation statistique stricte, qui risque, si l'on s'en tient à cela, d'attribuer aux 100 classes prédéfinies un nombre de pixels variables. Ainsi, une classe dont l'effectif en pixels est faible aura une valeur de fréquence faible par rapport à une classe dont l'effectif en pixels est élevé. Pour pallier ceci, l'utilisation de la phénologie est idéale puisque :

- Le seuil statistique qui aide à l'estimation des métriques de la phénologie est calculé spécifiquement pour chaque classe de  $Ra(\lambda)$ . Les seuils ne sont pas comparés entre eux et des fortes différences observées dans la valeur des seuils n'influencent pas le calcul des métriques de la phénologie de chaque classe de  $Ra(\lambda)$
- La majorité des métriques qui décrivent la phénologie sont des indices temporels

("moment" d'initiation, de maximum, de fin, durée).

*Nota bene* La valeur du maximum est une métrique classique dans l'étude de la phénologie (Platt et Sathyendranath, 2008). En revanche, cette métrique ne sera pas utilisée dans cette étude en raison de son absence d'objectivité pour des valeurs de fréquence (explication ci-dessus). En effet, l'utilisation de la valeur de maximum comme variable dans une classification pourrait engendrer un regroupement forcé des classes ayant les plus fortes fréquences si cette variable a un poids important, au détriment de regrouper les classes par la forme de leurs cycles temporels.

Cette version de l'utilisation des métriques de la phénologie (toutes, hormis la valeur du maximum), permet de décrire la forme absolue (ou normalisée) de la série temporelle.

### 3.1.3 Choix du seuil statistique pour la détermination des indices de la phénologie des classes de $Ra(\lambda)$

Une série de seuils est définie dans la littérature (Ji *et al.*, 2010) : la médiane, la moyenne et le maximum. Les seuils de types "maximum", pouvant être incrémentés (max -  $x\%$ ), sont rarement utilisés mais ont montré leur efficacité pour les séries temporelles extrêmes : forme en pic très "aplati" avec une variation de valeur faible entre l'initiation et le maximum (Delbart *et al.*, 2005) ou, au contraire, pour une saisonnalité très marquée avec une valeur de maximum très élevée (Platt *et al.*, 2009).

Dans cette étude, un éventail de seuils a été analysé afin de trouver le seuil le plus adéquat à la description de 100 séries temporelles de classes de  $Ra(\lambda)$  :

- Médiane +0 à +20% (incrément de 5%, soit 5 seuils de type "médiane")
- Moyenne +0 à +20% (incrément de 5%, soit 5 seuils de type "moyenne")
- Maximum -20 à -60% (incrément de 5%, soit 9 seuils de type "maximum")

Cette étape consiste à trouver le seuil qui décrit le mieux la série temporelle de fréquence de détection parmi les 19 seuils cités ci-dessus, et cela pour l'ensemble des 100 classes de  $Ra(\lambda)$ . Un seuil idéal doit décrire au mieux les variations de la fréquence, notamment en ce qui concerne le point d'initiation (Platt *et al.*, 2003; Wiltshire *et al.*, 2008). Le moment de l'initiation doit refléter une augmentation importante de la fréquence de détection de la classe de  $Ra(\lambda)$ . Ainsi, pour chacun des 19 seuils, le moment d'initiation a été calculé pour les 100 classes de  $Ra(\lambda)$ , et pour chaque année (19 seuils, 100 classes de  $Ra(\lambda)$  et 12 années, soit 22800 moments d'initiation estimés). Ensuite, il faut déterminer lequel de ces seuils décrit le mieux l'augmentation de fréquence de détection. Ce seuil doit donc décrire la plus forte vitesse de croissance de la fréquence, ce qui a été déterminé par la méthode

du point d'inflexion, également utilisée par Wiltshire *et al.* (2008). Dans cette étude, cette vitesse d'expansion<sup>3</sup> a été estimée comme la variation de la pente de la tangente entre le point précédant l'initiation et l'initiation. Plus cette pente de tangente est grande, plus la vitesse d'expansion est rapide, ce qui est la définition de l'initiation d'un événement biologique.

En utilisant le point d'inflexion, les seuils de type "médiane" sont plus adéquats dans 100% des cas testés par rapport aux seuils de type "moyenne" et "maximum". Il n'existe pas de différence entre les seuils "médiane +5%" et "médiane +10%", ceux-ci étant plus adéquats que les seuils médiane +15 à +20%. En effet, les variations de la pente de la tangente entre l'initiation et le point précédent étaient plus importantes pour les seuils "médiane +5 et +10%" (test Z unilatéral,  $\alpha = 0,05$ ). Afin d'être en adéquation avec un grand nombre d'études, le seuil "médiane +5%" a été choisi (Ji *et al.*, 2010). Ce seuil, étant également le plus utilisé, a montré son efficacité dans l'étude des fréquences d'apparition des anomalies de luminance, et dans la description de toutes sortes de séries temporelles (faibles à fortes amplitudes, cycles plus ou moins complexes (uni- ou bi-modal)).

#### **Résumé : seuil pour l'estimation des métriques de la phénologie**

Le seuil qui décrit le mieux les séries temporelles de fréquence de détection des classes d'anomalies de luminance est le seuil "**médiane +5%**". Ceci a été testé par la méthode du point d'inflexion (Wiltshire *et al.*, 2008). Ce seuil permet de décrire une grande diversité de types de séries temporelles (uni-modales, bi-modales, de faibles et fortes fréquences, cycliques ou plus complexes), et a donc été retenu pour la suite de notre étude.

Lorsque le seuil optimal a été déterminé, il convient alors de calculer les indices qui décrivent les séries temporelles pour chaque expansion de fréquence de chaque classe de  $Ra(\lambda)$  : moment de l'initiation (avec le seuil choisi), moment du maximum, moment de fin et durée d'expansion et de déclin du signal de fréquence de détection. Une méthode flexible a dû être développée afin de prendre en compte les spécificités observées sur le signal de fréquence.

---

3. En phénologie du phytoplancton, le terme de "vitesse de croissance du **bloom**" est utilisé. Ici, nous avons préféré utiliser le terme "vitesse de croissance de l'expansion de fréquence" car, bien que les classes de  $Ra(\lambda)$  soient liées à la présence du phytoplancton, ce lien n'est pas encore déterminé à ce stade de l'étude.

### 3.1.4 En pratique : développement d'un algorithme de calcul des métriques de la phénologie de manière flexible

Les méthodes phénologiques existantes (dont une liste succincte peut être trouvée dans Ji *et al.* (2010)) sont très performantes lorsque les séries temporelles sont unimodales, avec peu de valeurs extrêmes, et dont la cyclicité est préservée d'une année à l'autre. Pour les autres cas, il est nécessaire d'avoir une approche plus flexible avec moins d'hypothèses sur la distribution de la série temporelle (Ji *et al.*, 2010).

La méthode classique permet de considérer uniquement le pic principal (ayant le plus grand maximum), quel que soit le mode annuel de la série temporelle (unimodal, bimodal, trimodal). Pour répondre aux spécificités observées sur les fréquences des classes de  $Ra(\lambda)$ , nous avons eu besoin de développer une méthode plus flexible. En effet, il peut par exemple n'exister aucune expansion, ou une voire deux expansions par année pour une classe de  $Ra(\lambda)$  (voir Table 3.1), même si la majorité des classes de  $Ra(\lambda)$  présentent un seul pic par année. De plus, les fréquences de détection des anomalies de luminance présentent un signal relativement bruité. Les outils mathématiques classiques (e.g., dérivée en un point) sont donc difficilement utilisables sans lisser les séries temporelles au préalable. La méthode développée ici permet de prendre en compte plusieurs pics annuels grâce à une succession de critères (algorithme "*FindPheno*" programmé en R (Annexe 5, page 277) sur la base d'un programme "*FindPeak*"<sup>4</sup> (*non publié à ce jour*).) :

- Il ne peut pas exister plus de trois maxima dans une année. La valeur "trois maxima" est choisie car il existe des maxima (par exemple pour les séries temporelles hiver/été) qui peuvent être au début et à la fin du calendrier annuel. Une anomalie peut présenter un maximum en Janvier, puis en Juin, et un troisième maximum peut être détecté en Décembre (celui-ci correspondrait à un maximum de fréquence de l'année suivante). Dans ce cas, on peut observer trois maxima dans une même année calendaire mais en réalité seules deux périodes d'augmentation de la fréquence existent pour une année : en hiver et en été.
- Le maximum existe si la fréquence est supérieure à la médiane +5%, et qu'il existe au moins deux points précédents et deux points suivants dont la fréquence est inférieure à cette fréquence maximale.
- Le moment de l'initiation est déterminé de la même manière que dans la méthode classique : l'initiation débute lorsque la fréquence dépasse le seuil statistique après un minimum. L'initiation ne peut exister qu'entre un minimum et un maximum.
- Le moment de fin est déterminé de la même manière que dans la méthode classique :

---

4. Algorithme développé par Timothée Poisot permettant de trouver les minima et maxima locaux, disponible à l'url <https://rtricks.wordpress.com/2009/05/03/an-algorithm-to-find-local-extrema-in-a-vector/>

la fin est annoncée lorsque la fréquence redescend sous le seuil statistique, après un maximum. La fin de l'expansion ne peut exister qu'entre un maximum et un minimum.

- Un moment de fin ne peut pas être égal à un moment d'initiation (et *vice-versa*). Cette condition permet d'éviter, lorsque la fréquence est fortement bruitée (e.g., soubresaut brutal) que ce pic soudain soit considéré comme une réelle augmentation de fréquence.

TABLE 3.1 – Nombre d'expansion de la fréquence par année (0 à 3) enregistrées pour les 100 classes de  $Ra(\lambda)$  du capteur SeaWiFS sur 14 années (1997-2010). La moyenne du nombre d'expansion par années, toutes classes de  $Ra(\lambda)$  confondues, est de 1,2. Notez que le nombre d'expansion "3" correspond souvent à un signal situé en début ou fin d'année, qui peut donc être observé deux fois sur une année en fonction de sa temporalité : un maximum peut être détecté en janvier, puis en décembre (correspondant en réalité à une expansion de l'année suivante).

Nombre d'expansion du signal de fréquence	Occurrence pour SeaWiFS
0	194
1	773
2	383
3	50

Ces caractéristiques sont calculées sur les données de fréquences moyennées à 14 jours et non lissées (voir l'exemple ci-dessous). La résolution temporelle des indices de la phénologie sera donc de 14 jours dans la suite de cette étude, ce qui signifie, selon les critères définis ci-dessus qu'un pic sera considéré si l'expansion du signal de fréquence dure au moins un mois, ce qui est écologiquement représentatif pour prendre en compte un maximum de cas possibles, puisque la durée moyenne d'une efflorescence est de trois mois (e.g., Racault *et al.* (2012)).

## 3.2 Utilisation des métriques de la phénologie des anomalies de luminances comme nouvelles caractéristiques des ces spectres pour une classification

Après avoir calculé les indices de la phénologie pour chaque classe de  $Ra(\lambda)$ , ils ont été utilisés comme caractéristiques multivariées pour réaliser une classification (seconde étape de la classification des anomalies, après la carte auto-organisatrice). Pour cela, les variables suivantes ont été utilisées comme caractères observés : moment de l'initiation, moment du maximum, moment de la fin, durée d'augmentation et durée de déclin du signal. Ces cinq indices ont été calculés chaque année et pour chaque classe de  $Ra(\lambda)$  sur le pic annuel principal et secondaire. Par exemple, pour les données SeaWiFS, cela représente 140 variables (14 années \* 5 indices \* 2 pics).

Pour réaliser la classification des classes de  $Ra(\lambda)$  en fonction de ces indices de la phénologie, une matrice de distance a été calculée, puis une classification ascendante hiérarchique a été réalisée. Les *corrélations cophénétiqes* ont montré qu'une distance euclidienne et un algorithme de regroupement moyen était la combinaison qui respectait le mieux la structuration initiale des données.

Il faut en revanche prendre en compte le point évoqué ci-dessous avant le calcul des distances :

Les moments de l'initiation, du maximum, et de la fin représentent des points fixes dans le temps dont la valeur varie entre 1 (1 au 14 Janvier) et 26 (16 au 31 Décembre). Il est intuitif de dire que la fin du mois de Décembre et le début du mois de Janvier suivant sont des périodes similaires. Bien qu'appartenant à des années différentes, ces périodes sont en effet liées. Avec la notation présentée ci-dessus, attribuant le n°1 au début de Janvier et le numéro 26 à la fin Décembre, cette intuition n'est pas respectée. Pour pallier ceci, ce qui a un fort intérêt pour une méthode de classification basée sur une distance, une méthode de correction conditionnelle a été utilisée. Elle consiste à numéroter une fin d'année de manière négative (la quinzaine n°26 devient la quinzaine n°-1). Ceci est réalisé uniquement si cette correction conditionnelle réduit l'écart type de la numérotation (1 à 26) entre deux variables de même type au cours d'années consécutives. Pour comprendre ceci, un exemple détaillé est donné ensuite. Cette correction fait l'hypothèse qu'il existe une récurrence inter-annuelle des signaux de fréquences des classes d'anomalies de luminance, ce qui a bien été observé dans toute l'archive des fréquences des classes de  $Ra(\lambda)$ . L'exemple détaillé ensuite (page 68) montre l'intérêt d'utiliser ce type de correction pour le calcul des distances entre les classes  $Ra(\lambda)$  en fonction de leur phénologie. En effet, grâce à des valeurs volontairement (et intuitivement) quasi-identiques des métriques de la phénologie, l'exemple permet de comprendre l'impact de la correction conditionnelle sur la distance euclidienne.



**Exemple : la correction conditionnelle appliquée aux indices temporels** Soit une matrice pour 3 classes d’anomalies de luminance (A, B et C) et deux années (notée Y1 et Y2), pour lesquelles les moments d’initiations (Ini), de maximum (Max), de fin (Ter), de durée d’expansion (DE) et de déclin (DD) ont été estimées :

	←	—	Y1	—	→	←	—	Y2	—	→
	Ini	Max	Ter	DE	DD	Ini	Max	Ter	DE	DD
Ra( $\lambda$ ) A	1	6	12	6	7	26	6	12	7	7
Ra( $\lambda$ ) B	1	6	12	6	7	1	6	12	6	7
Ra( $\lambda$ ) C	26	3	8	4	6	26	3	8	4	6

Dans cet exemple, les Ra( $\lambda$ ) A et B sont quasi-identiques, excepté que la seconde initiation a lieu en quinzaine 26 pour le Ra( $\lambda$ ) A et en quinzaine 1 pour le Ra( $\lambda$ ) B. Ainsi, ils ont une phénologie très proche (identique la première année, une quinzaine plus tôt pour le Ra( $\lambda$ ) A la deuxième année). Le Ra( $\lambda$ ) C a une expansion beaucoup plus courte (maximum 1 mois et demi avant les autres, fin 2 mois avant les autres).

**Calcul de la distance euclidienne sans correction :** Le calcul de la distance euclidienne sans

	A	B
correction conditionnelle donne les valeurs suivantes :	B 25,01999	
	C 26,26785	36,19392

Ici, le Ra( $\lambda$ ) A est presque aussi similaire au Ra( $\lambda$ ) B qu’au Ra( $\lambda$ ) C. Les Ra( $\lambda$ ) B et C sont légèrement plus éloignés. Ce résultat n’est pas en adéquation avec ce qui est observé.

**Calcul de la distance euclidienne avec correction :** Afin de savoir s’il faut corriger le numéro de la quinzaine, les écarts types (SD) sont calculés entre les années pour chaque variable (hormis pour DE et DD, qui sont des durées), avant et après correction :

	Ini	Max	Ter		Ini	Max	Ter
SD <sub>A</sub> (Y1,Y2)	17,67	0	0	SD <sub>A</sub> (Y1,Y2)	<b>1,41</b>	19,09	19,09
SD <sub>B</sub> (Y1,Y2)	0	0	0	SD <sub>B</sub> (Y1,Y2)	19,09	19,09	19,09
SD <sub>C</sub> (Y1,Y2)	0	0	0	SD <sub>C</sub> (Y1,Y2)	19,09	19,09	19,09

Un seul écart-type est amélioré, pour l’initiation du Ra( $\lambda$ ) A (en gras dans le tableau). Ainsi, on corrige la valeur dans le tableau d’indice de la phénologie, la ligne qui concerne le Ra( $\lambda$ ) A devient :

	←	—	Y1	—	→	←	—	Y2	—	→
	Ini	Max	Ter	DE	DD	Ini	Max	Ter	DE	DD
Ra( $\lambda$ ) A	1	6	12	6	7	<b>-1</b>	6	12	7	7

Et ainsi, la distance euclidienne devient :

	A	B
B	2,24	
C	37,67	36,19392

Ici, les Ra( $\lambda$ ) A et B sont très similaires, alors que le Ra( $\lambda$ ) C est différent. Au vu des données initialement présentées, ce résultat correspond davantage à la réalité du paramètre étudié que le premier résultat.

Cette correction a été appliquée après le calcul des métriques de la phénologie et avant le calcul de la distance euclidienne entre les métriques des classes de  $Ra(\lambda)$ . Ces distances ont été utilisées pour grouper les classes de  $Ra(\lambda)$  en fonction de leurs indices phénologiques, et donc de leur saisonnalité, au moyen d'un algorithme de regroupement moyen (*UPGMA*).

La classification a été optimisée pour connaître la robustesse des groupes de  $Ra(\lambda)$  formés en fonction de leur phénologie (les groupes sont aussi appelés "branches", définies par des "nœuds") *via* une technique empruntée aux analyses phylogénétiques : le ré-échantillonnage par bootstrap multi-échelle (Shimodaira *et al.*, 2004).

Les techniques de "*bootstrap*" consistent à augmenter ou diminuer artificiellement le nombre d'observations prises en compte dans une classification (et ceci en dupliquant ou en sélectionnant des échantillons aléatoires). Le bootstrap est qualifié de multi-échelle car cette opération est réalisée pour les variables<sup>5</sup>, dont le nombre est réduit (par échantillonnage aléatoire), égal ou supérieur (par duplication aléatoire) au nombre de variables initiales (i.e., 10 métriques). Ainsi, 10 échelles ont été utilisées permettant de classer 5 à 15 variables. Pour chaque échelle, 1000 réplifications de la classification sont effectuées (nombre de réplifications à partir duquel les probabilités d'apparition des nœuds (voir ensuite) sont stables). On obtient ainsi 10 000 dendrogrammes pour une classification (soit 10 échelles répliquées 1000 fois). Il s'agit ensuite de construire un arbre en consensus majoritaire correspondant à une représentation de ces 10 000 dendrogrammes telle que leurs parties concordantes apparaissent clairement par rapport aux parties discordantes.

La robustesse de chaque nœud a été évaluée grâce à la valeur de bootstrap, qui représente la fréquence d'apparition du nœud à travers les 1000 réplifications. Le ré-échantillonnage par bootstrap multi-échelle a permis d'obtenir la probabilité "non biaisée" (*au pour approximately unbiased*) de l'existence de chaque nœud<sup>6</sup> (Shimodaira, 2002).

Une branche de la classification a été considérée comme robuste si la probabilité non biaisée était supérieure à 95 (soit une *p-value* > 0,95) (Suzuki et Shimodaira, 2006).

### 3.3 Résultats

#### 3.3.1 Les indices de la phénologie : exemple pour les classes de $Ra(\lambda)$ du capteur SeaWiFS sur la période 1998-2010

La Figure 3.5 montre les périodes d'initiation principales des 100 classes de  $Ra(\lambda)$  du capteur de couleur de l'océan SeaWiFS. Peu de classes de  $Ra(\lambda)$  ont une initiation variable d'une année à l'autre en considérant une période de 14 jours pour le calcul de phénologie.

5. **Rappel.** Les variables sont les suivantes : moment d'initiation, moment du maximum, moment de fin, durée d'expansion et durée de déclin (soit 5 métriques), au maximum pour deux expansions par an (soit 10 variables au total).

6. Cette valeur est dite "non biaisée" par rapport à la valeur de bootstrap car elle prend en compte la réduction ou l'augmentation du nombre de variable par rapport au jeu de données initial.

Cependant, c'est par exemple le cas de la classe de  $Ra(\lambda)$  n°30 qui présente une initiation du pic maximal instable, ayant lieu tantôt en Septembre, Octobre, ou Mai. Cette instabilité a en fait lieu pour les classes de  $Ra(\lambda)$  qui présentent deux pics de fréquence dans l'année, dont l'un ou l'autre peut être le pic maximal. *A contrario*, il existe également d'autres classes de  $Ra(\lambda)$  qui présentent cette distribution bi-modale, mais dont l'initiation du pic principal de fréquence est très stable. Sur la Figure 3.6.a, il est à noter que la plupart des initiations du pic maximal ont lieu principalement au mois de Septembre (quinzaine n°18). En effet, 58% des classes de  $Ra(\lambda)$  SeaWiFS ont une plus forte occurrence des initiations au mois de Septembre au cours des années (Figure 3.6.a).

Les périodes de fin sont aussi stables au cours des années pour les classes de  $Ra(\lambda)$  SeaWiFS (Figure 3.7). Cependant, une instabilité inter-annuelle apparaît pour les classes de  $Ra(\lambda)$  qui présentent un cycle d'augmentation de fréquence bi-modal (comme c'était le cas pour le moment d'initiation). Contrairement aux périodes d'initiations, les périodes de fin d'augmentation de la fréquence sont plus variées entre les classes de  $Ra(\lambda)$ . En effet, alors que 58% des classes de  $Ra(\lambda)$  ont une initiation du pic maximal de fréquence au mois de Septembre, les périodes de fin sont plus diversifiées ; les occurrences de fin d'augmentation les plus fréquentes étant 13% fin Mars et 11% fin Février (Figure 3.6.b).

Les moments de fréquences maximales pour les classes de  $Ra(\lambda)$  SeaWiFS sont également plus diversifiés que les initiations et les fin de pics (Figure 3.8). Pour les classes de  $Ra(\lambda)$ , la variabilité inter-annuelle suit les observations faites pour l'initiation et la fin du pic maximum de fréquence (c'est-à-dire, un moment du maximum plus variable pour les cycles de fréquences bimodaux). Les occurrences des maxima de fréquence sont plus importantes aux mois d'Août avec 14% des classes de  $Ra(\lambda)$  ayant leur moment de fréquence maximale ce mois (Figure 3.6.c).

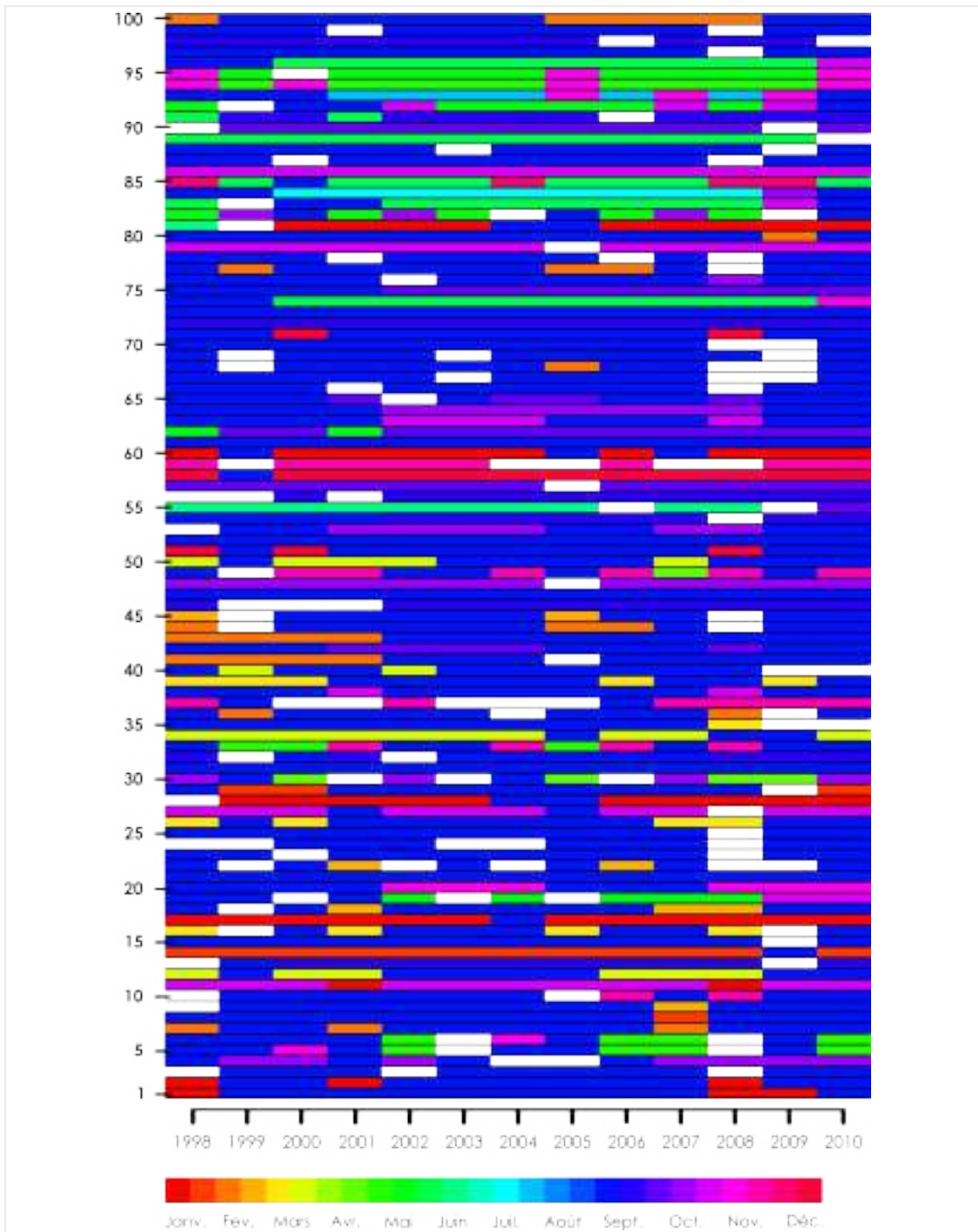


FIGURE 3.5 – Périodes d’initiation de l’augmentation de la fréquence des 100 classes de  $Ra(\lambda)$  du capteur de couleur de l’océan SeaWiFS (ordonnées) entre 1998 et 2010 (abscisse). Ces périodes sont représentées par une échelle de couleur. L’initiation du pic le plus important est montrée, c’est-à-dire l’augmentation de fréquence ayant la plus forte fréquence annuelle. L’année 1997 n’est pas représentée puisque elle n’est pas année complète. La couleur blanche est observée lorsque aucune augmentation de fréquence n’est détectée dans l’année.

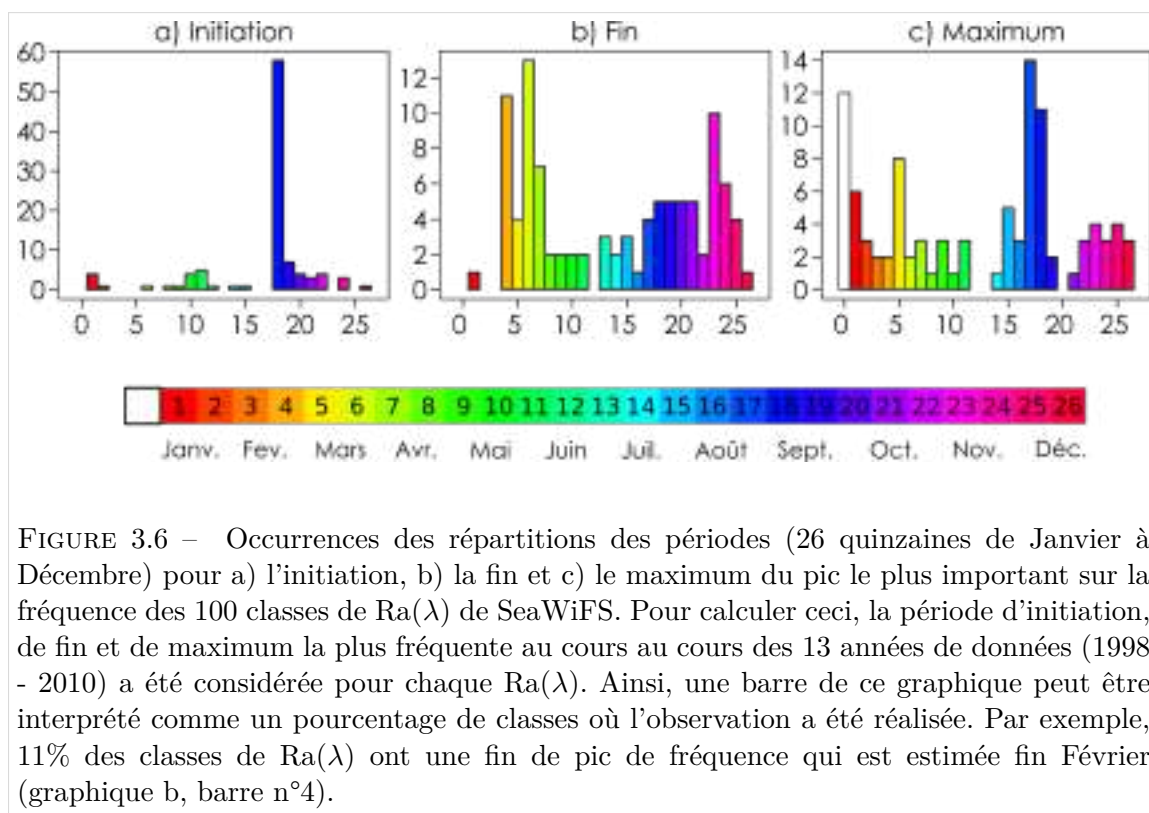


FIGURE 3.6 – Occurrences des répartitions des périodes (26 quinzaines de Janvier à Décembre) pour a) l’initiation, b) la fin et c) le maximum du pic le plus important sur la fréquence des 100 classes de  $Ra(\lambda)$  de SeaWiFS. Pour calculer ceci, la période d’initiation, de fin et de maximum la plus fréquente au cours au cours des 13 années de données (1998 - 2010) a été considérée pour chaque  $Ra(\lambda)$ . Ainsi, une barre de ce graphique peut être interprété comme un pourcentage de classes où l’observation a été réalisée. Par exemple, 11% des classes de  $Ra(\lambda)$  ont une fin de pic de fréquence qui est estimée fin Février (graphique b, barre n°4).

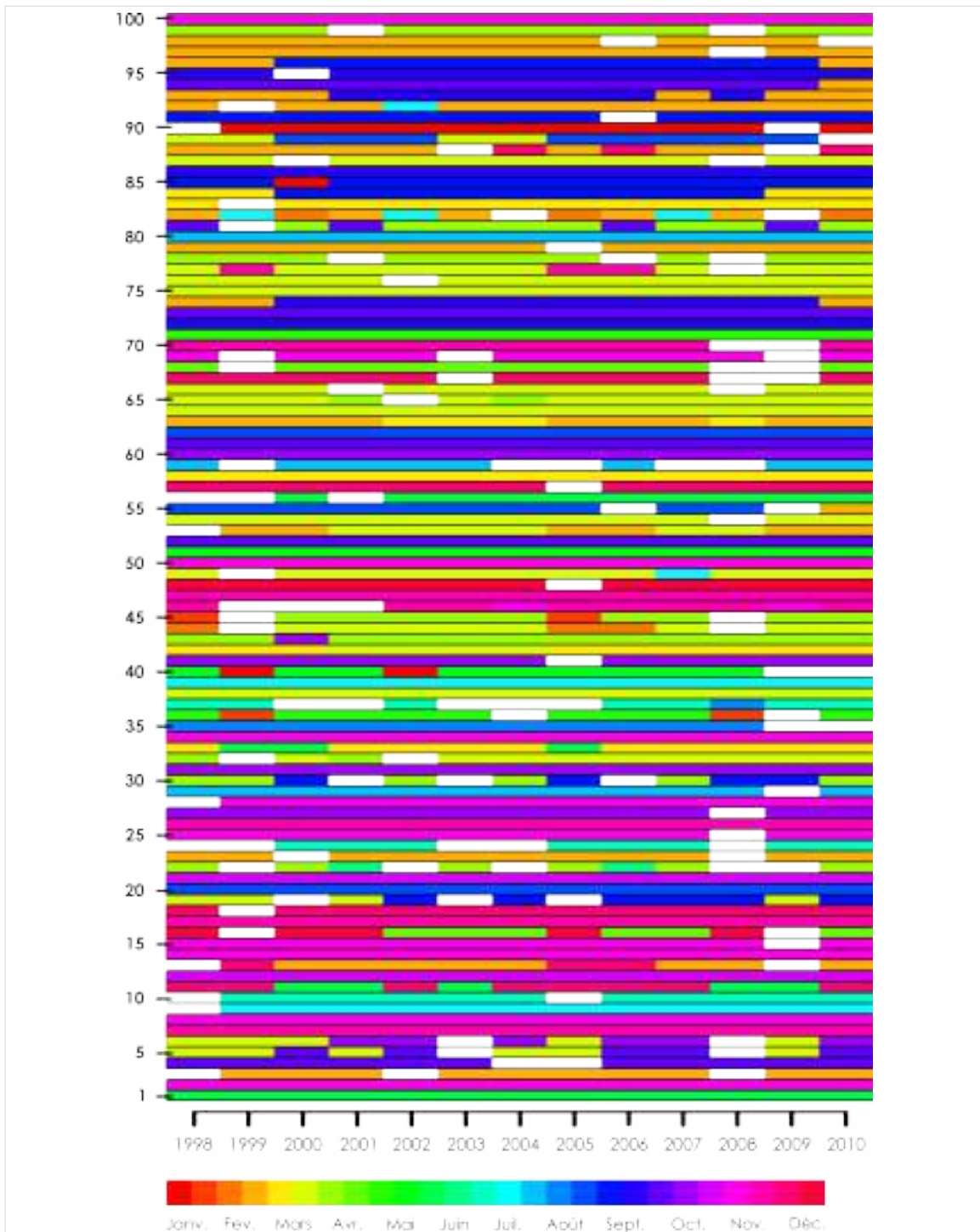


FIGURE 3.7 – Périodes de fin d’augmentation de la fréquence des 100 classes de  $Ra(\lambda)$  du capteur de couleur de l’océan SeaWiFS (ordonnées) entre 1998 et 2010 (abscisse). Ces périodes sont représentées par une échelle de couleur. La fin de l’augmentation la plus importante est montrée, c’est-à-dire l’augmentation de fréquence ayant la plus forte fréquence annuelle. L’année 1997 n’est pas représentée puisque elle n’est pas année complète. La couleur blanche est observée lorsque aucune augmentation de fréquence n’est détectée dans l’année.

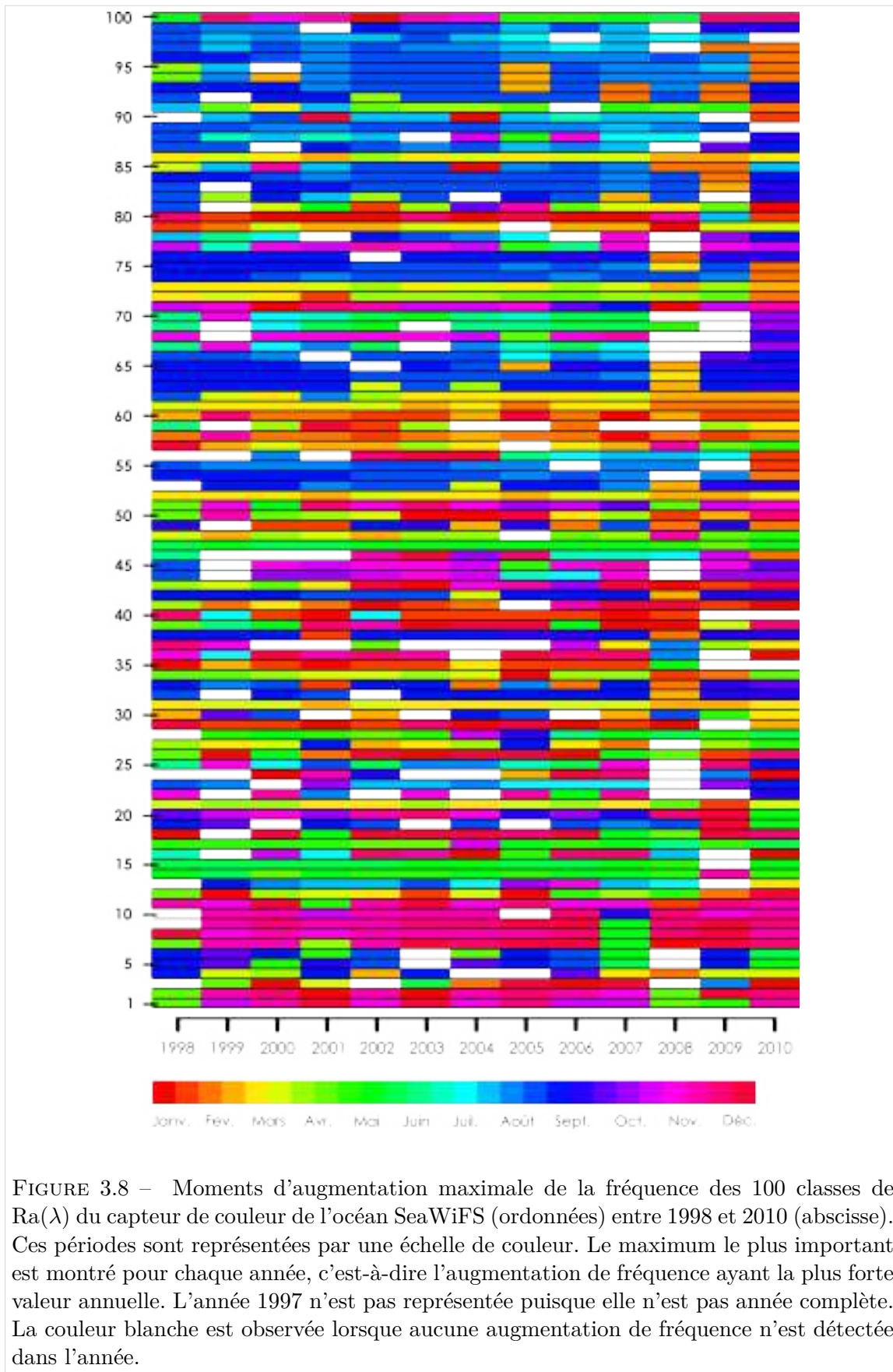


FIGURE 3.8 – Moments d’augmentation maximale de la fréquence des 100 classes de  $Ra(\lambda)$  du capteur de couleur de l’océan SeaWiFS (ordonnées) entre 1998 et 2010 (abscisse). Ces périodes sont représentées par une échelle de couleur. Le maximum le plus important est montré pour chaque année, c’est-à-dire l’augmentation de fréquence ayant la plus forte valeur annuelle. L’année 1997 n’est pas représentée puisque elle n’est pas année complète. La couleur blanche est observée lorsque aucune augmentation de fréquence n’est détectée dans l’année.

### 3.3.2 Classification des classes de $Ra(\lambda)$ grâce à leur phénologie

#### 3.3.2.1 Classification des 100 classes de $Ra(\lambda)$ issues de la carte auto-organisatrice

La Figure 3.9 montre le dendrogramme de classification au moyen d'un algorithme UPGMA et d'une distance euclidienne des 100 classes de  $Ra(\lambda)$  issues de la carte auto-organisatrice (Ben Mustapha *et al.*, 2014). D'après cette classification, et les probabilités associées, 18 nœuds sont considérés robustes selon les critères fixés par Shimodaira (2002) sur les probabilités d'apparition des nœuds. Le premier nœud formé à partir des données (les deux classes de  $Ra(\lambda)$  qui ont la distance euclidienne la plus petite) regroupe les classes de  $Ra(\lambda)$  n°21 et 61 (dont la distance euclidienne a été calculée à 19,03). Pour la suite des résultats, nous allons nous intéresser à trois nœuds : le premier formé, le quinzième (intermédiaire), et le 95<sup>ème</sup> nœud car il est le dernier regroupant deux classes de  $Ra(\lambda)$ . Ces trois nœuds sont choisis pour être détaillés dans les résultats car ils représentent des situations différentes, qui permettent d'illustrer le propos. Les distances euclidiennes entre ces classes de  $Ra(\lambda)$  basées sur leur phénologie sont montrées en Table 3.2.

TABLE 3.2 – Détails sur les classes de  $Ra(\lambda)$  SeaWiFS regroupées par les nœuds 1, 15 et 95 pendant la classification sur les indices phénologiques associées à la fréquence d'apparition de ces classes de  $Ra(\lambda)$ .

	Nœud 1	Nœud 15	Nœud 95
Classes de $Ra(\lambda)$ groupées par ce nœud	21 et 61	23 et 87	18 et 28
Distance euclidienne entre les $Ra(\lambda)$	19,03	66,19	169,24

Il est intéressant de se souvenir que les classes de  $Ra(\lambda)$  issues de la carte auto-organisatrice sont effectivement organisées, c'est-à-dire que, sur la base des caractéristiques des spectres (amplitude et forme), deux classes de  $Ra(\lambda)$  voisines dans la carte auto-organisatrice présentent des caractéristiques spectrales similaires. La Figure 3.10 montre la position dans la carte auto-organisatrice des classes de  $Ra(\lambda)$  regroupées dans les nœuds 1, 15 et 95. *A priori*, les classes de  $Ra(\lambda)$  qui ont les plus proches caractéristiques spectrales sont celles du dernier nœud formé (nœud 95, Figure 3.10, gauche).



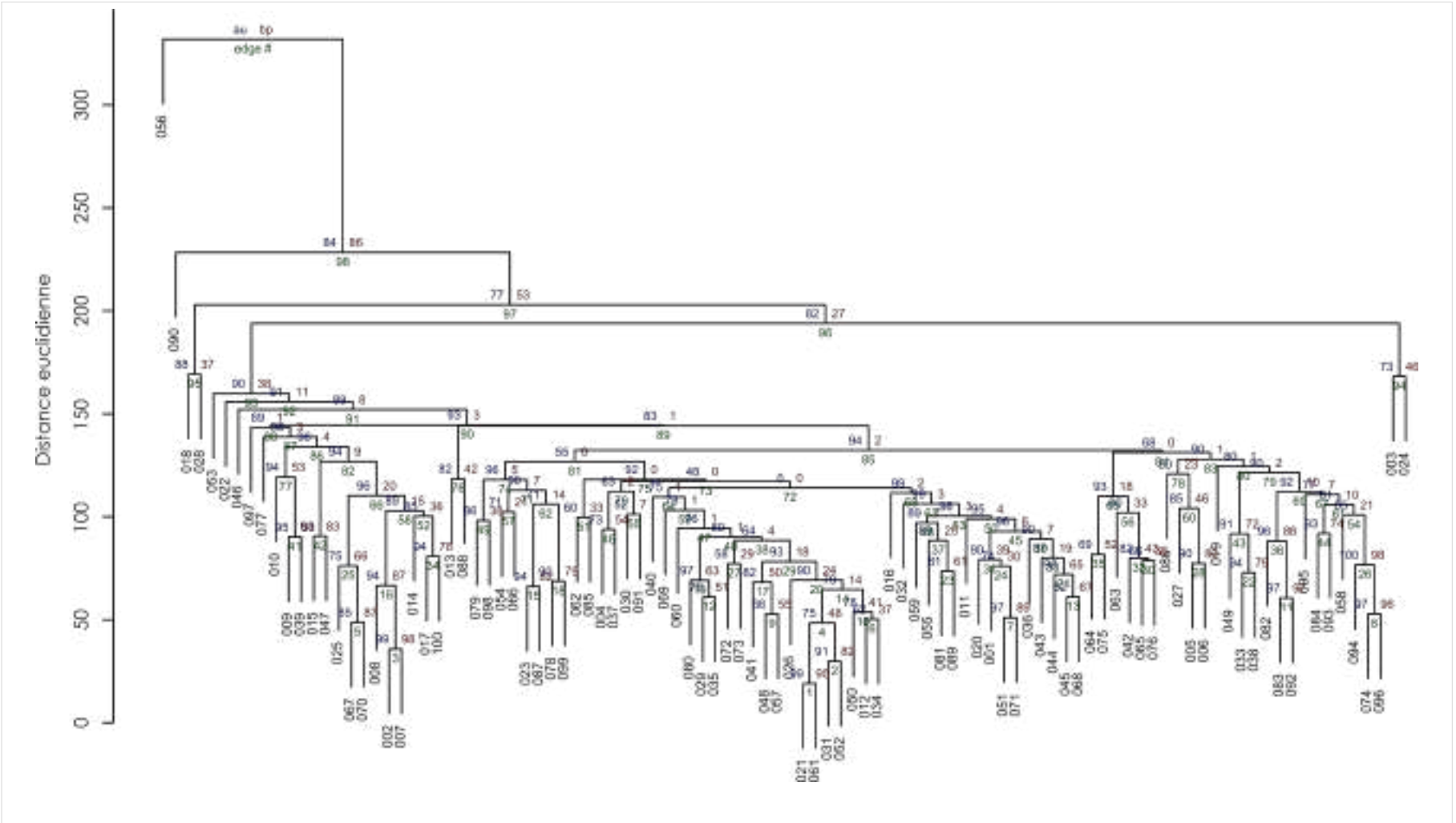


FIGURE 3.9 – Dendrogramme de la classification des indices de la phénologie des classes de  $Ra(\lambda)$  du capteur de couleur de l'océan SeaWiFS. Cette classification a été réalisée avec une méthode de *bootstrap-ressampling*, et les probabilités de bootstrap classique (*bp*, en rouge), "non biaisées" (*au*, en bleu) et le numéro du nœud indiquant l'ordre de formation des regroupements (*#edge*, en vert). Les nœuds dont la probabilité *au* est supérieure à 95 sont considérés comme robustes (Suzuki et Shimodaira, 2006).

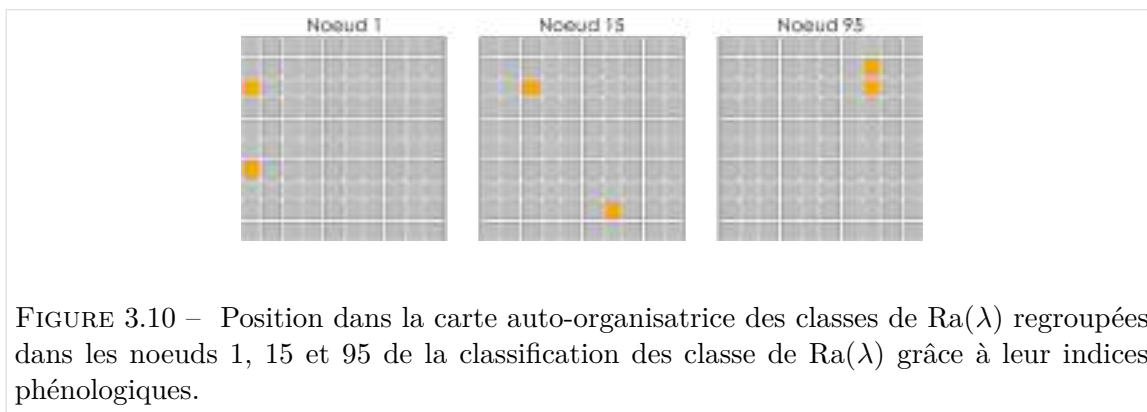


FIGURE 3.10 – Position dans la carte auto-organisatrice des classes de  $Ra(\lambda)$  regroupées dans les noeuds 1, 15 et 95 de la classification des classe de  $Ra(\lambda)$  grâce à leur indices phénologiques.

Afin de valider les groupes formés par cette méthode, une étape de contrôle *a posteriori* (non-automatisée) a été réalisée. En effet, il est assez difficile de connaître *a priori* le nombre de groupes qui devraient être formés à partir des 100 classes de  $Ra(\lambda)$ . Ainsi, la vérification s’est déroulée en 3 étapes :

- Les séries temporelles de fréquences des classes de  $Ra(\lambda)$  au sein des groupes formés sont-elles effectivement similaires ?
- Les spectres de ces classes de  $Ra(\lambda)$  sont-ils effectivement proches (forme et/ou amplitude) ?
- Les classes de  $Ra(\lambda)$  ont-elles une répartition spatiale similaire ?

### 3.3.2.2 Exemples de séries temporelles de fréquence des classes groupées

La Figure 3.11 montre les séries temporelles des fréquences de détection des signaux d’anomalies de luminosité de la classification *via* les indices phénologiques pour les trois exemples choisis.

Le noeud 1 (Figure 3.11.a), regroupant les classes de  $Ra(\lambda)$  *som* 21 et 61, dispose de deux séries temporelles de fréquence très similaires. D’après les distances euclidiennes, ce sont d’ailleurs les indices les plus similaires de toute la base de données de classes de  $Ra(\lambda)$ . Bien que les valeurs de fréquences soient éloignées d’un ordre de grandeur, ces deux séries temporelles ont une corrélation à  $r^2 = 0,90$ .

Le noeud 15 (Figure 3.11.b), qui regroupe les classes de  $Ra(\lambda)$  *som* 23 et 87, dispose également de deux séries temporelles de fréquence très similaires. La corrélation ici est plus forte que pour le noeud 1 ( $r^2 = 0,95$ ), ce qui s’explique par le fait que les deux séries temporelles ont des valeurs de fréquence du même ordre de grandeur.

Le nœud 95, qui regroupe les classes de  $Ra(\lambda)$  *som* 18 et 28 enregistre des séries temporelles assez différentes. En effet, ces deux séries temporelles sont assez bruitées, ce qui diminue la corrélation ( $r^2=0,53$ ). En revanche, de nombreuses augmentations et diminutions de fréquence ont lieu en même temps. La distance euclidienne est ici importante ce qui s'explique par un nombre de pics plus important pour l'une ou l'autre des classes de  $Ra(\lambda)$  selon les années. Par exemple, en 2007, deux pics sont enregistrés pour le  $Ra(\lambda)$  *som* 18 et un seul pour le  $Ra(\lambda)$  *som* 28, et c'est le contraire en 2009. C'est d'ailleurs ce qui est visible si l'on compare les distances entre les  $Ra(\lambda)$ , métriques par métriques : ces deux  $Ra(\lambda)$  sont proches en ce qui concerne les initiations et les fins de pic de fréquence (avec des distances euclidiennes de 33,7 et 34,6 respectivement) et moins proches pour les moments de maxima (distance euclidienne de 41,2).

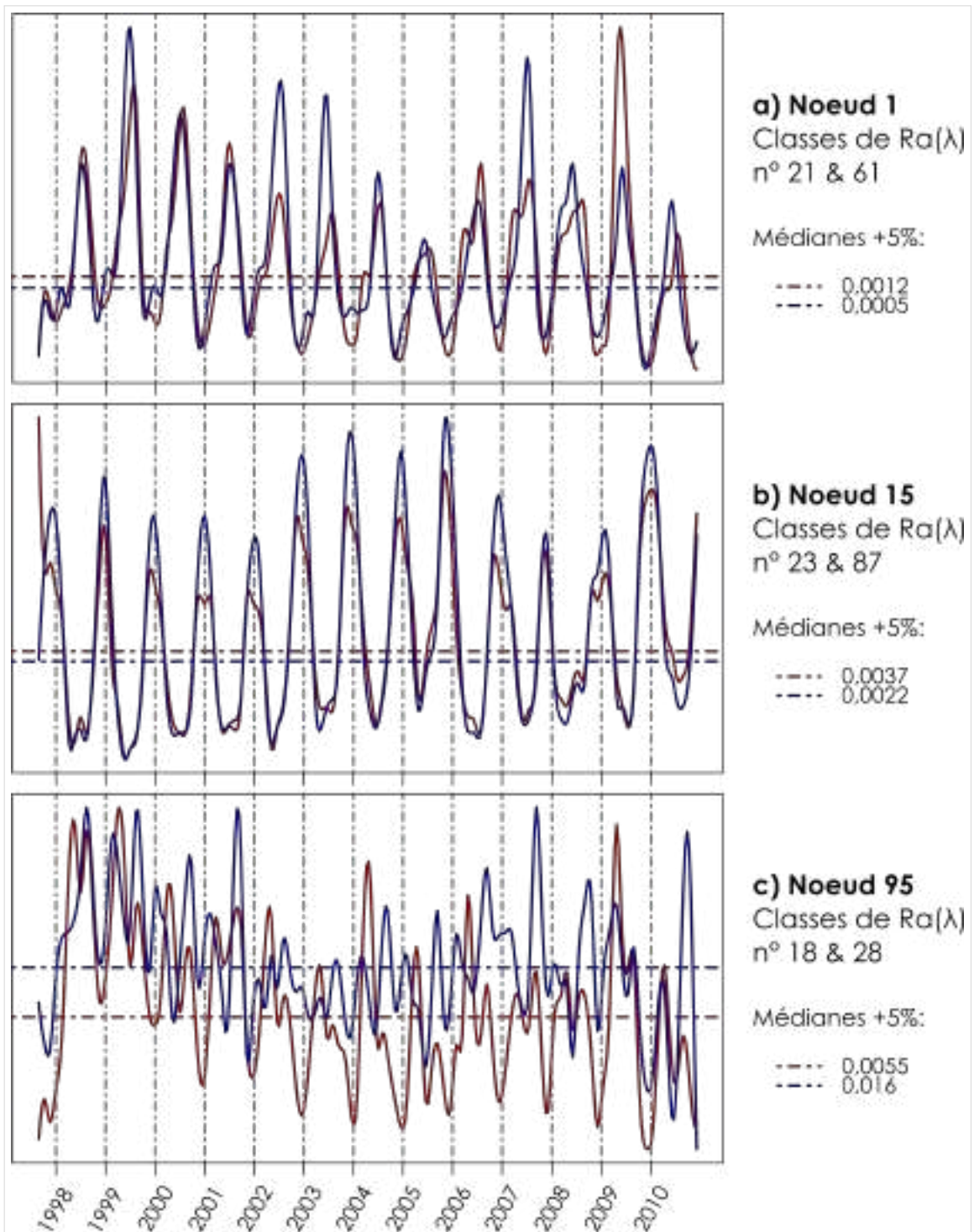


FIGURE 3.11 – Séries temporelles des fréquences d’occurrences à 15 jours moyennées sur le globe pour les classes d’anomalies de luminance SeaWiFS, classées en fonction de leur phénologie : a) noeud 1 de la classification ( $Ra(\lambda)$  21 et 61), b) noeud 15 de la classification ( $Ra(\lambda)$  23 et 87), et c) noeud 95 de la classification ( $Ra(\lambda)$  18 et 28). Les médianes tracées correspondent au seuil de décision utilisé pour estimer les métriques de la phénologie. Ces séries temporelles ont été normalisées par leur maximum pour leur représentation graphique afin de pouvoir les comparer visuellement de manière aisée.

### 3.3.2.3 Exemples de spectres des classes groupées

La Figure 3.12 montre les spectres référents des classes de  $Ra(\lambda)$  qui ont été groupées par la classification des indices de la phénologie pour les mêmes exemples que précédemment. Le nœud 1 regroupe deux classes de  $Ra(\lambda)$  qui ont la même forme (concave), mais une amplitude différente d'environ 0,5. Les nœuds 15 et 95 regroupent des spectres aux formes et amplitudes très similaires. Ceci était attendu pour le nœud 95 puisque les classes de  $Ra(\lambda)$  18 et 28 sont très proches dans la classification par la carte auto-organisatrice (Figure 3.10.c), mais on peut noter une légère différence pour la longueur d'onde 412 nm. En revanche, cette observation peut paraître plus surprenante pour le nœud 15, puisqu'il regroupe des anomalies de luminance assez éloignées dans la carte auto-organisatrice (Figure 3.10.b).

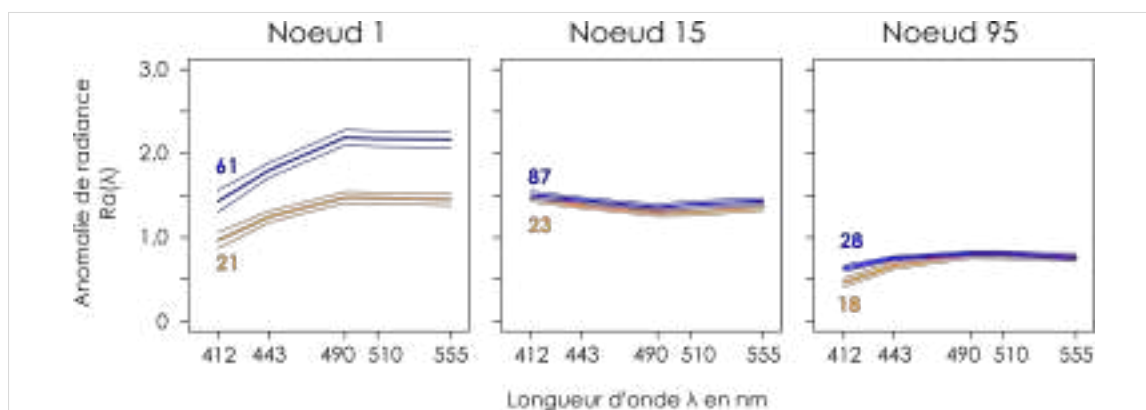


FIGURE 3.12 – Spectres de références des classes de  $Ra(\lambda)$  SeaWiFS qui ont été groupées selon leur phénologie. Le nœud 1 (gauche) groupe les classes de  $Ra(\lambda)$  n°21 et 61 ; le nœud 15 groupe les classes 23 et 87 et le nœud 95 groupe les classes 18 et 28.

### 3.3.2.4 Validation spatiale des regroupements de classes de $Ra(\lambda)$

Les Figures 3.13, 3.14 et 3.15 montrent respectivement les distributions spatiales (climatologies) des fréquences d'occurrence des classes de  $Ra(\lambda)$  groupées dans les nœuds 1, 15 et 95 de la classification des indices phénologiques.

Pour le nœud 1, groupant les classes de  $Ra(\lambda)$  21 et 61, les deux classes de  $Ra(\lambda)$  ont une répartition spatiale sensiblement similaire, hormis au mois de Janvier. En effet, un patch dans l'océan Austral apparaît pour la classe 21, et n'apparaît pas pour la classe 61 (uniquement quelques pixels). Ceci pourrait expliquer la différence observée dans l'amplitude des deux spectres de ces classes de  $Ra(\lambda)$ .

Pour le nœud 15, groupant les classes de  $Ra(\lambda)$  23 et 87, les patchs sont très similaires. La classe 23 est plus fréquente que la 87. Ces patterns sont localisés principalement dans l'océan Austral de Janvier à Avril, puis sur les côtes du Pacifique Nord et Sud et de

l’Afrique du Sud en Octobre.

Le noeud 95 regroupe des classes de  $Ra(\lambda)$  qui sont plus fréquentes que celles citées précédemment, et notamment à cause de la classe 28 qui a une fréquence médiane deux fois plus importante que la classe 18 (les deux classes de ce noeud). Les deux classes ont une distribution spatiale assez côtière, avec des incursions dans certains courants océaniques (Courant de Benguela par exemple en Atlantique Sud).

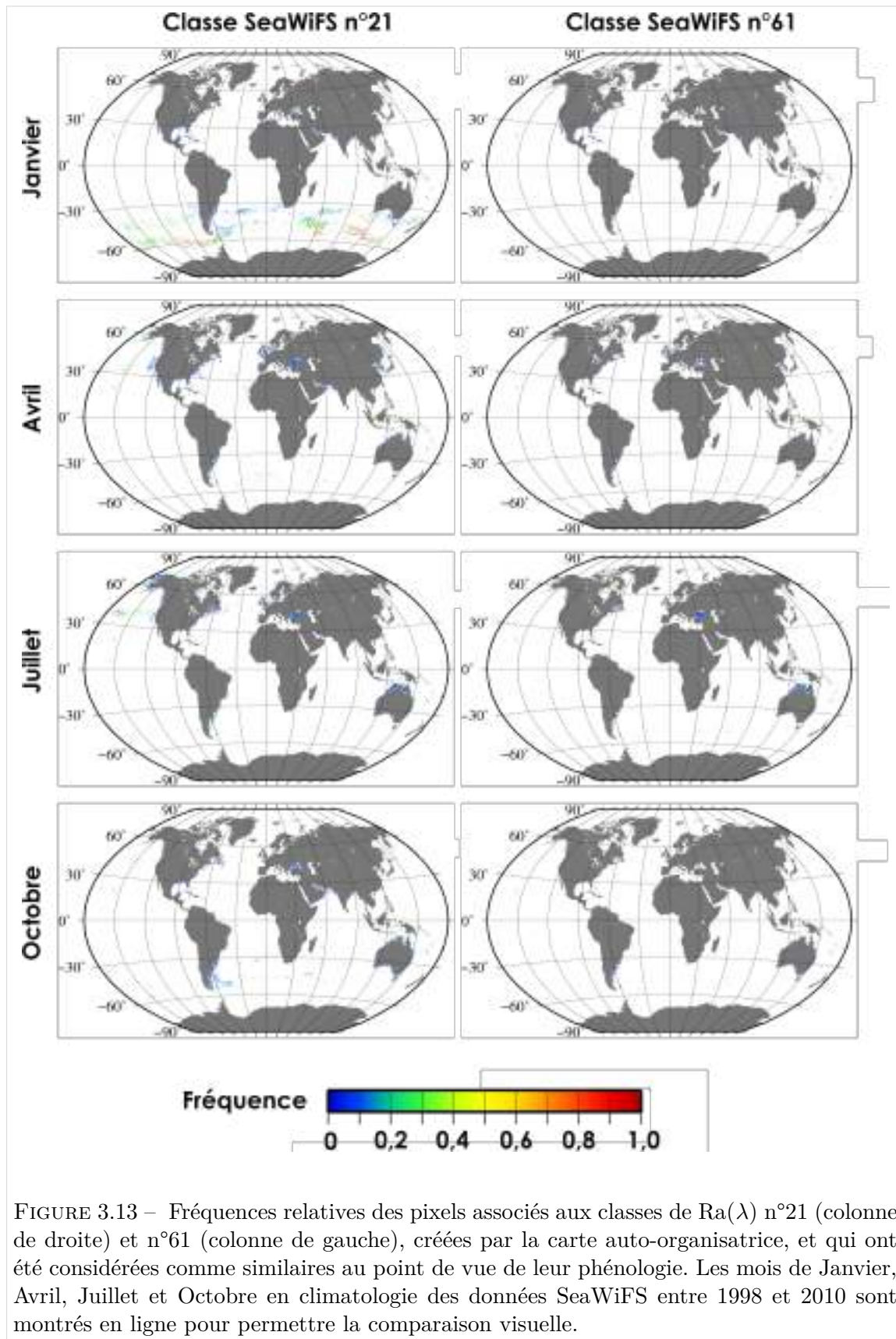


FIGURE 3.13 – Fréquences relatives des pixels associés aux classes de  $Ra(\lambda)$  n°21 (colonne de droite) et n°61 (colonne de gauche), créées par la carte auto-organisatrice, et qui ont été considérées comme similaires au point de vue de leur phénologie. Les mois de Janvier, Avril, Juillet et Octobre en climatologie des données SeaWiFS entre 1998 et 2010 sont montrés en ligne pour permettre la comparaison visuelle.

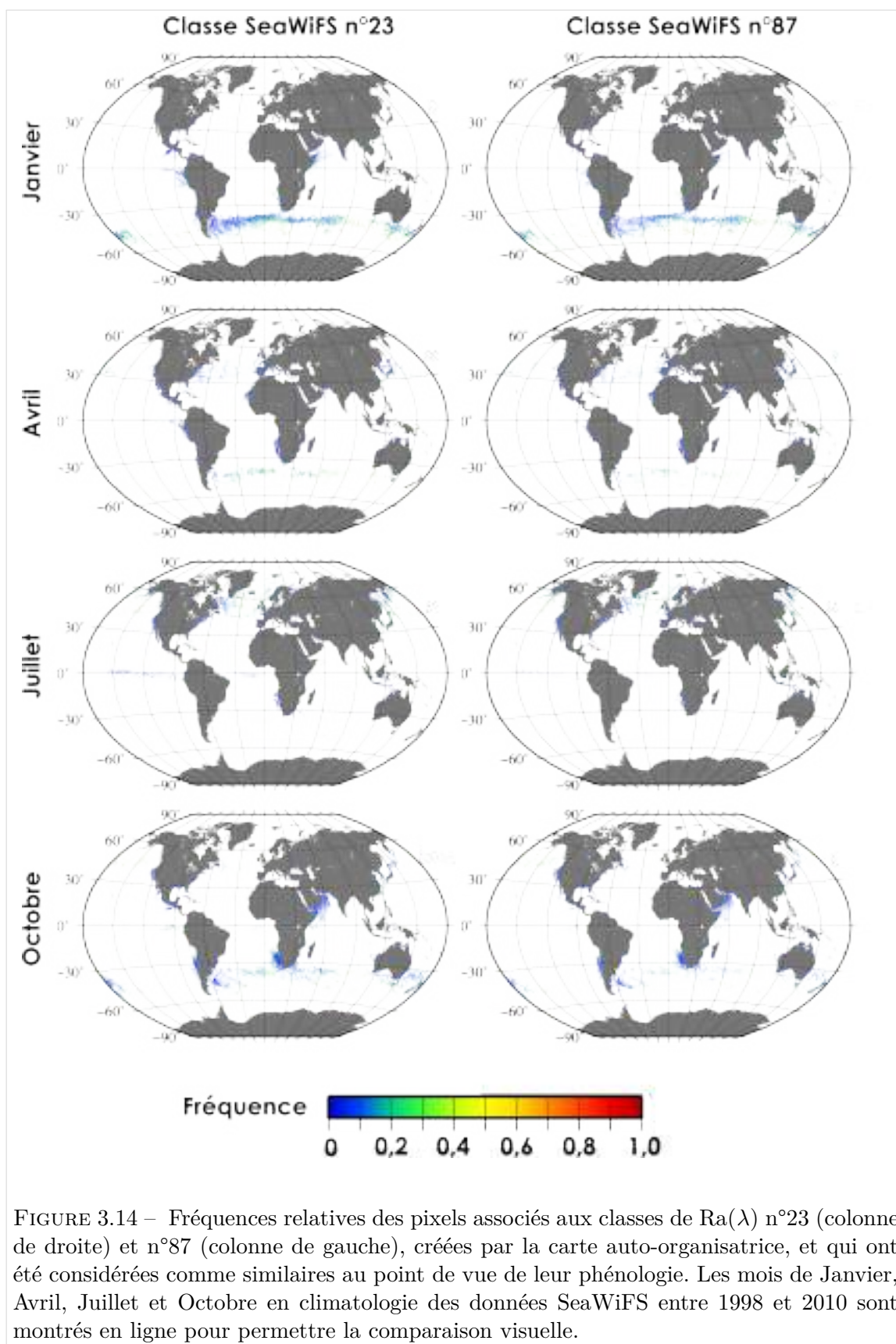


FIGURE 3.14 – Fréquences relatives des pixels associés aux classes de  $Ra(\lambda)$  n°23 (colonne de droite) et n°87 (colonne de gauche), créées par la carte auto-organisatrice, et qui ont été considérées comme similaires au point de vue de leur phénologie. Les mois de Janvier, Avril, Juillet et Octobre en climatologie des données SeaWiFS entre 1998 et 2010 sont montrés en ligne pour permettre la comparaison visuelle.



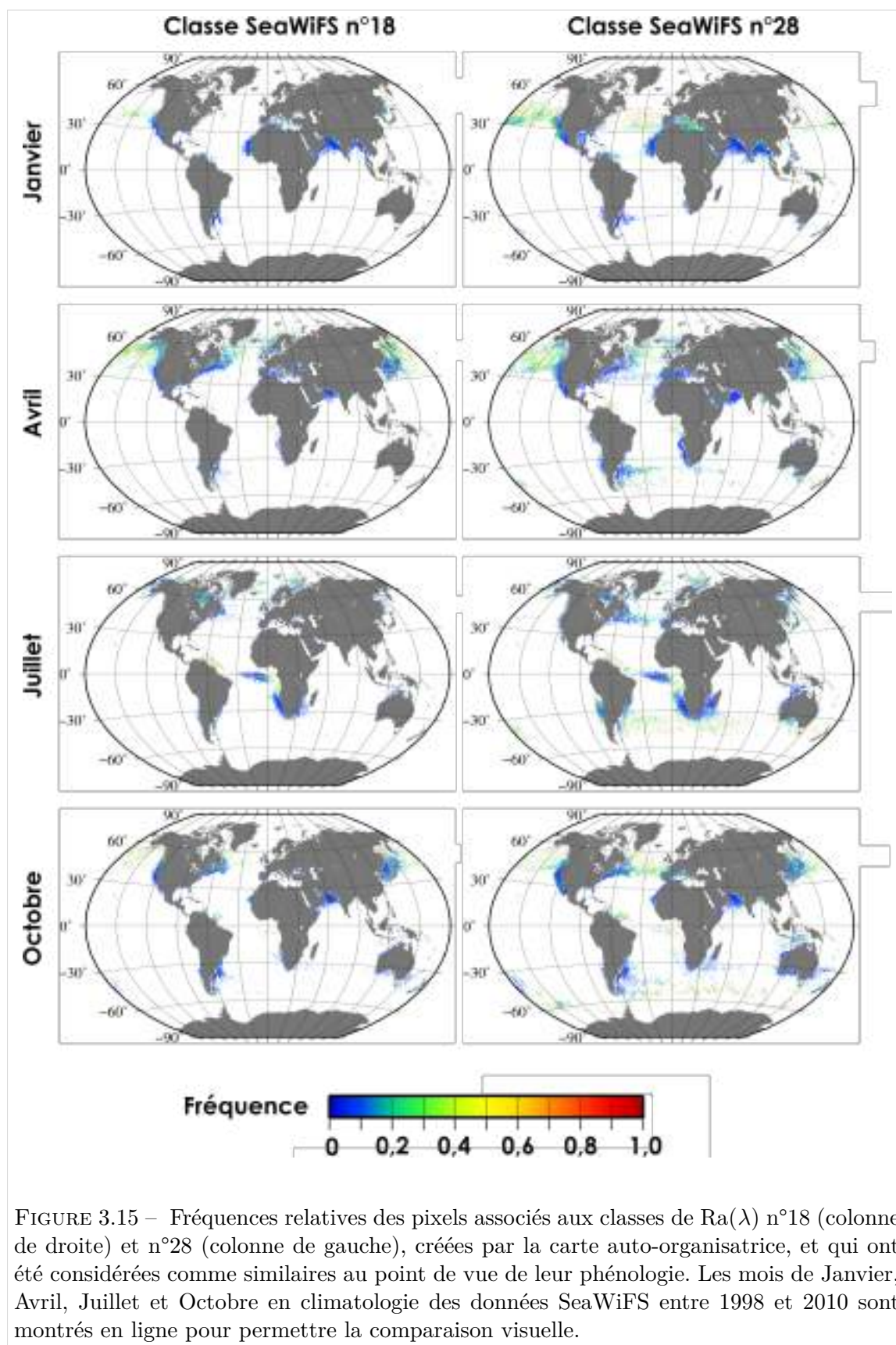


FIGURE 3.15 – Fréquences relatives des pixels associés aux classes de  $Ra(\lambda)$  n°18 (colonne de droite) et n°28 (colonne de gauche), créées par la carte auto-organisatrice, et qui ont été considérées comme similaires au point de vue de leur phénologie. Les mois de Janvier, Avril, Juillet et Octobre en climatologie des données SeaWiFS entre 1998 et 2010 sont montrés en ligne pour permettre la comparaison visuelle.

### 3.3.2.5 Diminution du nombre de classes de $Ra(\lambda)$ *via* la phénologie

Après avoir vérifié les caractéristiques précédentes (aspects temporel, spectral et spatial), les regroupements ont été (ou non) validés. Après ce regroupement pour les données SeaWiFS, il en résulte que 25 des 100 classes de  $Ra(\lambda)$  n'ont pas pu être groupées avec d'autres classes, soit car elles ne présentaient pas de caractéristiques similaires (20 parmi 25), soit parce que leur regroupement n'était pas rigoureux, i.e., faible valeur de probabilités de regroupement, mauvaise cohésion spatiale et spectrale (5 parmi 25). Ensuite, 30 classes de  $Ra(\lambda)$  forment un groupe avec une seule autre classe (soit 15 classes résultantes) ; 33 classes forment un groupe avec deux autres classes (soit 11 classes résultantes) ; et enfin, 12 classes forment des groupes de 4 classes (soit 3 classes résultantes). Ainsi, pour le capteur de couleur de l'océan SeaWiFS, le nombre de classes est passé de 100 (carte auto-organisatrice) à 54 qui comportent donc à la fois des caractéristiques spectrales différentes, mais également des distributions spatiales et temporelles différentes.

La Figure 3.16 montre les spectres référents de la carte auto-organisatrice organisés selon la nouvelle classification grâce à la phénologie (ces nouvelles classes seront appelées les classes *phénologie*, ou "SOM+P").

À partir de ces regroupements des classes de  $Ra(\lambda)$ , des cartes des classes les plus fréquentes peuvent être produites. Ces cartes peuvent être de l'échelle journalière (occurrence des classes) jusqu'à la climatologie. Un exemple de ces cartes est présenté en Figure 3.17 et l'intégralité des cartes en climatologie mensuelle est présentée en Annexe 7 (page 282). Ces cartes montrent que les structures spatiales sont effectivement conservées entre les deux classifications.

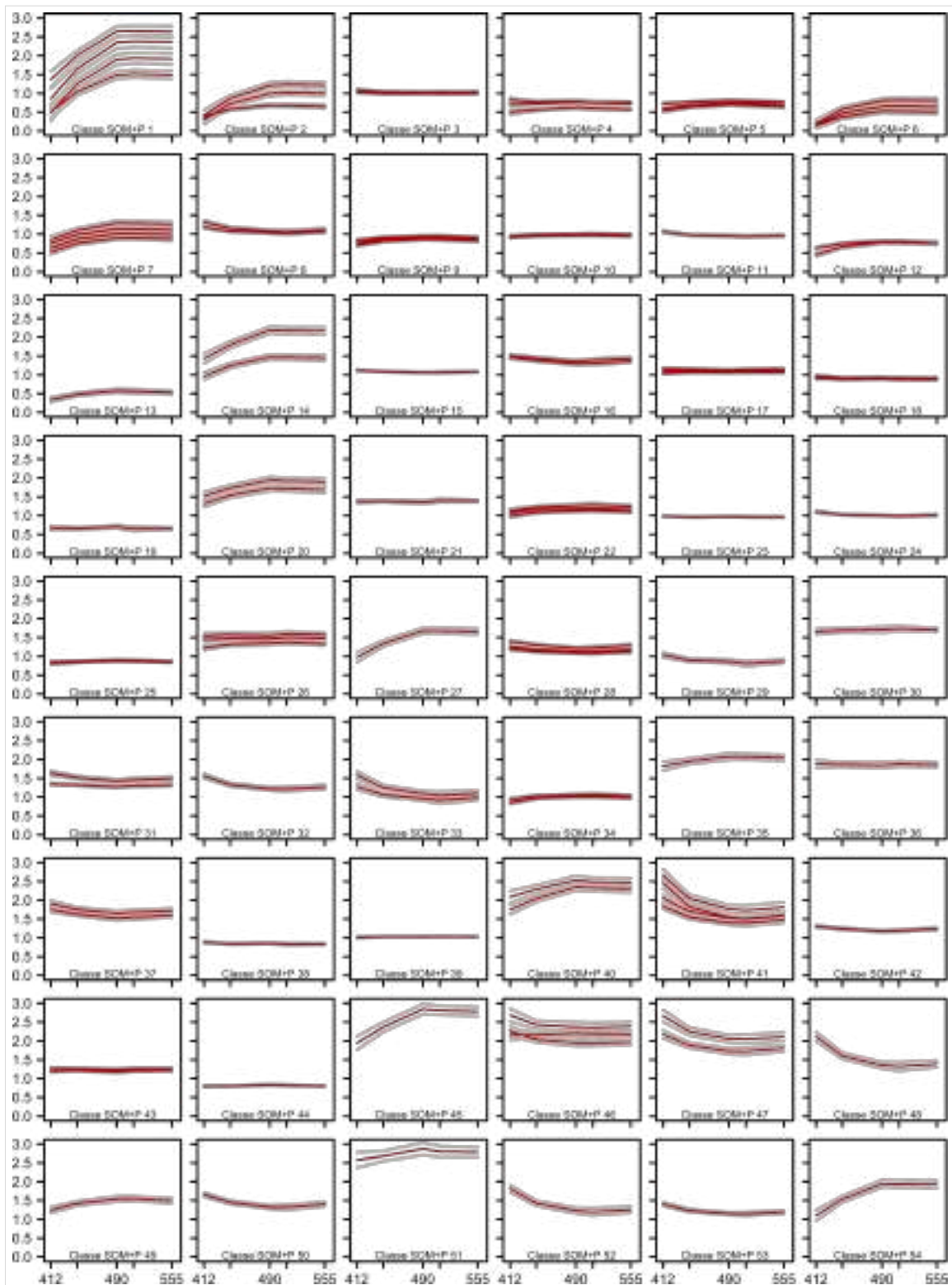
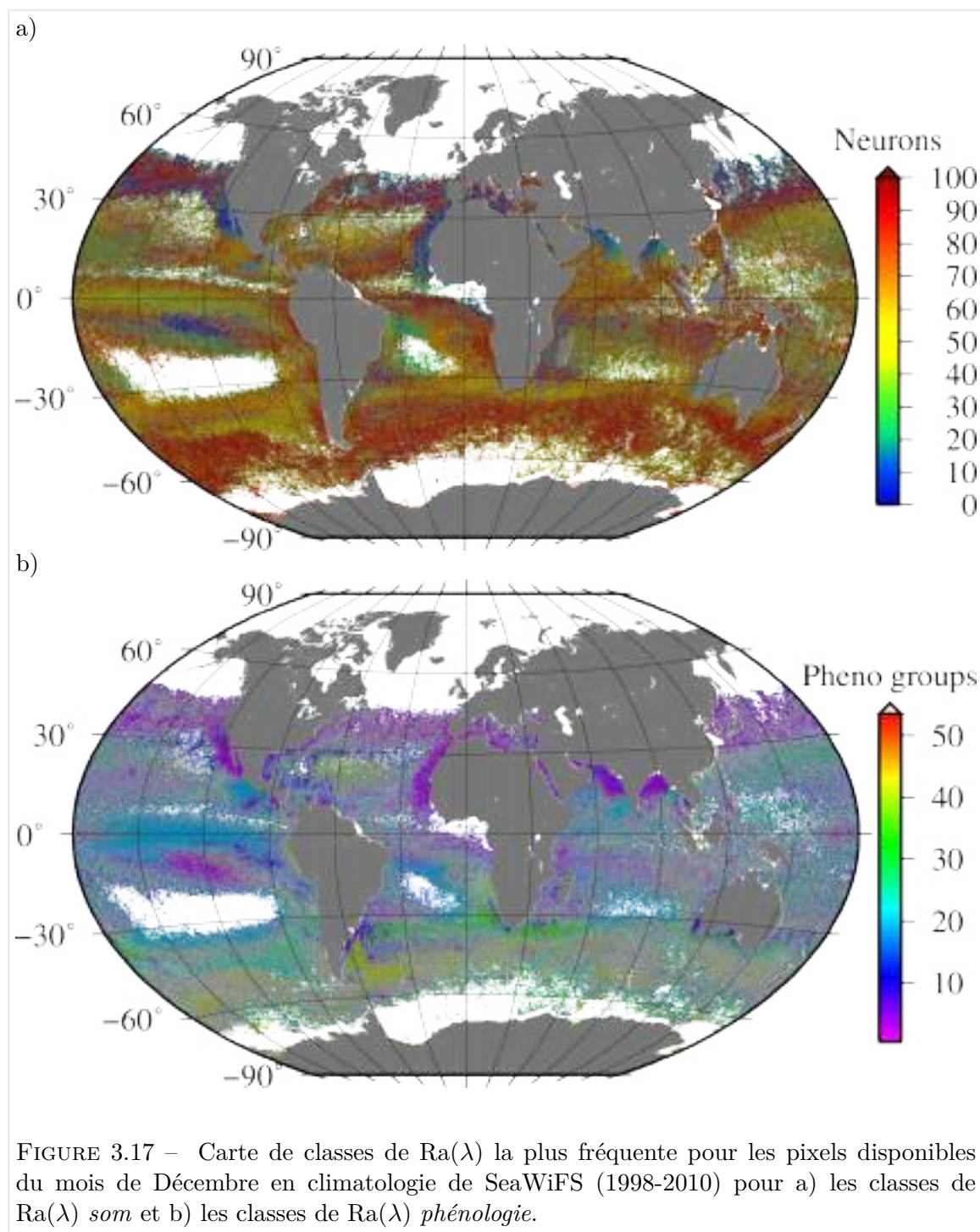


FIGURE 3.16 – Spectres référents des classes de  $R_a(\lambda)$  SeaWiFS (calculés par la carte auto-organisatrice) et groupées grâce à leur phénologie, dont le regroupement permet de former 54 classes de  $R_a(\lambda)$ . La ligne rouge représente les spectres référents de chaque classe et la ligne en pointillés l'écart standard.



### 3.4 Discussion

Pour cette étude, on disposait de séries temporelles de fréquence de détection à 14 jours pour une centaine de types d'anomalies PHYSAT identifiées grâce à une carte auto-organisatrice (Ben Mustapha *et al.*, 2014) appliquée aux données de couleur de l'océan SeaWiFS à l'échelle globale.

L'étude de la phénologie des fréquences de détection des anomalies PHYSAT a permis d'observer que celles-ci présentent des cycles saisonniers (i.e., augmentations de fréquence) variés dans leur mode (uni- ou bi-modal) et dans les périodes de maxima, d'initiations et de fins. Ces signaux présentent également des durées et des taux d'expansion variables. Ces observations sont en conformité avec ce qui est attendu pour des phénomènes biologiques en lien avec notre objet d'étude, le phytoplancton (e.g., Racault *et al.* (2012)).

L'idée d'utiliser les caractéristiques phénologiques pour regrouper *a posteriori* des signaux séparés arbitrairement, de part leur forme et leur amplitude en amont, est venue après observations des similarités pour certaines classes de  $Ra(\lambda)$ . Cependant, il convenait alors d'adapter les outils existant dans ce domaine à notre objet d'étude. En effet, les méthodes "classiques" d'estimation de la phénologie permettent d'appréhender les variations moyennes des fréquences d'anomalies pour l'expansion principale du signal mais elles ne sont pas adéquates pour une grande partie du jeu de données satellites qui présente des cycles bimodaux (51% des classes de  $Ra(\lambda)$  SeaWiFS ont une expansion bimodale). Une méthode phénologique plus flexible (basée sur une série de conditions détaillées précédemment) a donc été développée. L'utilisation des métriques de la phénologie dans le cadre de cette approche flexible permet une définition précise de la forme de la série temporelle des classes de  $Ra(\lambda)$ . Grâce aux exemples présentés auparavant, pour des classes de  $Ra(\lambda)$  qui ont pu être regroupées avec cette méthode, nous avons vu que les spectres des classes regroupées résultantes ("SOM+P") ont une variabilité de spectres tout-à-fait acceptable si on les compare avec les résultats empiriques des cas dominants définis par Alvain *et al.* (2005). Mais ce n'est pas vraiment le cas pour le groupe SOM+P n°1, où la variabilité de l'amplitude est assez élevée. En revanche, la forme très particulière, la plus concave qui est trouvée dans les données SeaWiFS, n'est trouvée dans aucune autre classe (ou alors la concavité est moins prononcée). Ce groupe est identifié principalement en Avril au niveau des côtes Nord-américaines, une région active et complexe, ce qui peut être un indice pour expliquer cette variabilité (qui restera à étudier plus précisément à l'avenir).

Cependant, pour deux classes de  $Ra(\lambda)$  dont les spectres sont proches, la classification *via* la carte auto-organisatrice aurait pu entraîner une séparation "aléatoire" en fonction des données qui lui sont présentées dans la phase d'apprentissage. Cela expliquerai pourquoi, dans certains cas (comme le nœud 95 présenté en résultat), les spectres référents soient très proches, et que la distribution spatiale également, alors qu'en se basant sur les séries

temporelles bruitées, il n'est pas aisé de faire le rapprochement entre ces deux classes de  $Ra(\lambda)$ . Dans ce genre de cas, l'utilisation des métriques de la phénologie est intéressante. En effet, les variables prises en compte pour l'estimation de la phénologie ne décrivent que les aspects principaux de la série temporelle : à partir de quand la fréquence augmente, à quelle vitesse et quand atteint-elle son maximum, à quelle vitesse et quand cette fréquence revient à un niveau basal.

Bien que nécessitant idéalement une étape finale non-automatisée<sup>7</sup> pour vérifier les regroupements des classes de  $Ra(\lambda)$  (en fonction à la fois de leurs caractéristiques spectrales, temporelles et spatiales), la classification à l'aide des outils de la phénologie établie précédemment semble être une solution adaptée pour regrouper *a posteriori* les classes de  $Ra(\lambda)$  formées par la carte auto-organisatrice. En effet, cela permet de réduire fortement le nombre de classes de  $Ra(\lambda)$ , en se basant sur des paramètres en lien avec l'objet d'étude, ce qui permettra par la suite d'optimiser les associations avec les données *in-situ*.

La seconde classification proposée en introduction, qui consistait à former des superclasses en groupant les classes de  $Ra(\lambda)$  en fonction de la forme et de l'amplitude du spectre référent de la classe donne parfois des résultats similaires à la classification *via* la phénologie (encadré en ligne pleine de la Figure 3.18), mais également des résultats différents (encadré en ligne pointillée de la Figure 3.18). Sur cette Figure, on peut remarquer qu'aussi bien pour la classification *via* la phénologie (Figure 3.18.a), que pour les super-classes spectrales (Figure 3.18.b), une certaine structure est conservée : des classes adjacentes ou proches sur la carte auto-organisatrice peuvent être regroupées par l'une ou l'autre des classifications.

Cependant, la classification en super-classes *via* la forme et l'amplitude des spectres ne respecte pas l'hypothèse faite pour cette étude : des classes de  $Ra(\lambda)$  sont "jumelles" si elles ont un spectre, une fréquence d'apparition et une répartition spatiale similaire. Par exemple, le groupe "spectral" (issu des super-classes) n°3 (Figure 3.18.b, chiffre en gras) regroupe un grand nombre de classes (10 classes de  $Ra(\lambda)$ ) : ces 10 classes ont une phénologie différente (6 classes *phénologie* différentes, Figure 3.18.b, chiffre en gras), mais également des répartitions spatiales différentes (cartes en Annexe 6, page 279). Ainsi, la comparaison de ces deux classifications permet de mettre en évidence qu'il ne faut pas juste prendre en compte la forme et l'amplitude du spectre d'anomalie de luminance, mais aussi sa distribution spatio-temporelle pour un regroupement.

Cette classification *via* les outils de la phénologie a été testée dans cette thèse pour les capteurs SeaWiFS et MODIS à l'échelle globale et également à l'échelle régionale (voir Chapitre 5). Elle devient une étape supplémentaire dans la chaîne de traitement des données d'anomalies de luminance dans le cadre de la nouvelle version de PHYSAT.

---

7. Actuellement, cette étape n'est pas automatisée, mais elle est automatisable. L'automatisation n'a pas été réalisée pour cette thèse puisque l'étape de regroupement n'est nécessaire qu'une fois par capteur.

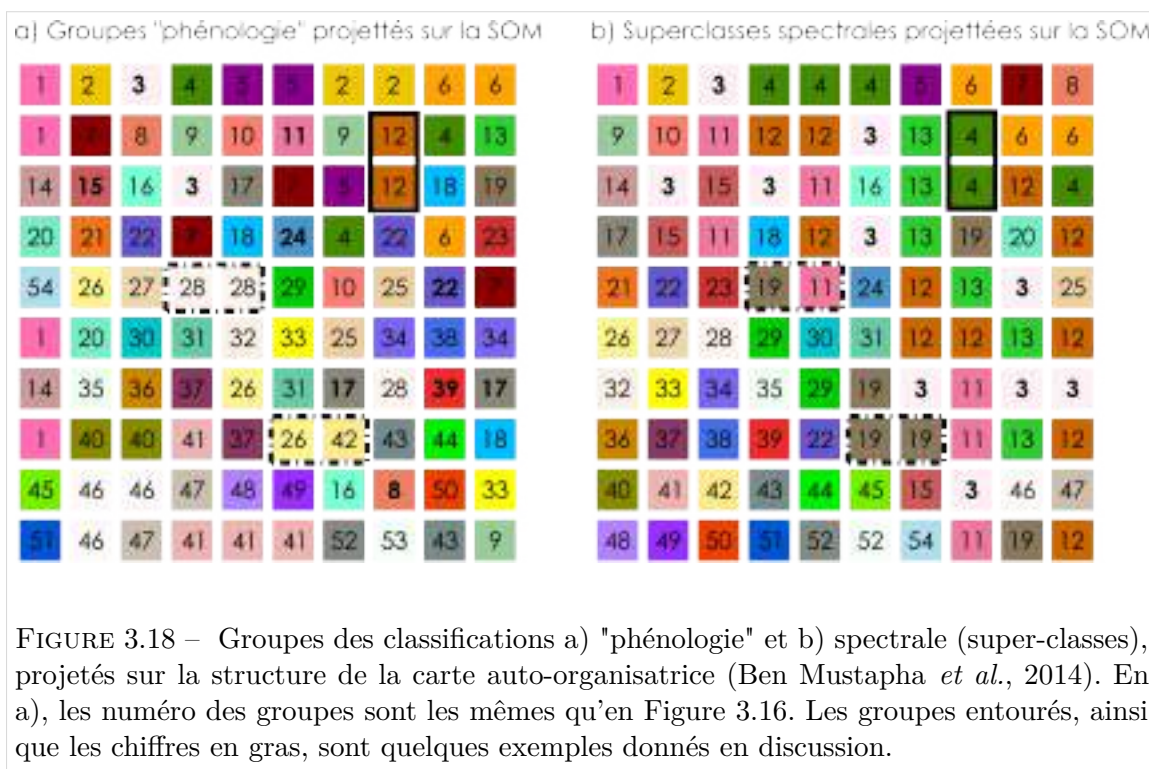


FIGURE 3.18 – Groupes des classifications a) "phénologie" et b) spectrale (super-classes), projetés sur la structure de la carte auto-organisatrice (Ben Mustapha *et al.*, 2014). En a), les numéros des groupes sont les mêmes qu'en Figure 3.16. Les groupes entourés, ainsi que les chiffres en gras, sont quelques exemples donnés en discussion.

### 3.5 Conclusion et perspectives

Dans ce Chapitre, une classification *a posteriori* a été réalisée pour regrouper les classes de  $Ra(\lambda)$  issues d'une précédente classification par une carte auto-organisatrice en fonction de leur fréquence d'occurrence. Cette classification présente des avantages :

- Cela évite de diviser un signal de  $Ra(\lambda)$  uniquement sur des critères statistiques basés sur la forme et l'amplitude du spectre, ce qui était fait auparavant avec l'utilisation unique de la carte auto-organisatrice. En conséquence, les chances d'obtenir des meilleurs résultats lors d'une calibration des signaux de  $Ra(\lambda)$  sont potentiellement augmentées, aussi bien au niveau de la calibration que de la validation. En effet, la réduction du nombre de classes de  $Ra(\lambda)$  limite les chances de dispersion artificielle d'une même situation phytoplanctonique dans plusieurs classes au sein des mesures coïncidentes.
- Cela permet aussi de distinguer deux  $Ra(\lambda)$  qui auraient des spectres similaires en terme de forme et d'amplitudes, mais qui aurait des comportements d'occurrence différents (aussi bien temporellement que spatialement). Cela permet de garder la possibilité d'identifier plus de cas de "confusion", i.e., deux groupes différents, qui selon la situation<sup>8</sup>, pourraient présenter la même anomalie de luminance, comme

8. Pour rappel, les cas ayant été signalés comme potentiellement confus se réfèrent à l'étude théorique de Alvain *et al.* (2012) (Section 2.2.4, Chapitre 2). En effet, en théorie, il est possible que le couplage des variations des *IOPs* engendre une même anomalie de luminance.

l'ont mis en évidence Alvain *et al.* (2012), mais soit, à terme, identifiable grâce aux critères supplémentaires issus des outils de phénologie.

- Enfin, lors d'une calibration, si les mesures coïncidentes sont insuffisantes pour labelliser une classe de  $Ra(\lambda)$  (pas assez de mesures coïncidentes par exemple), et que cette classe est étroitement liée à une autre classe (phénologie, spectre, distribution spatiale) qui a pu être labellisée, cela permet également d'étendre la **labellisation** à cette classe. Une validation sera alors effectuée *a posteriori* afin de vérifier cette labellisation étendue lorsque de nouvelles mesures coïncidentes seront disponibles.

Cette classification a été testée sur plusieurs capteurs de couleur de l'océan, à l'échelle régionale et globale. Une limitation pourrait exister pour les applications régionales, où il faut vérifier auparavant que la fréquence d'apparition n'est pas biaisée par un manque de couverture des données à certaines périodes de l'année (au niveau des hautes latitudes par exemple). Ces outils pourront être intégrés à la chaîne de traitement des données de PHYSAT et les classes de  $Ra(\lambda)$  ainsi regroupées seront mises à disposition des utilisateurs afin d'optimiser les labellisations à venir.





# Détection des assemblages phytoplanctoniques à l'échelle globale par PHYSAT

---

## Problématique de ce Chapitre

Dans le Chapitre 2 (page 27), nous avons détaillé l'explication théorique du lien entre le signal de la méthode PHYSAT et la présence d'un groupe phytoplanctonique spécifique (cas de dominance). Cette explication théorique a permis de montrer deux choses :

- Les signaux de PHYSAT sont liés à des caractéristiques du phytoplancton. En effet, des facteurs comme la forme, la taille, le contenu cellulaire et les matières dissoutes associées, influent sur *IOPs*, et leur combinaison engendre des anomalies de luminance différentes (forme et amplitude)
- Les signaux utilisés dans l'approche PHYSAT pourraient donc en théorie être associés à des situations phytoplanctoniques au-delà des seuls cas de dominance

La question posée dans ce Chapitre est donc :

**Est-il possible, en se basant sur les résultats du Chapitre 3 (page 55) et sur une analyse plus précise des données *in-situ*, d'associer les anomalies de luminance PHYSAT à la présence de situations phytoplanctoniques plus diversifiées que la dominance ?**

Pour répondre à cette question, nous allons présenter une version alternative de la calibration de PHYSAT (par rapport à Alvain *et al.* (2005)), afin d'apprécier le potentiel de détection des assemblages phytoplanctoniques à l'échelle globale.

## Introduction

Un *assemblage* phytoplanctonique est un regroupement d'espèces<sup>1</sup>, qui co-existent sur une même fenêtre temporelle dans une même zone géographique (celles-ci pouvant être en mouvement) ; car elles possèdent une physiologie, morphologie et / ou d'autres facteurs qui induisent une réponse similaire face aux variations des conditions environnementales.

Dans la nature, ces assemblages sont souvent récurrents (Margalef, 1978) et la proportion des *populations* (i.e., les individus d'une espèce<sup>1</sup>) est variable dans une communauté en fonction de la modification du milieu ; il y a une succession d'espèces au sein d'un assemblage.

Il existe de nombreuses observations d'assemblages phytoplanctoniques récurrents dans les océans ouverts ou en milieux côtiers. Voici deux exemples bien documentés :

- Des petites cellules phytoplanctoniques telles que les *Prochlorococcus-like* et les *Synechococcus-like* sont connues pour préférer des environnements similaires, oligotrophes, comme les gyres océaniques (e.g. Dandonneau et Niang (2007)).
- Les diatomées et les dinoflagellés partagent également des environnements similaires, mésotrophe ou eutrophe, souvent côtiers (e.g., Hinder *et al.* (2012)), bien que les diatomées préfèrent une colonne d'eau mélangée et les dinoflagellés une colonne d'eau stratifiée (Margalef, 1978).

Dans certains cas, un type phytoplanctonique dans l'assemblage se met à croître plus rapidement qu'un autre, car il est plus efficace dans le milieu ; c'est une situation de dominance, liée à une interaction de compétition pour le milieu (l'espèce la plus efficace domine), ou à l'intervention d'un prédateur (l'espèce la moins consommée domine).

Lors de la première mise en place de l'algorithme PHYSAT (Alvain *et al.*, 2005), les signaux d'anomalie de luminance ont été mis en relation avec un état de dominance établi au préalable sur les concentrations en *pigments biomarqueurs* de la campagne *in-situ* GEP&CO (Dandonneau *et al.*, 2004; Alvain *et al.*, 2005). Cela signifie que les cas de dominance ont été recherchés dans les mesures coïncidentes, puis un signal spécifique à ces cas a été déduit empiriquement. L'objectif à l'époque était de rechercher la présence

---

1. "Espèce" sera utilisé dans le manuscrit pour définir des sous unités directes d'un assemblage, sachant que ces espèces peuvent aussi correspondre à des groupes d'espèces, de genres ou de groupes (fonctionnels/de tailles etc.)

---

de signaux spécifiques et les cas de dominance étaient un choix le plus évident pour cette étude exploratoire (*Alvain, communication personnelle*).

La dominance *in-situ* a été évaluée grâce à une étude de la littérature, permettant de définir des seuils sur les concentrations en pigments dits "biomarqueurs" à partir desquels le type de phytoplancton était dominant. En établissant ces seuils de dominance sur les mesures coïncidentes (voir Chapitre 2), *Alvain et al.* (2005) ont restreint l'information potentiellement associable aux données satellites d'anomalies de luminance. Cependant, les signaux satellites d'anomalie de luminance étaient observés bruts (les  $Ra(\lambda)$  calculés), et non en classes comme actuellement avec la carte auto-organisatrice (**SOM**). Lors de la première mise en place de la méthode (*Alvain et al.*, 2005), le format "classes", nécessaire à l'établissement d'un lien empirique entre les occurrences de deux types de données, a donc dû être recherché en amont dans les données *in-situ* (les données *in-situ* ont été regroupées en classes *via* les informations qui concernent la dominance). Cette étape de recherche de classes dans les données *in-situ* a également permis de s'affranchir de la quantité de phytoplancton, pour se concentrer sur l'établissement du lien entre sa qualité (présence d'un type de phytoplancton particulier) et la présence d'une anomalie de luminance spécifique.

À ce jour (2017), plusieurs changements ont eu lieu (présentés dans les Chapitres précédents) et facilitent une nouvelle tentative de calibration pour les données d'anomalies de luminance, sans se limiter *a priori* aux cas de dominance :

- Grâce à l'utilisation du réseau de neurones pour classer en amont les anomalies de PHYSAT en 100 classes (*Ben Mustapha et al.*, 2014), la calibration des anomalies dans PHYSAT est rendue plus facile. En effet, l'information croisée correspond au maximum à 100 classes de  $Ra(\lambda)$  \*  $x$  groupes *in-situ* (chaque spectre de  $Ra(\lambda)$  observé étant attribué à une classe), alors qu'elle était de  $\infty$  \*  $x$  groupes avant la classification mathématique des spectres de  $Ra(\lambda)$  (chaque spectre étant unique, comme s'il était lui-même sa propre classe de  $Ra(\lambda)$ ). De plus, ces classes *som* ont été regroupées pour former des classes de  $Ra(\lambda)$  "**phénologie**" pour lesquelles les occurrences temporelles et spatiales sont proches, ainsi que les caractéristiques des spectres (Chapitre 3). L'utilisation de ces classes de  $Ra(\lambda)$  évite la définition, *a posteriori*, de critères liés à la spécificité des formes et amplitudes des  $Ra(\lambda)$  ;
- Le dernier processing des données de couleur de l'océan (R2014.0, Franz (2009)) permet d'augmenter le nombre de données colocalisées entre les données *in-situ* et les signaux de PHYSAT<sup>2</sup> ;
- Plusieurs campagnes océanographiques (ou bases de données) ont collecté ces types d'échantillons depuis 2005 : OOMPH, OISO, MAREDAT, COLIBRI, ICOTA, CRO-

---

2. R2014. Meilleurs algorithmes de corrections atmosphériques, meilleures estimations des concentrations en chlorophylle *a*, réductions des biais

ZEX, KEOPS, BOUM, HALIFAX, et ont mesuré les concentrations pigmentaires par **HPLC**. Cela a pour effet d'augmenter le nombre potentiel de mesures colocalisées. Des détails sur ces campagnes sont disponibles en Annexe 8 (page 285).

En utilisant ces nouvelles informations, nous avons évalué la possibilité de calibrer PHYSAT en y intégrant une diversité plus grande de groupes phytoplanctoniques, au-delà de la dominance. Ce Chapitre sera divisé en trois suivant le schéma en Figure 2.1 (page 28 au Chapitre 2) : pré-traitement de l'information *in-situ* (Section 4.1.1, ensuite), pré-traitement de l'information de couleur de l'océan (Section 4.1.2, à partir de la page 103), et mise en relation des deux niveaux d'information au moyen des mesures colocalisées (Section 4.1.3, à partir de la page 104).

## 4.1 Matériels et méthodes

### 4.1.1 Analyse des données *in-situ* : Étape 1

#### 4.1.1.1 Les concentrations pigmentaires

Pour la nouvelle calibration de PHYSAT, les données pigmentaires, et plus spécifiquement les **pigments biomarqueurs**, dont les concentrations sont estimées par **HPLC**, restent d'actualité (comme dans la version Alvain *et al.* (2005), voir Chapitre 2). En effet, ce sont les données relatives au phytoplancton qui sont les plus communes et diversifiées spatialement et temporellement. Elles sont souvent mesurées lors de campagnes traversant plusieurs océans.

Des campagnes ont notamment estimé les concentrations des pigments biomarqueurs suivants : Péridinine (abrégé *Peri*, biomarqueur des dinoflagellés), Fucoxanthine (abrégé *Fuco*, biomarqueur des diatomées), Zéaxanthine (abrégé *Zea*, biomarqueur des *Synechococcus*-like), 19'-HexanoxyFucoxanthine (abrégé *HF19*, biomarqueur des Nanoeucaryotes), Divinyl-Chlorophylle-*a* (abrégé *Div-chla*, biomarqueur des *Prochlorococcus spp.*).

La définition des assemblages grâce aux concentrations pigmentaires est une étape délicate. En effet, la concentration en pigments n'est pas le reflet d'une abondance de telle ou telle espèce (ou groupe d'espèces). Alvain *et al.* (2005) ont pallié ce problème en utilisant des seuils sur les concentrations en pigments relatives aux concentrations en chlorophylle-*a* (Équation 4.1). Ces seuils statistiques ont été évalués à partir de la littérature pour permettre la distinction des cas de dominances, c'est-à-dire les cas pour lesquels le groupe phytoplanctonique représente plus de 50% de la concentration en chlorophylle-*a* (Alvain *et al.*, 2006, 2008). Ces seuils sont présentés en Table 4.1 (page 97).

$$[\text{Pigment}]_{\text{Rel}} = \frac{[\text{Pigment}]}{[\text{Chl-}a] + [\text{Div. Chl-}a]} \quad (4.1)$$

Où  $[\text{Pigment}]_{\text{Rel}}$  est la concentration relative du pigment biomarqueur,  $[\text{Pigment}]$  est la concentration mesurée du pigment biomarqueur et  $[\text{Chl-}a] + [\text{Div. Chl-}a]$  est la somme des concentrations observées de chlorophylle-*a* et de Div. Chl-*a*.

TABLE 4.1 – Seuils sur les concentrations relatives en pigments biomarqueurs utilisés pour associer empiriquement un inventaire de pigments à un groupe de phytoplancton dominant spécifique. Les valeurs en gras correspondent aux biomarqueurs de chaque groupe nommé en première colonne. Extrait de Alvain *et al.* (2005).

	Div.a	Périd.	Fuco.	HF'19	Zéa.
Diatomées	<0,40	<0,10	> <b>0,18</b>	-	<0,20
Prochlorococcus	> <b>0,40</b>	<0,10	-	-	>0,35
Nanoecaryotes	<0,40	<0,10	-	> <b>0,14</b>	<0,20
<i>Synechococcus-Like</i>	<0,40	<0,10	-	-	> <b>0,20</b>
Dinoflagellés	<0,40	> <b>0,10</b>	-	-	<0,20

#### 4.1.1.2 Informations générales sur les données *in-situ* de cette étude

Le jeu de données pigmentaire actuel comporte 2444 points *in-situ*, dont 1,68% des échantillons a déjà été utilisé pour la précédente calibration de PHYSAT lors de la recherche des cas de dominance (issus de campagnes GEP&CO, Alvain *et al.* (2005)).

Dans ce jeu de données, les dominances ont été calculées de la même manière que pour Alvain *et al.* (2005). Le diagramme en Figure 4.1 montre la répartition de ces dominances dans les données *in-situ* actuelles. Sur cette Figure, la majorité des cas de dominances sont trouvés pour les nanoecaryotes (24,1%), puis *Synechococcus-Like* (16,1%), les diatomées (10,5%), les *Prochlorococcus spp.* (9,5%) et les dinoflagellés (4,4%). Il est noté que 35,4 % des échantillons ne présentent pas de cas de dominance. Ce sont ces échantillons qui sont particulièrement intéressants puisqu'ils représentent la partie utilisable pour l'étude du potentiel de détection d'assemblages non dominants par la méthode PHYSAT.

Les 2444 échantillons sont issus d'une base de données plus grande, pour laquelle les points suivants ont été écartés :

- Les outliers ont été supprimés. Ce sont les concentrations en pigments biomarqueurs en marge, déterminées comme les 5% supérieurs de la distribution totale du pigment considéré. De même, les concentrations en pigments biomarqueurs supérieures aux concentrations en chlorophylle-*a* sont considérés comme des outliers.

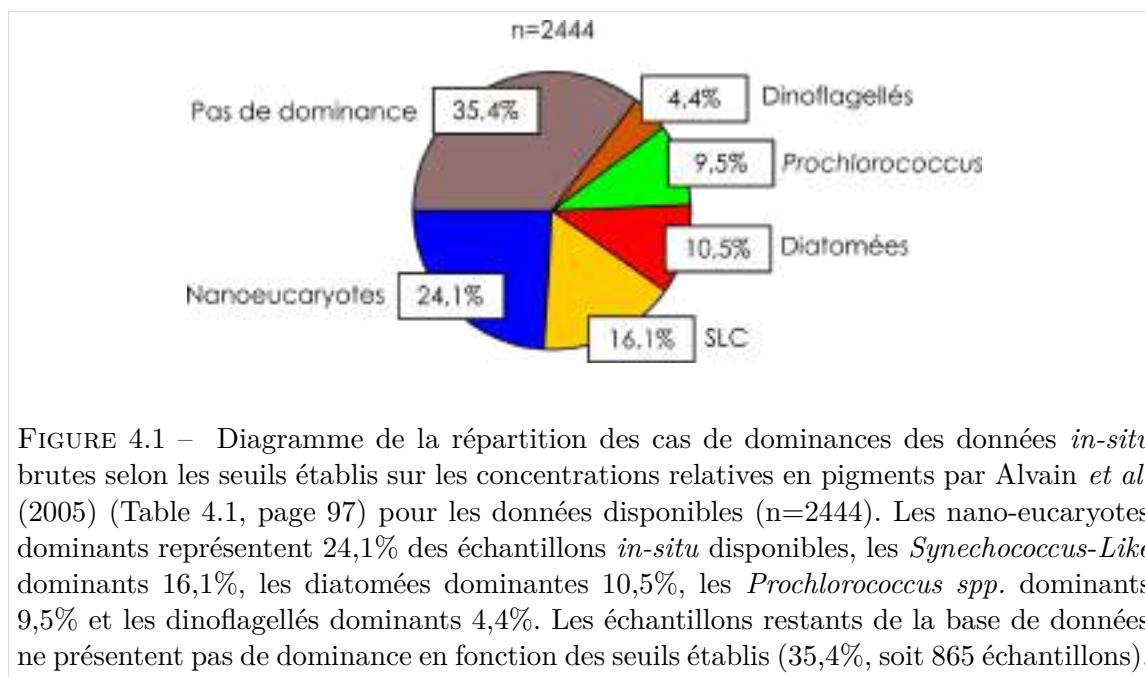


FIGURE 4.1 – Diagramme de la répartition des cas de dominances des données *in-situ* brutes selon les seuils établis sur les concentrations relatives en pigments par Alvain *et al.* (2005) (Table 4.1, page 97) pour les données disponibles ( $n=2444$ ). Les nano-eucaryotes dominants représentent 24,1% des échantillons *in-situ* disponibles, les *Synechococcus-Like* dominants 16,1%, les diatomées dominantes 10,5%, les *Prochlorococcus spp.* dominants 9,5% et les dinoflagellés dominants 4,4%. Les échantillons restants de la base de données ne présentent pas de dominance en fonction des seuils établis (35,4%, soit 865 échantillons).

- Les données situées dans les zones dites "côtières" d'après la classification des provinces de Longhurst (Figure 4.2; Longhurst (2010)) ont été supprimées. En effet, nous ne nous intéressons qu'aux zones d'océans ouverts et du Cas I (masque prédéfini pour PHYSAT, Chapitre 2).

Nous rappelons que le but de ce Chapitre est d'associer les mesures *in-situ* à des anomalies de luminance spécifiques, classées *via* la carte auto-organisatrice et la phénologie. Pour cela, il existe deux manières de procéder :

- Puisque l'information satellite est déjà ordonnée en classes, nous pouvons exploiter les données *in-situ* telles qu'elles sont. Dans ce cas, pour chaque classe d'anomalie de luminance, il faudra analyser les données *in-situ* correspondantes, et trouver la ressemblance entre ces données *in-situ a posteriori* (i.e., ces données forment-elles un ensemble cohérent?). Cette approche ne permet pas réellement de réaliser une validation sans donner de critères *a posteriori*. En effet, des critères doivent permettre d'évaluer si l'ensemble des données *in-situ* associées à une classe de  $Ra(\lambda)$  est cohérent (par exemple des seuils ou des ratios de concentrations relatives en pigments biomarqueurs).
- Les données *in-situ* peuvent également être ordonnées en classes en amont, afin de faciliter la recherche de correspondances lors de la mise en parallèle des données *in-situ* et des données satellites (il suffit de rechercher dans ce cas les paires récurrentes

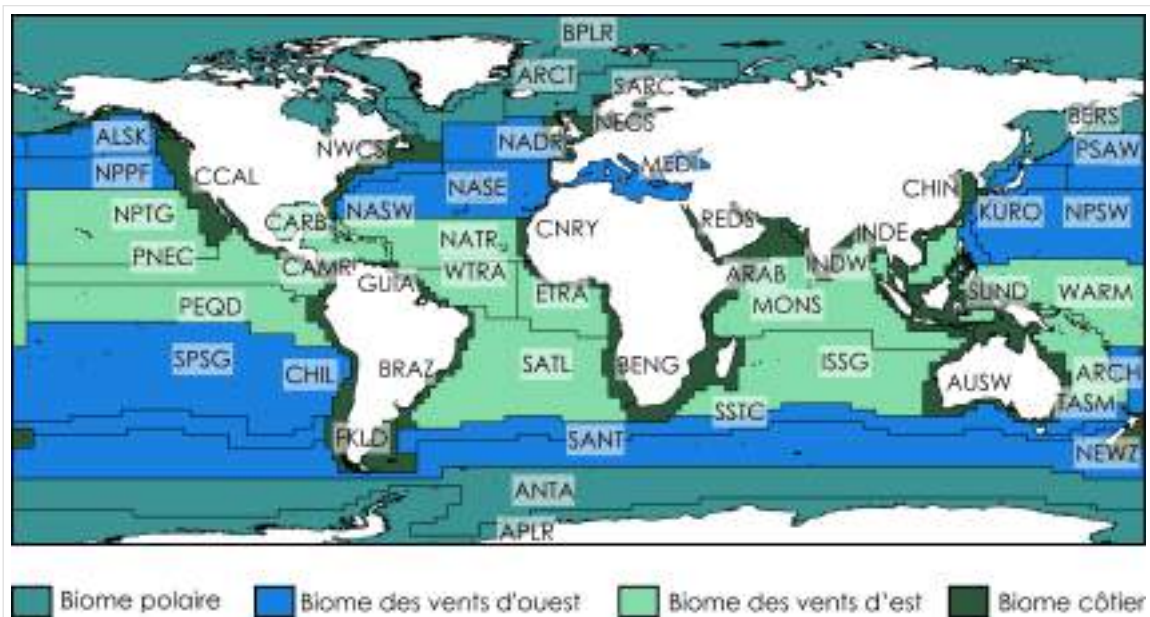


FIGURE 4.2 – Provinces de Longhurst. Les abréviations du nom des provinces peuvent être trouvées en Annexe 9 (page 287), conformément à Longhurst (2010). Les provinces sont groupées en quatre grands biomes : les provinces polaires, les provinces des vents d’Ouest, les provinces des vents d’Est, et les provinces côtières (incluant les plateaux continentaux). Notez que les limites géographiques des provinces peuvent être modifiées au fil du temps (en fonction des saisons ou inter-annuellement, Reygondeau *et al.* (2013)). Ici, les provinces fixées selon (Longhurst, 2010) sont représentées. Les données *in-situ* se trouvant dans les provinces appartenant au biome côtier n’ont pas été utilisées dans cette étude.

de classes), et de définir des critères de classification plus robustes<sup>3</sup>. Dans ce second cas, les critères permettant d’identifier des ressemblances entre les échantillons *in-situ* sont déterminés statistiquement en amont sur un nombre de données plus grand qu’en considérant uniquement les mesures coïncidentes. Ainsi, une validation des associations entre classe de  $Ra(\lambda)$  par satellite et classe de phytoplancton *in-situ* est aisée.

Une classification fine des échantillons *in-situ* disponibles a donc été réalisée (deuxième cas ci-dessus), afin de déterminer des critères de ressemblance entre les échantillons en amont de l’étape d’association avec les classes d’anomalies de luminance. Cette classification a été réalisée sur l’ensemble des données *in-situ* (les 2444 échantillons), afin de prendre en compte toute la diversité des cas phytoplanctoniques disponibles, et de définir des critères à partir d’un grand ensemble de données.

3. En effet, lorsque les critères sont définis *a posteriori* des associations avec les  $Ra(\lambda)$  (premier cas de la liste), ces critères ne concernent que les mesures coïncidentes, et donc un nombre d’observations assez faible si on le compare à l’ensemble des données *in-situ* disponibles, ce qui peut engendrer une mauvaise caractérisation de cette information.



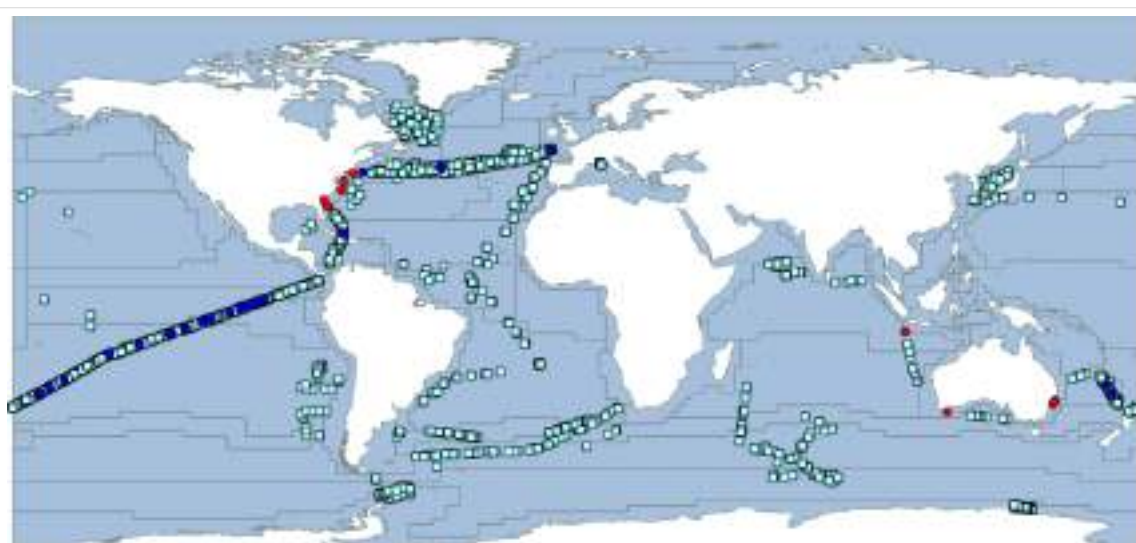


FIGURE 4.3 – Répartition des données *in-situ* utilisées dans cette étude pour définir des profils pigmentaires. Les carrés bleu clair (n=2444 échantillons) sont les échantillons inclus dans l'étude alors que les points ronds (n=41 échantillons) sont les échantillons utilisés lors de la calibration de 2005 (Alvain *et al.*, 2005). Les ronds rouges (n=11 échantillons) n'ont pas été considérés dans cette étude (côtiers selon les critères actuels) et les ronds bleus sont inclus dans la définition des profils pigmentaires (n=30 échantillons).

Cette étape est légèrement différente de celle réalisée dans le cadre de l'étude de Alvain *et al.* (2005). En effet, comme décrit dans le Chapitre 2, Alvain *et al.* (2005) ont utilisé la base de données GEP&CO, qui comportait 176 mesures coïncidentes avec les anomalies de luminance SeaWiFS, et dont 41 échantillons présentaient une dominance selon les critères en Table 4.1 (page 97). Ainsi, sur une base de données de 1502 échantillons à l'époque, issus des campagnes GEP&CO, uniquement 41 échantillons ont servi à la recherche d'un lien empirique entre les données *in-situ* et les données satellites (soit 23 % des mesures coïncidentes (n=176), ou 2,7% du nombre d'échantillons total (n=1502)). Cela représente une sous-exploitation des données *in-situ*, qui peut potentiellement entraîner une connaissance biaisée sur le milieu *in-situ*<sup>4</sup>.

Parmi les 2444 échantillons pour lesquels nous souhaitons définir des profils pigmentaires "types" *via* une classification, 30 échantillons avaient déjà été utilisés lors de la précédente calibration (Alvain *et al.*, 2005), ce qui représente 1,23% des données de cette étude. Ces 30 échantillons sont mis en évidence sur la carte en Figure 4.3 (bleu foncé). Onze échantillons utilisés dans Alvain *et al.* (2005) n'ont pas été utilisés cette fois-ci, car ils ont été considérés comme côtiers, en accord avec la définition des biomes océaniques de Longhurst (2010) (Figure 4.3, échantillons rouges).

4. Ce n'est pas le cas pour Alvain *et al.* (2005, 2008) car la dominance est évaluée sur des critères extrinsèques au jeu de données (i.e., basés sur la littérature), mais qui serait le cas dans cette étude particulière, sans critères pré-définis *a priori*.

Dans cette étude, en utilisant une classification de tous les échantillons disponibles (et pas uniquement les mesures coïncidentes), nous avons, dans un premier temps, cherché les nuances des situations phytoplanctoniques pour des données similaires à celles ayant servi à la calibration des anomalies de luminance en 2005. Pour déterminer ces nuances, nous avons donc besoin d'un maximum d'information afin d'observer les cas récurrents de composition phytoplanctonique au-delà de la dominance. Ensuite, après avoir analysé cette base de données dans son intégralité, les mesures coïncidentes seront effectuées afin de trouver les correspondances (Étape 3, Section 4.1.3, page 104).

#### 4.1.1.3 Analyse des échantillons *in-situ* en fonction de leur composition pigmentaire, au-delà de la dominance

L'intérêt des classifications statistiques lors de la recherche de correspondances réside dans le fait d'augmenter les chances de trouver des paires "gagnantes", c'est-à-dire des paires identifiées de manière récurrente. Ici, nous cherchons à établir une classification détaillée des échantillons *in-situ*, regroupant les échantillons en fonction de leur composition pigmentaire, sans *a priori* sur le fait qu'ils soient ou non associés, à terme, à une mesure satellite (mesures coïncidentes).

Il existe plusieurs types de classifications pouvant répondre à la question suivante : **existe-t-il des structures communes dans les échantillons *in-situ* observables à partir de la concentration relative des pigments biomarqueurs ?**

Nous avons choisi d'utiliser ici aussi une carte auto-organisatrice (Kohonen, 2013), qui a montré un fort potentiel pour la classification des concentrations en pigments (Chazottes *et al.*, 2006), couplé à une classification ascendante hiérarchique comme conseillée dans Niang *et al.* (2003) ou Chazottes *et al.* (2007), permettant de réduire le nombre de groupes issus de la carte auto-organisatrice (soit statistiquement (Chazottes *et al.*, 2007), soit à partir d'une expertise (Niang *et al.*, 2003)). En effet, une carte auto-organisatrice (Kohonen, 2013) peut être considérée comme un outil d'analyse automatique des propriétés statistiques d'un jeu de données. Elle permet une synthèse objective et interprétable (grâce à son aspect visuel et sa notion de voisinage) de données complexes et variables. Contrairement à des méthodes classiques (de type classification ascendante hiérarchique), les classifications de type carte auto-organisatrice permettent de mieux appréhender la contribution des petites similarités (ou dissimilarités). Ainsi, la carte auto-organisatrice forme des classes dont les caractéristiques intra-classes sont très similaires, et les caractéristiques inter-classes sont statistiquement différentes, tout en respectant une notion de "voisinage", c'est-à-dire que deux classes proches dans l'espace de la carte auto-organisatrice sont également proches dans leurs caractéristiques. Les autres méthodes de classification ne permettent pas ou peu d'extraire finement ce type de classes. Par exemple, la distance de similarité (ou de dissimilarité) est évaluée entre les paires d'observations : les faibles et les fortes distances vont

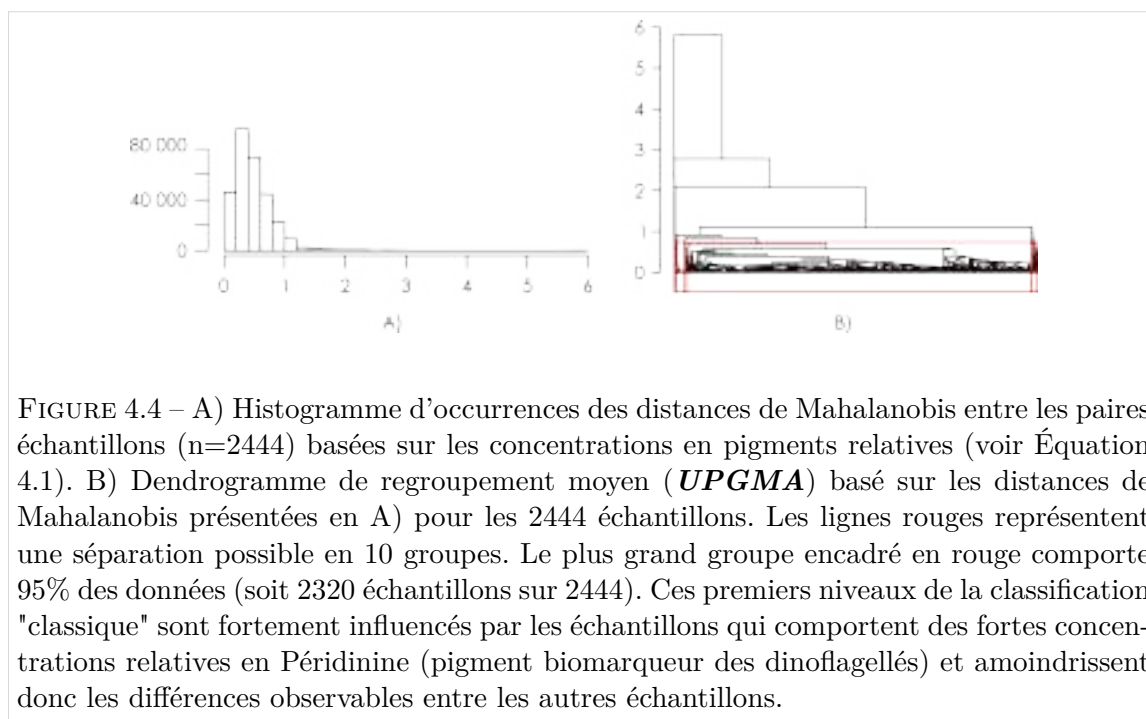


FIGURE 4.4 – A) Histogramme d'occurrences des distances de Mahalanobis entre les paires d'échantillons ( $n=2444$ ) basées sur les concentrations en pigments relatifs (voir Équation 4.1). B) Dendrogramme de regroupement moyen (*UPGMA*) basé sur les distances de Mahalanobis présentées en A) pour les 2444 échantillons. Les lignes rouges représentent une séparation possible en 10 groupes. Le plus grand groupe encadré en rouge comporte 95% des données (soit 2320 échantillons sur 2444). Ces premiers niveaux de la classification "classique" sont fortement influencés par les échantillons qui comportent des fortes concentrations relatives en Périidine (pigment biomarqueur des dinoflagellés) et amoindrissent donc les différences observables entre les autres échantillons.

donc influencer fortement la classification (Figure 4.4). Contrairement à cela, la classification *via* une carte auto-organisatrice nécessite la construction d'un espace de référence à partir des données, où chaque observation permettra d'améliorer cet espace de référence. La distance n'est plus calculée entre les paires d'observations, mais entre une observation et un espace de référence. D'ailleurs, une séparation en 10 groupes potentiels (montré en Figure 4.4, B.) n'est pas évidente au moyen d'une méthode de classification hiérarchique ascendante classique (abrégé CAH). En effet, pour une CAH, le premier groupe formé regroupe environ 95% des données *in-situ*, ce qui apparaît comme un manque de diversité à la vue de la Figure 4.1, qui montre qu'au moins 64,6 % des données présentaient des dominances phytoplanctoniques, et donc une diversité dans les concentrations pigmentaires assez élevée, qui n'apparaît pas avec les outils de regroupement et les distances de similarités ou de dissimilarités classiques. La CAH utilisée sur nos données (Figure 4.4) ne permet, au neuvième niveau de recouplement (10 groupes d'échantillons), que de distinguer que 5% des échantillons forment à eux seuls 9 groupes distincts tant ils sont différents du reste de la base de données.

C'est pour cette raison que nous avons opté pour le choix d'une classification couplée entre une carte auto-organisatrice et une classification hiérarchique *post*-carte auto-organisatrice pour classer les données *in-situ* (d'après Kohonen (2013) et Chazottes *et al.* (2007)). Les analyses des différentes topologies des cartes auto-organisatrices rendent compte d'une meilleure structuration des données pour une carte de 25 classes dans une configuration 5x5, avec une fonction de voisinage gaussienne (Kohonen, 2013). Dans l'étape de construction de la carte auto-organisatrice, le jeu de données est présenté 70 fois (itérations) à la

carte, ce qui permet à chaque fois d'améliorer la représentativité des concentrations de références par rapport aux données initiales. Cela permet de diminuer les erreurs de classifications, mesurées par les indices suivants : les *erreurs de quantification vectorielle*, l'*erreur topologique*, et le *paramètre de distorsion*.

*Nota bene* L'intégration des échantillons déjà utilisés en 2005 n'est pas gênante pour le type d'étude que nous allons réaliser. En effet, ces données ne représentent que 1,23% des données utilisées pour construire la nouvelle classification ; et n'auront donc pas un grand impact sur ce type de classification non linéaire.

### **Résumé : Étape 1, analyse des données *in-situ***

Les données de concentration pigmentaires (pigments biomarqueurs) ont été standardisées par rapport aux concentrations en chlorophylle-*a* et divinyl chlorophylle-*a* (Alvain *et al.*, 2005). Elles sont ensuite classées grâce à une carte auto-organisatrice (Niang *et al.*, 2003; Chazottes *et al.*, 2007; Kohonen, 2013) en 25 classes, permettant d'estimer des concentrations de références pour chaque classe et chaque pigment biomarqueur. Ces concentrations de références sont ensuite elles-mêmes groupées pour réduire le nombre de classes de la carte auto-organisatrice de façon à s'approcher de la diversité écologique recherchée<sup>a</sup>. Dans cette dernière étape, 10 groupes pigmentaires sont obtenus (voir en Résultats).

<sup>a</sup>. En effet, à partir des concentrations de 5 pigments biomarqueurs, on ne s'attend pas à trouver 25 groupes phytoplanctoniques de composition différentes.

Une fois que les données *in-situ* ont été classées en groupes de composition phytoplanctonique similaire d'après leur concentration relative en pigments biomarqueurs, les données *in-situ* sont prêtes à être mises en parallèle avec les données d'anomalies de luminance (Étape 3, mesures coïncidentes, page 104), après avoir pré-traité l'information satellite (Étape 2, ci-dessous).

#### 4.1.2 Données satellites et pré-traitement de l'information satellite : Étape 2

Les données satellites utilisées dans ce Chapitre sont issues du capteur de couleur de l'océan SeaWiFS. SeaWiFS est considéré comme le meilleur capteur de couleur de l'océan, ayant une qualité de radiométrie, et une longue durée d'observation (1997-2010) qui induit de meilleures chances d'obtenir des mesures coïncidentes *in-situ* pour notre étude. Le

capteur de couleur SeaWiFS est, à ce jour, considéré comme une référence en couleur de l'océan (Couto *et al.*, 2016). C'est également avec ce capteur que la méthode PHYSAT a été initialement développée (Alvain *et al.*, 2005). L'utilisation des données SeaWiFS dans cette étude permettra donc des comparaisons avec la méthode initiale.

Cette partie est basée sur le traitement des données satellites pour PHYSAT détaillées aux Chapitres 2 et 3. Brièvement, voici les étapes résumées :

- Les anomalies de luminance  $Ra(\lambda)$  sont calculées à partir des  $nLw(\lambda)$  de SeaWiFS (niveau 3 journalier) et d'une table de référence  $nLw_{ref}(\lambda, Chl-a)$  construite spécifiquement pour le capteur de couleur de l'océan SeaWiFS (Chapitre 2 et Alvain *et al.* (2005)).
- Les anomalies de luminance  $Ra(\lambda)$  sont classées grâce à une carte auto-organisatrice en fonction de leurs formes et amplitudes en 100 classes spectrales (Chapitre 2 et Ben Mustapha *et al.* (2014)).
- Les classes de  $Ra(\lambda)$  provenant de la carte auto-organisatrice sont classées en fonction de leurs saisonnalités de détection (Chapitre 3), *via* des outils dérivés de l'étude de la phénologie.

#### **Résumé : Étape 2, pré-traitement des données satellites**

Les  $Ra(\lambda)$  sont calculés à partir des  $nLw(\lambda)$  de SeaWiFS (niveau 3) et d'une table de référence  $nLw_{ref}(\lambda, Chl-a)$  construite spécifiquement pour le capteur de couleur de l'océan SeaWiFS (Chapitre 2 et Alvain *et al.* (2005)). Les  $Ra(\lambda)$  sont classés grâce à une carte auto-organisatrice en fonction de leurs amplitudes et formes spectrales (Chapitre 2 et Ben Mustapha *et al.* (2014)). Les classes de  $Ra(\lambda)$  provenant de la carte auto-organisatrice sont classées en fonction de leurs saisonnalités de détection (Chapitre 3). Il en résulte, à l'échelle globale, 54 classes d'anomalies ayant des caractéristiques spectrales, spatiales et temporelles spécifiques pour le capteur SeaWiFS.

En parallèle de l'analyse des données *in-situ* (Étape 1, Section précédente), les anomalies de luminance sont calculées puis classées avant de pouvoir être associées aux mesures *in-situ* dans l'étape 3 (mesures coïncidentes, ci-dessous).

### 4.1.3 Établissement du lien empirique entre les données *in-situ* et les anomalies de luminance : Étape 3

Afin d'établir les liens empiriques, des mesures colocalisées (aussi appelées "matchups") sont effectuées entre les données *in-situ* analysées au préalable et les anomalies de luminance PHYSAT. Conformément aux procédures classiques (Werdell et Bailey, 2005), le pixel dont les coordonnées spatiales (ou "mesure coïncidente centrale"), ainsi que le pavé de pixels alentours (ou "mesures coïncidentes étendues") sont considérés pour le jour de la mesure *in-situ* (soit  $J_0$ ), la veille ( $J_{-1}$ ) et le lendemain ( $J_{+1}$ ). Ainsi, 27 pixels composent la mesure coïncidente selon la Figure 4.5.

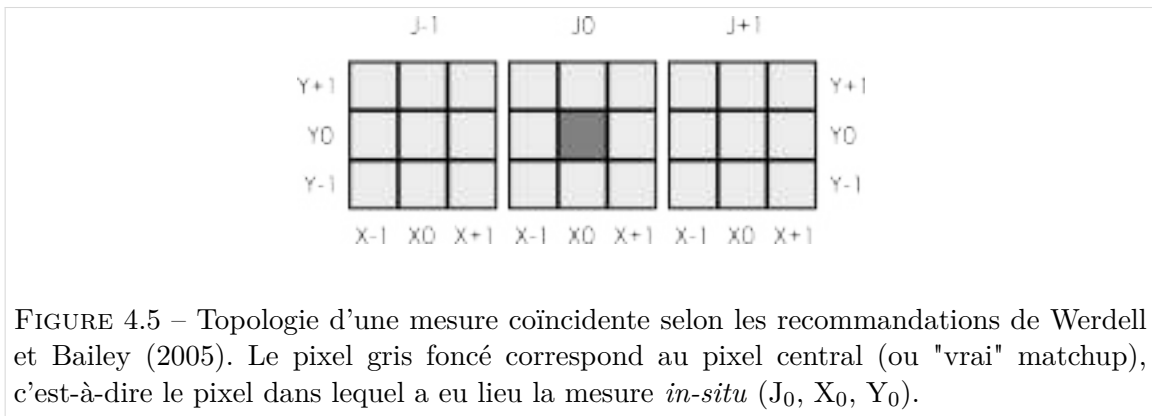


FIGURE 4.5 – Topologie d'une mesure coïncidente selon les recommandations de Werdell et Bailey (2005). Le pixel gris foncé correspond au pixel central (ou "vrai" matchup), c'est-à-dire le pixel dans lequel a eu lieu la mesure *in-situ* ( $J_0, X_0, Y_0$ ).

#### 4.1.3.1 Statistiques sur les matchups

Pour une mesure coïncidente (27 pixels), plusieurs variables sont disponibles :

1. Les  $nLw(\lambda)$  (la luminance) comme information "de base" de la couleur de l'océan
2. Les  $Ra(\lambda)$  calculés qui sont disponibles si  $nLw(\lambda)$  et concentration en chlorophylle-*a* sont disponibles, et correspondent aux critères définis au Chapitre 2
3. Les classes de  $Ra(\lambda)$  *som* qui sont disponibles si le  $Ra(\lambda)$  calculé ci-dessus correspond (en forme et amplitude) à des caractéristiques connues par la carte auto-organisatrice, et est donc affecté à une classe de  $Ra(\lambda)$  (Ben Mustapha *et al.*, 2014)
4. Les classes de  $Ra(\lambda)$  *phénologie* (Chapitre 3) pour lesquelles les caractéristiques temporelles et spatiales sont prises en compte au-delà des caractéristiques spectrales déterminées par les classes de  $Ra(\lambda)$  *som*

*Nota bene* À ce stade de l'étude, nous avons préféré conserver à la fois les classes de

$Ra(\lambda)$  som ainsi que les classes de  $Ra(\lambda)$  phénologie. En effet, le travail effectué dans ce Chapitre permet d'évaluer l'adéquation de ces nouvelles classes de  $Ra(\lambda)$  phénologie avec les groupes phytoplanctoniques, au-delà des caractéristiques inhérentes aux classes de  $Ra(\lambda)$  (i.e., spectres, temporalité, distribution spatiale).

À partir des 2444 échantillons *in-situ*, 907 mesures coïncidentes avec le premier niveau d'information (nLw( $\lambda$ )) ont été trouvées, ce qui représente 37.1% du jeu de données *in-situ* de base. Ce nombre est important puisqu'en moyenne, on observe 10% de mesures coïncidentes pour un jeu de données *in-situ*.

Cependant, le nombre de mesures coïncidentes tombe à 26 matchups lorsque l'on considère la mesure coïncidente centrale pour laquelle l'information "classe de  $Ra(\lambda)$ " est disponible (ce qui fait 1% de la base de données *in-situ* initiale). L'ensemble du nombre de mesures coïncidentes par catégories (deux niveaux de mesures coïncidentes et trois variables) est présenté dans la Table 4.2.

TABLE 4.2 – Information croisée sur les matchups montrant le nombre de mesures coïncidentes disponibles pour le jeu de données *in-situ* croisé avec les données du capteur de couleur SeaWiFS, et ce, pour plusieurs niveaux d'information décrits ci-dessus.

	nLw( $\lambda$ )	Ra( $\lambda$ )	Classe de Ra( $\lambda$ )
Matchup central	113	91	26
Matchup étendu	907	818	624

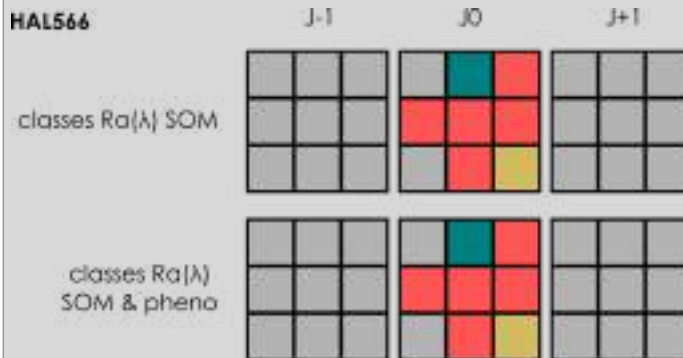
La Table 4.2 montre l'intérêt d'utiliser les matchups "étendus" (i.e., les 27 pixels présentés en Figure 4.5), comme la procédure de Werdell et Bailey (2005) le conseille. En effet, si on considère la mesure coïncidente centrale, il n'y a que 26 mesures coïncidentes disponibles pour la calibration des classes d'anomalies de PHYSAT. Cependant, afin de conserver un signal cohérent, il existe plusieurs critères additionnels à appliquer pour conserver ou non ces mesures coïncidentes "étendues".

Dans les pages suivantes (pages 107 - 109), différents cas de mesures coïncidentes "étendues" à conserver sont présentés au travers d'exemples.

Ainsi, si le matchup central est disponible, et que l'information "classe de  $Ra(\lambda)$ " est disponible, il est également vérifié que les pixels alentours sont homogènes en termes de classes de  $Ra(\lambda)$ , de valeur d'épaisseur optique à 865 nm et de concentration en chlorophylle-*a*. En effet, s'il existe une grande différence au sein de ces valeurs, cela peut laisser sous-entendre un problème de correction atmosphérique, ou un bord de masque (nuages, glint par exemple). Si l'homogénéité des pixels autour du matchup central est vérifiée, l'information "classe de  $Ra(\lambda)$ " est alors conservée.

**Exemple de matchup central pouvant être conservé :** Le point *in-situ* "HAL566" a été échantillonné le 29 avril 2003 par la campagne HALIFAX dans l'océan Atlantique-Nord (latitude 43,40 et longitude -57,47).

Cet échantillon *in-situ* possède un matchup central ainsi que 7 informations au total sur les 27 informations potentielles. Ces 7 informations sont toutes obtenues pour le  $J_0$  (voir Figure ci-dessous). Au total, trois classes de  $Ra(\lambda)$  ont pu être observées pour ce point *in-situ* : les classes SOM n°18<sub>som</sub> (4 pixels), n°26<sub>som</sub> (1 pixel) et n°39<sub>som</sub> (1 pixel), ce qui correspond aux classes "phénologiques" n°12<sub>pheno</sub>, n°7<sub>pheno</sub> et n°6<sub>pheno</sub>. Les pixels gris ne comportent pas d'informations.



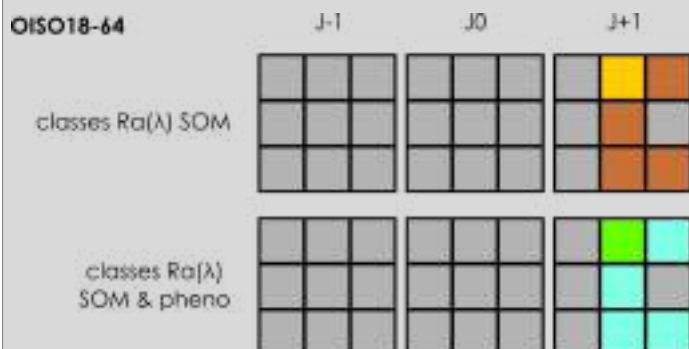
En moyenne, la chlorophylle-*a* SeaWiFS est estimée à  $1,06 \text{ mg.m}^{-3}$ ,  $\pm 0,45$  ( $1,6 \text{ mg.m}^{-3}$  *in-situ*). La moyenne de l'épaisseur optique Tau à 865 nm est quant à elle de  $0,061 \pm 0,004$ . Ce pixel peut ainsi être considéré comme homogène. Le matchup central est conservé tel quel (n°18<sub>som</sub> et n°12<sub>pheno</sub> seront donc considéré pour ce point *in-situ*).



Dans le cas où le matchup central ( $J_0$ ,  $X_0$ ,  $Y_0$ ) n'est pas disponible, où si l'homogénéité des pavés de pixels n'est pas vérifiée, nous nous intéressons alors aux informations disponibles sur le matchup "étendu". Cette fois-ci, nous prenons en compte l'information la plus fréquente pour les classes de  $Ra(\lambda)$ , en essayant d'être le plus proche possible de  $J_0$ .

**Matchups étendus, Cas 1 : la classe la plus fréquente est conservée** Le point *in-situ* "OISO18-64" a été échantillonné le 23 décembre 2009 par la campagne OISO dans le sud de l'Océan Indien (latitude -34,99 et longitude 53,48).

Cet échantillon *in-situ* ne possède pas de matchup central, mais 5 informations étendues au total sur les 27 informations potentielles. Ces 5 informations sont toutes obtenues pour le  $J_{+1}$  (voir Figure ci-dessous). Au total, deux classes de  $Ra(\lambda)$  ont pu être observées pour ce point *in-situ* : les classes SOM n°33<sub>som</sub> (4 pixels) et n°76<sub>som</sub> (1 pixel), ce qui correspond aux classes "phénologiques" n°22<sub>pheno</sub> et n°26<sub>pheno</sub>. Les pixels gris ne comportent pas d'informations.

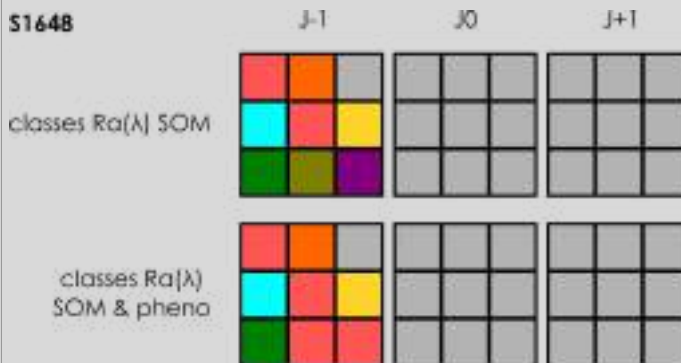


En moyenne, la chlorophylle-*a* SeaWiFS est estimée à  $0,14 \text{ mg.m}^{-3}$ ,  $\pm 0,02$  ( $0,094 \text{ mg.m}^{-3}$  *in-situ*). La moyenne de l'épaisseur optique Tau à 865 nm est quant à elle de  $0,10 \pm 0,005$ . Dans ce cas, il paraît logique de conserver pour ce point *in-situ* les informations n°33<sub>som</sub> et n°22<sub>pheno</sub>.

Dans certains cas, l'homogénéité du pavé de pixel n'est pas vérifiée sur les classes SOM de  $Ra(\lambda)$ . Il y a alors deux solutions : soit il y a homogénéité sur classes phénologiques (on conserve alors l'information phénologique), soit non (et on ne conserve pas le matchup).

**Matchups étendus, Cas 2 : la classe phénologique la plus fréquente est conservée** Le point *in-situ* "Q1648" a été échantillonné le 20 avril 2002 (base de données NOMAD) dans l'océan Atlantique (latitude 32,59 et longitude -15,94).

Cet échantillon *in-situ* ne possède pas de vrai matchup, mais 8 informations étendues au total sur les 27 informations potentielles. Ces 8 informations sont toutes obtenues pour le  $J_{-1}$  (voir Figure ci-dessous). Au total, sept classes de  $Ra(\lambda)$  ont pu être observées pour ce point *in-situ* : les classes SOM n°67<sub>som</sub> (2 pixels), puis un pixel pour chacune de ces classes de  $Ra(\lambda)$  : n°25<sub>som</sub>, n°33<sub>som</sub>, n°49<sub>som</sub>, n°58<sub>som</sub>, n°70<sub>som</sub>, n°78<sub>som</sub> ; ce qui correspond aux classes "phénologiques" n°17<sub>pheno</sub> (4 pixels), n°22<sub>pheno</sub> (2 pixels), puis un pixel des classes *phéno* suivantes : n°34<sub>pheno</sub> et n°43<sub>pheno</sub>. Les pixels gris ne comportent pas d'informations.



En moyenne, la chlorophylle-*a* SeaWiFS est estimée à  $0,19 \text{ mg.m}^{-3}$ ,  $\pm 0,03$  ( $0,12 \text{ mg.m}^{-3}$  *in-situ*). La moyenne de l'épaisseur optique Tau à 865 nm est quant à elle de  $0,024 \pm 0,003$ . Dans cette situation, les classes de  $Ra(\lambda)$  issues de l'algorithme SOM ne permettent pas de trancher sur l'information la plus pertinente. En revanche, les classes de  $Ra(\lambda)$  issues de la classification *via* la phénologie montre tout de même que 4 pixels sur les 8 informations disponibles comportent la même classe de  $Ra(\lambda)_{pheno}$ . Ici, la classe n°17<sub>pheno</sub> est donc conservée.

À partir de l'exercice de matchups, complété par les sélections présentées ci-dessus, 123 mesures colocalisées considérées comme fiables ont pu être trouvées, soit directement à  $J_0$ ,  $X_0$ ,  $Y_0$  (26 points), soit indirectement *via* les informations satellites les plus fréquentes (97 points). En 2005, l'algorithme PHYSAT a été développé à partir de 176 mesures colocalisées, mais uniquement 41 mesures présentaient une dominance, et étaient donc utilisables pour la calibration initiale, qui visait à montrer la présence de signaux pour les cas de dominance uniquement (*via* les enveloppes de  $Ra(\lambda)$  réalisées intuitivement comme expliqué au Chapitre 2). Ainsi, nous disposons de trois fois plus de mesures qu'en 2005 pour tester la calibration au-delà des cas de dominance.

*Nota bene* Au sein des données de mesures coïncidentes, il n'a pas été rencontré de données *in-situ* à la fois proches dans le temps et dans l'espace, ce qui évite une éventuelle autocorrélation spatiale ou temporelle. En effet, aucune mesure coïncidente ne provient du même fichier journalier SeaWiFS. Ainsi, une association entre données *in-situ* et données satellites ne pourra pas être due au fait que l'ensemble des mesures coïncidentes ont été collectée en même temps (dans une même masse d'eau par exemple), et les chances que les associations potentielles soient dues "au hasard" diminuent.

#### 4.1.3.2 Établissement du lien empirique entre classes de $Ra(\lambda)$ et information *in-situ* : étape 3

##### Procédure de calibration classique

Les 123 mesures coïncidentes sont séparées aléatoirement en deux parties : 80% des mesures servent à la calibration (étape pour trouver les liens empiriques) et 20% des mesures serviront à la validation. Cette étape est classiquement utilisée dans une procédure visant à établir et valider des liens empiriques.

À partir du jeu de données de calibration, une classe de  $Ra(\lambda)$  spécifique est considérée "labellisée" (i.e., un nom lui est attribué) si au moins 50% des informations *in-situ* correspondent à une composition *in-situ* similaire (définie en Section 4.1.1). Ce seuil de 50% est classique dans la procédure PHYSAT (Chapitre 2 et Alvain *et al.* (2005)). Pour une classe de  $Ra(\lambda)$ , le nombre de mesures coïncidentes doit être au moins de 5 (4 échantillons pour la calibration et 1 échantillon pour la validation).

Après avoir réalisé l'étape de calibration, une validation est réalisée grâce au set de validation. Pour cela, les classes de  $Ra(\lambda)$  labellisées dans l'étape de calibration seront recherchées dans les données de validation. Idéalement, toutes les occurrences d'une classe de  $Ra(\lambda)$  dans les données de validation devraient être associées au même groupe phyto-planctonique que celui associé dans l'étape de calibration, donnant un score de validation de 100%.

**Répétition de la procédure de calibration et validation**

Nous pourrions considérer la méthode terminée à ce stade (approches utilisées dans les précédentes versions Alvain *et al.* (2005), Ben Mustapha *et al.* (2014)). Cependant, afin d’apprécier la variabilité qui peut exister dans les données colocalisées, cette procédure validation et calibration est répétée 400 fois. En effet, il existe une variabilité dans les mesures coïncidentes qui ne sont pas toujours réparties équitablement au niveau spatio-temporel<sup>5</sup>, ce qui est d’autant plus vrai pour les grands jeu de données de mesures coïncidentes. Le nombre de procédure,  $n=400$ , permet que chaque mesure coïncidente se trouve au moins une fois dans les données de validation.


De plus, une répétition de la procédure permet d’évaluer la robustesse de la calibration *via* le calcul du taux de calibration pour les classes de  $Ra(\lambda)$  (nombre de fois où une classe est labellisée pour chaque assemblage phytoplanctonique).

**Les classes de  $Ra(\lambda)$  utilisées pour la calibration**

Deux jeux de données de classes de  $Ra(\lambda)$  ont été testés ici : les classes de  $Ra(\lambda)$  *som*, issues de la carte auto-organisatrice (Ben Mustapha *et al.*, 2014), et les classes de  $Ra(\lambda)$  *phénologie*, issues du regroupement des classes de  $Ra(\lambda)$  *som* *via* les outils de la phénologie (et qui ont donc les aspects temporels, spectraux et spatiaux similaires, voir Chapitre 3). Cela permet de comparer la précédente calibration de PHYSAT (Alvain *et al.*, 2008; Ben Mustapha *et al.*, 2014) avec cette nouvelle calibration, mais aussi d’étudier l’impact des deux changements majeurs issus de cette thèse (regroupement des classes de  $Ra(\lambda)$  *via* la phénologie, et utilisation de données *in-situ* non dominantes en termes de groupes phytoplanctoniques), puis d’évaluer l’ensemble des nouveautés.

---

5. Par exemple, certaines mesures de l’équateur ont plus de chance d’obtenir des mesures coïncidentes que les mesures aux pôles, ce qui est facilement explicable par les couvertures nuageuses ou l’angle solaire.

 **Résumé : Recherche du lien empirique entre les classes de  $Ra(\lambda)$  et les informations *in-situ***

Les mesures coïncidentes sont réalisées entre les données *in-situ*, regroupées en 10 classes pigmentaires, et les données satellites, regroupées en classes (spectrales et distributions spatio-temporelle). Deux types de liens empiriques sont recherchés entre les données *in-situ* et i) les classes de  $Ra(\lambda)$  issues de la carte auto-organisatrice et ii) les classes de  $Ra(\lambda)$  *som* regroupées grâce à leur phénologie et présentant donc des caractéristiques spatiales, spectrales et temporelles similaires. À partir de 80% des mesures coïncidentes, une classe de  $Ra(\lambda)$  (*som* ou *phénologie*) est considérée labellisée si au moins 50% des informations *in-situ* correspondent à une composition *in-situ* similaire. Les 20% de données restantes servent à la validation. Cette procédure est répétée pour évaluer le taux de calibration et les scores de validation en fonction des séparations aléatoires des données de calibration et de validation.

## 4.2 Résultats

### 4.2.1 Les assemblages phytoplanctoniques déterminés *in-situ*

La Figure 4.6 montre les concentrations pigmentaires des points *in-situ* classés par une carte auto-organisatrice, puis groupés par une classification hiérarchique (abrégée *CAH* par la suite). Nous remarquons que les classes adjacentes, qu'elles soient définies par la carte auto-organisatrice ou par la *CAH*, sont proches en termes de concentrations pigmentaires. Au sein des classes de la carte auto-organisatrice, les concentrations en pigments ont une variabilité assez faible (Table 4.3), ce qui montre le potentiel d'un algorithme de classification de type carte auto-organisatrice (Kohonen, 2013).

La carte auto-organisatrice a été effectuée 500 fois. En effet, en fonction du sous-ensemble de données présentées à la carte, celle-ci peut différer dans sa structure (i.e., les classes ne sont pas toujours au même endroit sur la carte, car l'espace de référence initial dépend des données présentées dans la phase d'apprentissage). Dans cette optique, une mesure de stabilité des classes a été calculée : c'est le pourcentage de cas pour lesquels les données *in-situ* se voient groupées dans une même classe, peu importe la topologie de la carte auto-organisatrice. Au total, 66% des données *in-situ* se retrouvaient toujours affectées ensemble dans la même classe pigmentaire. Il a été remarqué que l'instabilité est principalement liée aux classes n°6, 11, 12, 13, 16, 17, pour lesquelles les points *in-situ* peuvent être parfois dans une de ces classes et parfois dans une autre (couleur bleu claire en Figure 4.6).

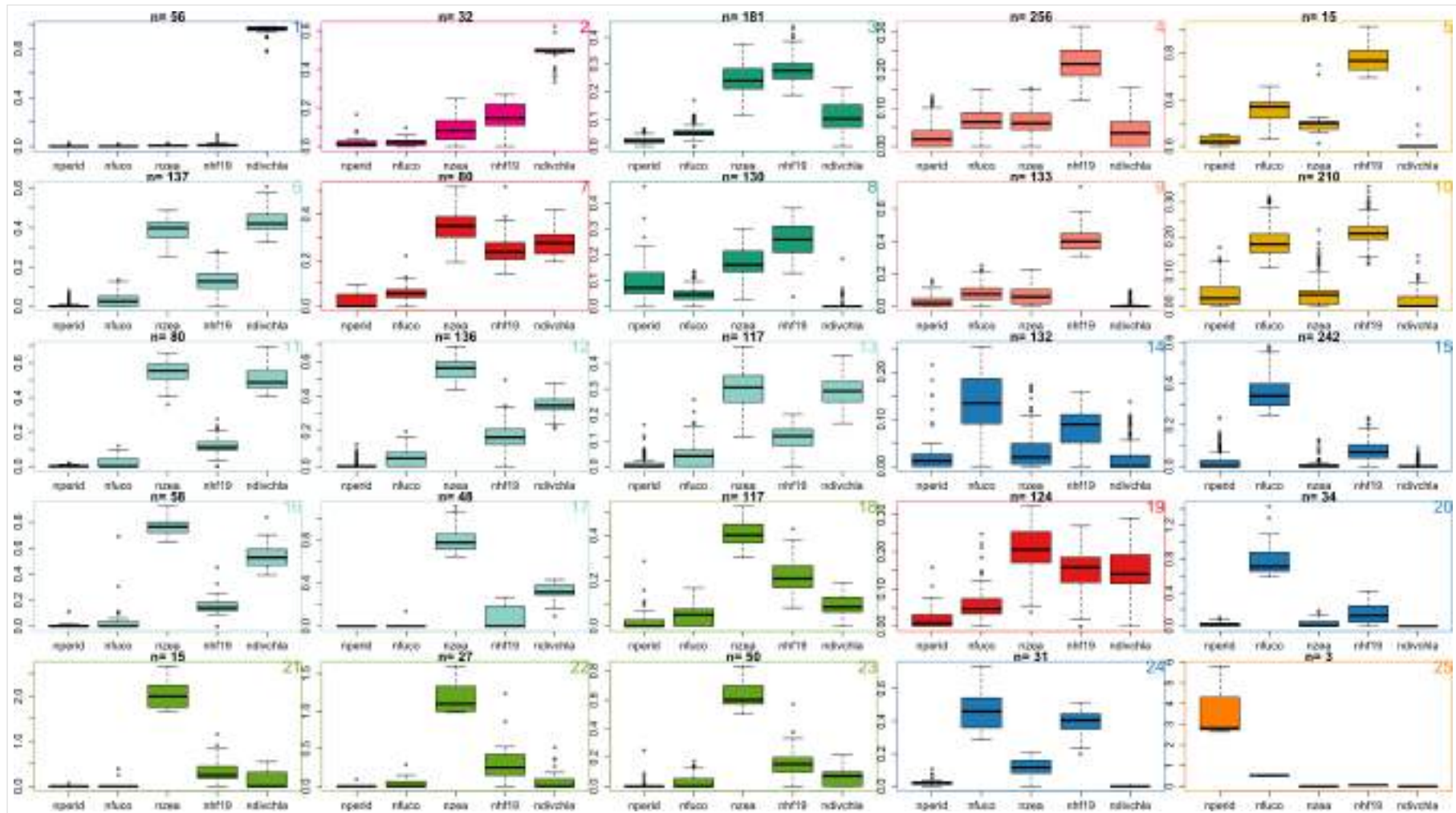


FIGURE 4.6 – Concentrations relatives *in-situ* des cinq pigments biomarqueurs au sein des 25 classes déterminées grâce à un algorithme de classification de type carte auto-organisatrice. Les couleurs correspondent aux groupes formés d'après la classification hiérarchique sur les concentrations de références de chaque classe. Le nombre "n" au dessus de chaque graphique est le nombre d'observations *in-situ* assignées à cette classe.

TABLE 4.3 – Concentrations relatives moyennes des cinq pigments biomarqueurs utilisés dans cette étude au sein des 25 classes pigmentaires définies par la carte auto-organisatrice réalisée à partir des données *in-situ* (n=2444), ainsi que les écarts-types des distributions de ces pigments entre parenthèses.

Groupe <i>SOM</i> pigments	Péridine	Fucoxanthine	Zeaxanthine	HF19	Divinyl-Chla
1	0,002 ( $\pm 0,005$ )	0,002 ( $\pm 0,004$ )	0,007 ( $\pm 0,005$ )	0,014 ( $\pm 0,019$ )	0,955 ( $\pm 0,041$ )
2	0,02 ( $\pm 0,033$ )	0,021 ( $\pm 0,02$ )	0,091 ( $\pm 0,072$ )	0,157 ( $\pm 0,069$ )	0,482 ( $\pm 0,06$ )
3	0,019 ( $\pm 0,015$ )	0,051 ( $\pm 0,024$ )	0,243 ( $\pm 0,051$ )	0,279 ( $\pm 0,043$ )	0,104 ( $\pm 0,052$ )
4	0,025 ( $\pm 0,029$ )	0,066 ( $\pm 0,029$ )	0,064 ( $\pm 0,035$ )	0,219 ( $\pm 0,044$ )	0,037 ( $\pm 0,036$ )
5	0,048 ( $\pm 0,04$ )	0,32 ( $\pm 0,112$ )	0,235 ( $\pm 0,18$ )	0,757 ( $\pm 0,135$ )	0,052 ( $\pm 0,133$ )
6	0,006 ( $\pm 0,015$ )	0,028 ( $\pm 0,032$ )	0,39 ( $\pm 0,056$ )	0,129 ( $\pm 0,058$ )	0,428 ( $\pm 0,055$ )
7	0,025 ( $\pm 0,03$ )	0,056 ( $\pm 0,035$ )	0,348 ( $\pm 0,065$ )	0,245 ( $\pm 0,057$ )	0,276 ( $\pm 0,055$ )
8	0,095 ( $\pm 0,075$ )	0,044 ( $\pm 0,03$ )	0,172 ( $\pm 0,056$ )	0,259 ( $\pm 0,06$ )	0,006 ( $\pm 0,021$ )
9	0,031 ( $\pm 0,035$ )	0,085 ( $\pm 0,052$ )	0,067 ( $\pm 0,055$ )	0,404 ( $\pm 0,067$ )	0,009 ( $\pm 0,021$ )
10	0,035 ( $\pm 0,038$ )	0,187 ( $\pm 0,045$ )	0,036 ( $\pm 0,036$ )	0,213 ( $\pm 0,038$ )	0,015 ( $\pm 0,023$ )
11	0,003 ( $\pm 0,006$ )	0,022 ( $\pm 0,03$ )	0,547 ( $\pm 0,064$ )	0,115 ( $\pm 0,052$ )	0,507 ( $\pm 0,071$ )
12	0,009 ( $\pm 0,022$ )	0,046 ( $\pm 0,046$ )	0,558 ( $\pm 0,057$ )	0,169 ( $\pm 0,076$ )	0,35 ( $\pm 0,047$ )
13	0,013 ( $\pm 0,027$ )	0,045 ( $\pm 0,047$ )	0,3 ( $\pm 0,07$ )	0,105 ( $\pm 0,054$ )	0,291 ( $\pm 0,056$ )
14	0,02 ( $\pm 0,034$ )	0,137 ( $\pm 0,064$ )	0,035 ( $\pm 0,042$ )	0,08 ( $\pm 0,042$ )	0,018 ( $\pm 0,032$ )
15	0,025 ( $\pm 0,042$ )	0,355 ( $\pm 0,08$ )	0,007 ( $\pm 0,017$ )	0,075 ( $\pm 0,047$ )	0,006 ( $\pm 0,015$ )
16	0,007 ( $\pm 0,021$ )	0,036 ( $\pm 0,099$ )	0,766 ( $\pm 0,064$ )	0,156 ( $\pm 0,068$ )	0,541 ( $\pm 0,085$ )
17	0 ( $\pm 0$ )	0,003 ( $\pm 0,02$ )	0,799 ( $\pm 0,113$ )	0,094 ( $\pm 0,104$ )	0,326 ( $\pm 0,067$ )
18	0,019 ( $\pm 0,035$ )	0,053 ( $\pm 0,043$ )	0,406 ( $\pm 0,054$ )	0,214 ( $\pm 0,065$ )	0,09 ( $\pm 0,054$ )
19	0,018 ( $\pm 0,024$ )	0,055 ( $\pm 0,046$ )	0,208 ( $\pm 0,057$ )	0,15 ( $\pm 0,052$ )	0,151 ( $\pm 0,06$ )
20	0,026 ( $\pm 0,027$ )	0,796 ( $\pm 0,196$ )	0,038 ( $\pm 0,053$ )	0,145 ( $\pm 0,118$ )	0 ( $\pm 0$ )
21	0,004 ( $\pm 0,017$ )	0,068 ( $\pm 0,144$ )	2,063 ( $\pm 0,312$ )	0,36 ( $\pm 0,35$ )	0,123 ( $\pm 0,188$ )
22	0,003 ( $\pm 0,017$ )	0,039 ( $\pm 0,072$ )	1,194 ( $\pm 0,2$ )	0,295 ( $\pm 0,278$ )	0,087 ( $\pm 0,136$ )
23	0,016 ( $\pm 0,052$ )	0,028 ( $\pm 0,049$ )	0,631 ( $\pm 0,089$ )	0,155 ( $\pm 0,105$ )	0,067 ( $\pm 0,064$ )
24	0,025 ( $\pm 0,023$ )	0,468 ( $\pm 0,128$ )	0,117 ( $\pm 0,058$ )	0,392 ( $\pm 0,077$ )	0 ( $\pm 0$ )
25	3,743 ( $\pm 1,778$ )	0,518 ( $\pm 0,03$ )	0 ( $\pm 0$ )	0,049 ( $\pm 0,013$ )	0 ( $\pm 0$ )

---

Les résultats du regroupement des classes de la carte auto-organisatrice *via* une CAH sont présentés en Figure 4.7. Les concentrations relatives des échantillons *in-situ* des dix classes pigmentaires sont un peu plus variables que dans le cas des classes de la carte auto-organisatrice, ce qui était attendu. En revanche, les profils pigmentaires des inter-classes sont différents : chacune des classes possède sa spécificité en termes de concentrations relatives en pigments biomarqueurs.

Les cartes de localisation des échantillons *in-situ* sont présentées en Figure 4.8. Certaines classes pigmentaires ont une localisation qui semble être spécifique à la vue des données disponibles (par exemple la classe n°5, en marge continentale), alors que d'autres semblent plus ubiquistes (par exemple la classe n°6, en océans ouverts).

Le problème d'instabilité des groupes observé sur la carte auto-organisatrice est résolu par la CAH, où cette fois-ci 96% des données se retrouvent toujours ensemble dans la même classe. L'instabilité des classes CAH est elle liée aux classes CAH n°5 et 8, où quelques échantillons sont parfois dans l'une ou l'autre de ces deux classes en fonction de différents essais de classifications, ce qui montre l'intérêt d'utiliser la classification en deux étapes (couplage carte auto-organisatrice et CAH), mais également de répéter ces procédures, afin d'en évaluer la qualité.



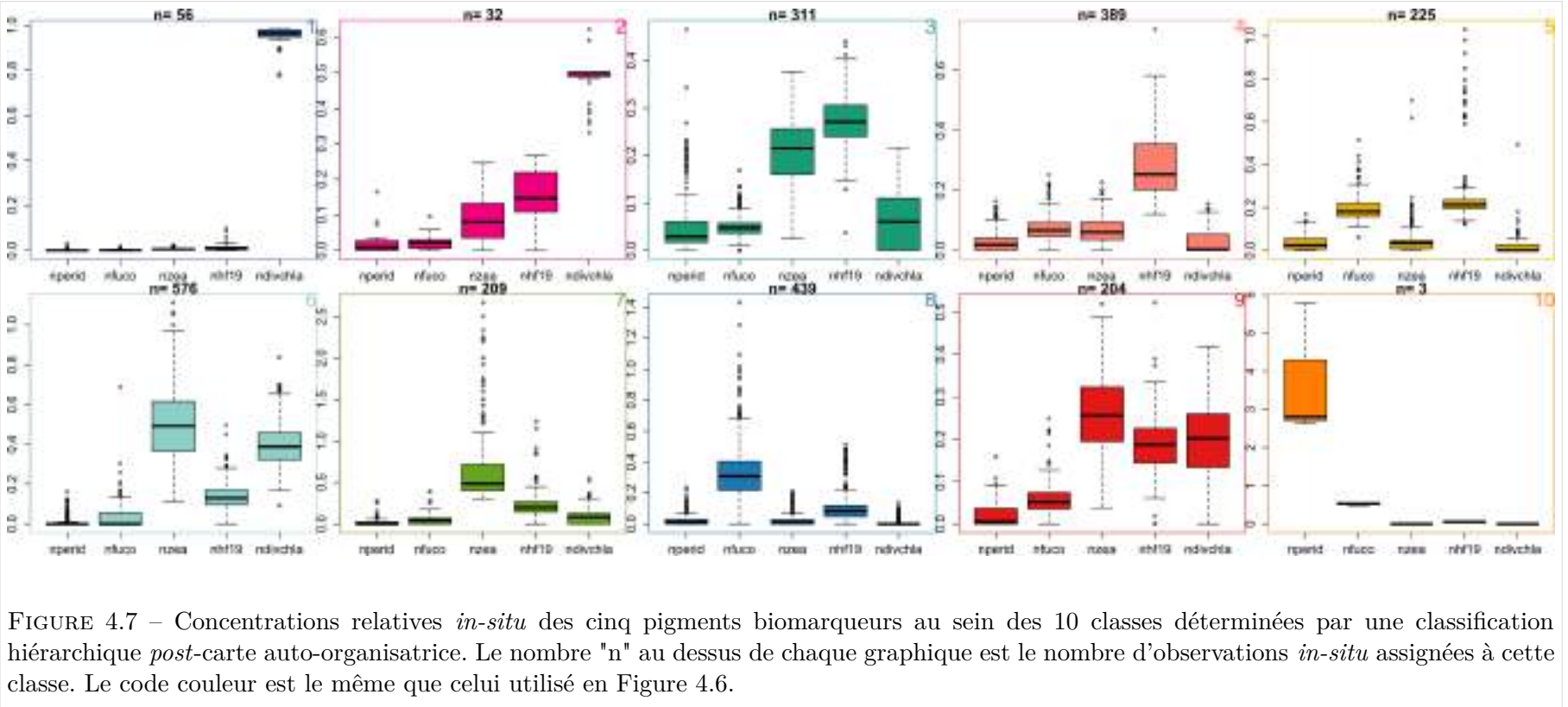


FIGURE 4.7 – Concentrations relatives *in-situ* des cinq pigments biomarqueurs au sein des 10 classes déterminées par une classification hiérarchique *post-carte* auto-organisatrice. Le nombre "n" au dessus de chaque graphique est le nombre d'observations *in-situ* assignées à cette classe. Le code couleur est le même que celui utilisé en Figure 4.6.

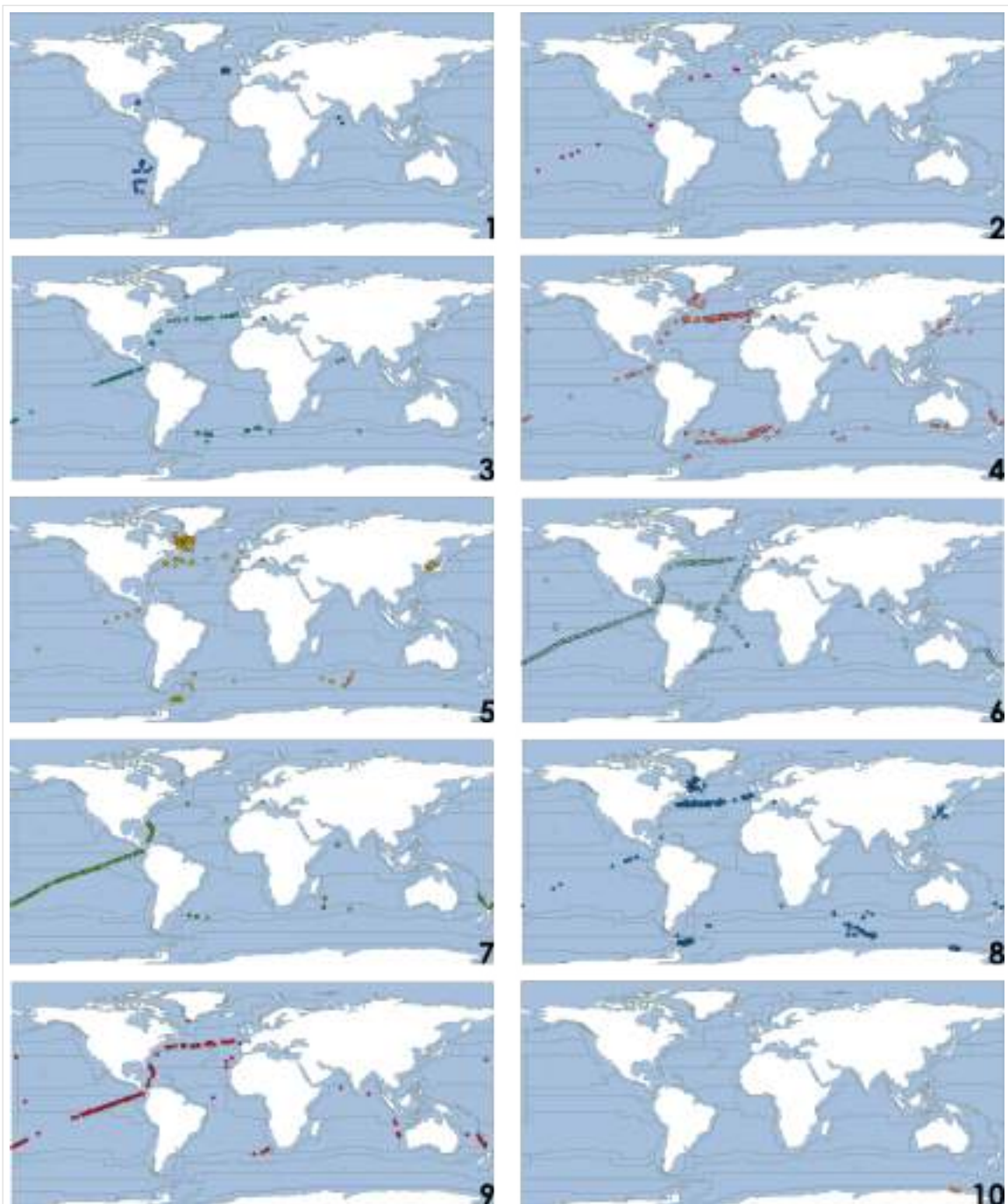


FIGURE 4.8 – Cartes de localisation des points *in-situ* appartenant aux groupes pigmentaires issus de la CAH post-carte auto-organisatrice. Les numéros des groupes, en bas à droite de chaque carte, correspondent aux mêmes numéros que ceux de la Figure 4.7, où figure également le nombre d'échantillons.

Une fois que les échantillons *in-situ* ont été classés, il est possible de tenter de leur donner "un nom". Les noms donnés ensuite aux 10 assemblages sont issus de recherches bibliographiques sur les pigments (Jeffrey *et al.*, 1997), mais également de l'interprétation des résultats par les auteurs des campagnes *in-situ*, ou de données complémentaires *in-situ* réalisées par les auteurs des campagnes (comptages, observations sur le terrain), e.g., Boyd *et al.* (1999), Gall *et al.* (2001), Dandonneau *et al.* (2004), Claustre *et al.* (2004), Peloquin *et al.* (2013).

#### **Classe 1 : les *Prochlorococcus* dominants**

La classe 1 possède une forte concentration en Divinyl-Chlorophyll-*a*, et une très faible concentration de tous les autres pigments biomarqueurs. La Divinyl-Chlorophyll-*a* est biomarqueur des Prochlorophytes du type *Prochlorococcus spp.* (Jeffrey *et al.*, 1997), qui ici, d'après les seuils définis en Table 4.1 (page 97), sont dominants.

#### **Classe 2 : assemblage de *Prochlorococcus* non dominants**

La classe 2 possède également une assez forte concentration en Divinyl-Chlorophyll-*a* (bien que plus faible que la classe 1), mais cette fois-ci avec une proportion non négligeable de 19'-HexanoyloxyFucoxanthine (HF19) (biomarqueur des nanoeucaryotes). Les seuils sur les concentrations relatives de ce pigment n'atteignent pas la dominance. Ici, d'après également les informations fournies sur les bases de données, nous aurions plutôt affaire à des *Prochlorococcus* non dominants, probablement mélangés à d'autres cellules de type nanoeucaryotes.

#### **Classe 3 : assemblage de nanoeucaryotes et *Synechococcus spp.***

La classe 3 est représentée par des concentrations en Zéaxanthine et HF19 suggérant la co-dominance des nanoeucaryotes et des *Synechococcus* selon les seuils définis en Table 4.1 (page 97).

#### **Classe 4 : Nanoeucaryotes dominants**

Pour la classe 4, la concentration en HF19 dépassant le seuil de dominance, ainsi que les faibles concentrations relatives des autres pigments biomarqueurs suggèrent que ces échantillons sont dominés par des cellules de type nanoeucaryotes.

#### **Classe 5 : assemblage de diatomées et de nanoeucaryotes**

Dans la classe 5, les seuils de dominance sur les concentrations relatives en pigments biomarqueurs sont dépassés pour la Fucoxanthine et le HF19, ce qui suggère une cohabitation de diatomées et nanoeucaryotes.

**Classe 6 : assemblage de *Synechococcus* et *Prochlorococcus***

Les échantillons qui ont été assignés à la classe 6 montre une plus grande quantité de Zéaxanthine et de Divinyl-Chlorophyll-*a*. Quelques points *in-situ* déplacent les distributions de Fucoxanthine et de HF19 vers le haut, mais la plupart des échantillons (91%) ont des concentrations de ces biomarqueurs très faibles (respectivement de 0,013 et 0,11 mg.m<sup>-3</sup> de moyenne pour ces 91%). Dans cette classe, il semble que certains échantillons "perturbent" la variabilité observée dans les concentrations (mais n'est pas étonnant si l'on considère le grand nombre d'échantillons qui compose cette classe). Ceci est gardé en mémoire pour l'étape suivante (Étape 3, calibration des Ra( $\lambda$ )).

**Classe 7 : *Synechococcus spp.* dominants**

Les échantillons de la classe 7 présentent une forte concentration en Zéaxanthine, dépassant les seuils de dominance, ce qui suggère que ces échantillons sont dominés par les *Synechococcus*.

**Classe 8 : diatomées dominantes**

Les échantillons de la classe 7 présentent une forte concentration en Fucoxanthine, dépassant les seuils de dominance, ce qui suggère que ces échantillons sont dominés par les diatomées.

**Classe 9 : assemblage "nanoeucaryotes, *Synechococcus* et *Prochlorococcus*"**

Les concentrations en Zéaxanthine, HF19 et Divinyl-Chlorophyll-*a* de ces échantillons suggèrent que ces échantillons *in-situ* sont composés d'un mélange de nanoeucaryotes, *Synechococcus spp.* et *Prochlorococcus spp.*

**Classe 10 : les dinoflagellés dominants**

La concentration relative très importante en Péridinine montre que les 3 échantillons assignés à cette classe sont caractérisés par une dominance de dinoflagellés.

 **Résumé : Les assemblages déterminés *in-situ***

Les 2444 échantillons *in-situ* ont permis de caractériser dix groupes phytoplanctoniques à partir de leurs signatures pigmentaires, dont certains sont des assemblages (i.e., comportent plusieurs groupes phytoplanctoniques) et d'autres présentent des dominances (i.e., comporte un groupe phytoplanctonique principal). Ces dix situations ont été déterminées à partir d'un couplage de classifications de type carte auto-organisatrice suivie d'une classification ascendante hiérarchique, et ils sont stables (i.e., les échantillons sont fidèles à un groupe d'échantillons).

Les 2444 échantillons *in-situ* étudiés dans l'étape 1 ont permis de définir les groupes phytoplanctoniques dont on va pouvoir rechercher un signal éventuel d'anomalie de luminance spécifique. Pour cela, les mesures coïncidentes ont été réalisées, et le lien potentiel entre ces groupes phytoplanctoniques *in-situ* et les anomalies de luminance a été recherché.

#### 4.2.2 Lien empirique entre les données *in-situ* et les anomalies de luminance

Après avoir identifié les situations phytoplanctoniques *in-situ*, la recherche des mesures coïncidentes a été effectuée avec les données satellites. Pour rappel, 123 matchups ont été trouvés entre les données satellites et les données *in-situ*. Dans ces mesures coïncidentes, uniquement 45 classes de  $Ra(\lambda)$  *som* (représentant 29 classes de  $Ra(\lambda)$  *phénologie* d'après la classification de la phénologie) étaient présentes pour ces données *in-situ* (de 1 à 10 occurrences).

*Nota bene* Nous rappelons que les données d'anomalies de luminance de SeaWiFS ont déjà fait l'objet d'une calibration dont les résultats sont publiés dans Alvain et al. (2008) et Ben Mustapha et al. (2014). Pour cette dernière version, au total, 74 classes de  $Ra(\lambda)$  avaient été associées à des groupes dominants (voir la Figure 3.1) par projection de la précédente **labellisation** (Alvain et al., 2008) sur les  $Ra(\lambda)$  classés par une carte auto-organisatrice (Ben Mustapha et al., 2014), et 26 n'avaient pas pu être associés à des groupes phytoplanctoniques (explications au Chapitre 2). Le fait que les classes de  $Ra(\lambda)$  dans les mesures coïncidentes ne représentent que 45% de la diversité des classes de  $Ra(\lambda)$  pose un problème. En effet, nous savons à l'avance que, peu importe le nombre d'association que l'on pourra trouver au sein des mesures coïncidentes, nous ne pourrions labelliser au maximum que 45% de la base de données de  $Ra(\lambda)$ . Bien que la couverture spatiale de la nouvelle version de PHYSAT sera ainsi diminuée, cette approche est plus rigoureuse que la projection des groupes phytoplanctoniques de Alvain et al. (2008) sur la carte auto-organisatrice (Ben Mustapha et al., 2014), et permettra également d'effectuer une validation.

Ainsi, les résultats présentés dans les parties suivantes ("variabilité de la calibration" et "validation") ne sont portés que sur les 32 classes d'anomalies pour lesquelles les mesures coïncidentes étaient suffisantes pour obtenir une nouvelle labellisation (ainsi, 13 classes de  $Ra(\lambda)$  ne comportaient pas assez de matchups sur les 45).

Dans cette partie (pages 120 à 126), nous avons présenté les résultats classe de  $Ra(\lambda)$  par classe de  $Ra(\lambda)$ . Pour la suite des résultats ("Distribution spatiale" et "Spectres de  $Ra(\lambda)$ "), le couplage des labellisations est présenté, puisque les résultats seront donnés par situation phytoplanctonique (donc pour l'ensemble des classes de  $Ra(\lambda)$  associées à ces cas).

**Les nouvelles labellisations par rapport à Alvain *et al.* (2008); Ben Mustapha *et al.* (2014) : 3 classes de Ra( $\lambda$ ) *som***

Selon les critères de labellisation, trois classes de Ra( $\lambda$ ) ont été associées à un groupe phytoplanctonique alors qu'elles ne l'avaient jamais été auparavant. Il s'agit des classes de Ra( $\lambda$ ) SeaWiFS n°48, 74 et 94 issues de la carte auto-organisatrice des Ra( $\lambda$ ), et pour lesquelles les classes 74 et 94 font parties de la même classe de Ra( $\lambda$ ) *phénologie*.

La Table 4.4 regroupe les informations de labellisation concernant ces trois classes de Ra( $\lambda$ ). Pour la classe n°*som* 48, 100% des 400 procédures distinctes de calibration ont donné le même label gagnant : le groupe *Prochlorococcus* non dominants. Dans les données de validation, on trouve en moyenne 72,33% du temps ce label (sur les 6 matchups, 1 correspond à un mélange de diatomées et de nanoeucaryotes). Ainsi, la classe de Ra( $\lambda$ ) *som* n°48 est confondue dans 27,67% des identifications avec l'assemblage diatomées et nanoeucaryotes.

Les classes de Ra( $\lambda$ ) *som* n°74 et 94 ne comportent pas assez de matchups à elles seules pour être labellisées puisque le minimum est fixé à 5 matchups. Ici, elles ont donc été labellisées grâce à leur ressemblance en termes de phénologie, de distribution spatiale et de caractéristiques spectrales (elles représentent la classe *phénologie* n°41). En couplant ces deux classes de Ra( $\lambda$ ), 100% des procédures de calibration distinctes ont donné comme label l'assemblage de diatomées et de nanoeucaryotes (avec 6 mesures coïncidentes cette fois-ci), avec un taux de validation de 100% également. En effet, toutes les données *in-situ* associées à la classe *phénologie* n°41 comportent cet assemblage.

**Confirmation des labellisations existantes par rapport à Alvain *et al.* (2008); Ben Mustapha *et al.* (2014) : 5 classes de Ra( $\lambda$ )**

La Table 4.5 regroupe les informations de labellisation concernant les 5 classes de Ra( $\lambda$ ) pour lesquelles le label de Alvain *et al.* (2008) projeté sur la carte auto-organisatrice de Ben Mustapha *et al.* (2014) a pu être retrouvé. Pour ces 5 classes, le label initial était *Prochlorococcus* dominants. Ces 5 classes ont ici été labellisées comme du *Prochlorococcus* également mais non dominant selon les données *in-situ*.

La classe de Ra( $\lambda$ ) *phénologie* n°18 (classes Ra( $\lambda$ ) *som* n°29, 35, et 80) a été labellisée dans 100% des cas comme étant des *Prochlorococcus* (Table 4.5). Le score de validation montre que 86,1 % des données *in-situ* de validation identifient également la présence de *Prochlorococcus*, alors que les mauvaises identifications (13,9% des données *in-situ* de validation) sont liées à la présence de l'assemblage *Prochlorococcus* et *Synechococcus-like*, toutes trouvées pour la classe de Ra( $\lambda$ ) *som* n°35.

La classe de Ra( $\lambda$ ) *som* n°40 a été labellisée dans 71,7% des procédures de calibration comme étant du *Prochlorococcus*. Le score de validation de cette classe est de 60,8% de

TABLE 4.4 – Informations de labellisation concernant les nouvelles labellisation de classes de  $Ra(\lambda)$  par rapport aux labellisations de Alvain *et al.* (2008) et Ben Mustapha *et al.* (2014). Les classes "SOM" correspondent aux numéros des classes issues de la classification par la carte auto-organisatrice. Les classes "phénologie" correspondent aux numéros des classes issues de la classification post-carte auto-organisatrice en fonction de la temporalité des classes de  $Ra(\lambda)$ . Le nombre de matchups correspond au nombre de mesures coïncidentes pour chacune de ces classes "SOM", qui permet d'associer à la classe de  $Ra(\lambda)$  un groupe identifié dans la colonne labellisation. Au cours des 400 itérations de la procédure de labellisation, un pourcentage de labellisation peut être calculé, correspondant au nombre de fois où le label (colonne "labellisation") a effectivement été trouvé pour cette classe de  $Ra(\lambda)$ . Enfin, le score de validation moyen correspond au nombre de fois en moyenne (sur les itérations) où le label a été retrouvé dans les données de validation. Les classes de  $Ra(\lambda)$  suivies d'une astérisque ont été labellisée grâce à leur ressemblance spectrales et leur comportement d'apparition spatio-temporel (les mesures coïncidentes sont alors couplées pour la calibration et la validation et les 3 dernières colonnes sont identiques).

Classe "SOM"	Classe "Phénologie"	Nombre de matchups	Labellisation	% de labellisation	Score de validation
48	25	6	<i>Prochlorococcus</i> non dominant	100	72,33
74*	41	3	Mixte diatomées et nanoeucaryotes	100	100
94*	41	3	Mixte diatomées et nanoeucaryotes	100	100

TABLE 4.5 – Informations de labellisation concernant les confirmations des labellisations existantes de classes de  $Ra(\lambda)$  par rapport aux labellisations de Alvain *et al.* (2008) et Ben Mustapha *et al.* (2014). Pour le reste de la légende, se référer à la Table 4.4, page 122.

Classe "SOM"	Classe "Phénologie"	Nombre de matchups	Labellisation	% de labellisation	Score de validation
29*	18	1	<i>Prochlorococcus</i> non dominant	100	86,11
35*	18	8	<i>Prochlorococcus</i> non dominant	100	86,11
40	23	5	<i>Prochlorococcus</i> non dominant	71,66	60,83
57	25	5	<i>Prochlorococcus</i> non dominant	100	85,66
80*	18	4	<i>Prochlorococcus</i> non dominant	100	86,11

bonnes identifications, les erreurs étant liées à la présence de l'assemblage *Prochlorococcus* et nanoeucaryotes pour 39,2% des données *in-situ* de validation.

Enfin, la classe de  $Ra(\lambda)$  som n°57 a été labellisée dans 100% des procédures de calibration comme étant du *Prochlorococcus*. Le score de validation est en moyenne de 85,7% de bonnes identifications, les erreurs étant liées à la présence de l'assemblage de diatomées et de nanoeucaryotes pour en moyenne 14,3% des données *in-situ* de validation.

Pour l'ensemble des informations données ci-dessus, les erreurs dans la validation sont représentées par un seul groupe phytoplanctonique à chaque fois. Les erreurs d'identifications trouvées pour l'assemblage *Prochlorococcus* et nanoeucaryotes et l'assemblage *Prochlorococcus* et *Synechococcus-like* sont assez courantes, puisque ces assemblages sont souvent situés dans des conditions environnementales semblables (e.g., Dandonneau *et al.* (2004)). En revanche, l'erreur est plus importante (bien que le pourcentage d'erreur soit plus faible pour ce cas précis) lorsque l'on confond la présence de *Prochlorococcus* avec la présence de l'assemblage diatomées et nanoeucaryotes, censé plutôt occuper des environnements méso-à eu-trophe (Armbrust, 2009).



**Changement de la labellisation existante par rapport à Alvain *et al.* (2005) et Ben Mustapha *et al.* (2014) : 10 classes de Ra( $\lambda$ )**

Certaines classes de Ra( $\lambda$ ) *som* avaient été labellisées par projection des groupes phytoplanctoniques dominants de Alvain *et al.* (2008) sur la carte auto-organisatrice (Ben Mustapha *et al.*, 2014). Bien que certains de ces labels aient été validés grâce à l'utilisation de données *in-situ*, ce type de labellisation est moins rigoureux qu'une approche de calibration/validation *in-situ* telle qu'effectuée dans Alvain *et al.* (2005) ou ici. C'est pour cette raison que, lorsque le label est différent dans notre étude, nous avons considéré que ce label était plus robuste que celui donné par projection de l'ancienne labellisation (Ben Mustapha, 2013). Les informations de ces changements de label sont données dans la Table 4.6 (page 125).

La classe de Ra( $\lambda$ ) *som* n°3 a été labellisée dans 100% des procédures de calibration comme étant du *Prochlorococcus*. Le score de validation de cette classe est de 84,66% de bonnes identifications, les erreurs étant liées à la présence de l'assemblage *Prochlorococcus* et nanoeucaryotes pour 15,34% des données *in-situ* de validation. Les observations sont identiques pour le Ra( $\lambda$ ) *som* n°24 avec un score de validation moyen de 72,33% et donc 27,67% de mauvaises identifications liée à l'assemblage de *Prochlorococcus* et nanoeucaryotes. Ces deux classes de Ra( $\lambda$ ) avaient été labellisées comme du *Synechococcus spp.-like* dominant.

La classe de Ra( $\lambda$ ) *phénologie* n°4 (classes *som* n°4, 37) a été labellisée dans 70% des cas comme étant un assemblage de diatomées et de nanoeucaryotes. Le score de validation montre que 57,33 % des données *in-situ* de validation montrent également la présence de diatomées et de nanoeucaryotes, alors que les mauvaises identifications (42,67% des données *in-situ* de validation) sont liées à la présence de diatomées dominantes.

La classe de Ra( $\lambda$ ) *phénologie* n°5 (classes *som* n°5, 6, 27) a été labellisée dans 100% des cas comme étant un assemblage de diatomées et de nanoeucaryotes. Le score de validation montre que 71,5 % des données *in-situ* de validation montrent également la présence de diatomées et de nanoeucaryotes, alors que les mauvaises identifications (28,5% des données *in-situ* de validation) sont liées à la présence de nanoeucaryotes dominants.

La classe de Ra( $\lambda$ ) *som* n°36 a été labellisée dans 100% des procédures de calibration comme étant du *Prochlorococcus*. Le score de validation de cette classe est de 70% de bonnes identifications, les erreurs étant liées à la présence de l'assemblage diatomées et nanoeucaryotes (15% des erreurs) et de nanoeucaryotes dominants (15% des erreurs).

La classe de Ra( $\lambda$ ) *som* n°44 ne possède pas assez de matchups pour être labellisée (3 mesures coïncidentes alors qu'il en faut 5). Cependant, elle est la seule classe qui présente une association potentielle avec l'assemblage de dinoflagellés dominants (2 échantillons *in-*

*situ* pour la calibration), avec des erreurs liées au groupe des diatomées dominantes. Bien que le pourcentage de calibration et de validation ne soit pas suffisant pour estimer que la classe de Ra( $\lambda$ ) *som* n°44 soit effectivement liée à la présence de dinoflagellés dominants, il est intéressant pour la suite de conserver cette labellisation, tout en gardant cette limitation à l'esprit, afin de voir comment se comporte cette classe, car aucune association entre les Ra( $\lambda$ ) et les dinoflagellés n'avait pu être trouvée par le passé. Ainsi, lorsque nous étudierons les résultats des "dinoflagellés", nous rappellerons ce fait.

### **Labellisation grâce aux ressemblances des classes (phénologie, caractéristiques de spectres et distribution spatiale) : 13 classes de Ra( $\lambda$ ) *som***

Dans cette partie, l'intérêt du regroupement des classes de Ra( $\lambda$ ) post carte auto-organisatrice est présenté. En effet, lorsqu'une classe de Ra( $\lambda$ ) *som* peut trouver un label grâce aux mesures coïncidentes, nous pouvons vérifier que les classes de Ra( $\lambda$ ) *phénologie* trouvent aussi le même label. En revanche, ces labels peuvent être validés, ou non, en fonction de l'existence de mesures coïncidentes pour les autres classes de Ra( $\lambda$ ) *som* qui appartiennent à la même classe de Ra( $\lambda$ ) *phénologie*.

La classe de Ra( $\lambda$ ) *phénologie* n°2 regroupe les Ra( $\lambda$ ) *som* n°2, 7 et 8. La labellisation a montré que les Ra( $\lambda$ ) 7 et 8 étaient associés à la présence de nanoeucaryotes dominants. Par extension, la classe de Ra( $\lambda$ ) 2 a aussi été considérée comme étant caractérisée par des nanoeucaryotes dominants. De même, la classe de Ra( $\lambda$ ) *phénologie* n°6 regroupe les Ra( $\lambda$ ) n°9, 10 et 39 et les Ra( $\lambda$ ) 9 et 39 sont labellisés nanoeucaryotes dominants. Par extension, la classe Ra( $\lambda$ ) 10 a été labellisée nanoeucaryotes dominants.

La classe de Ra( $\lambda$ ) *phénologie* n°4 regroupe les Ra( $\lambda$ ) n°4, 19 et 37. Les Ra( $\lambda$ ) 4 et 37 ont été labellisés comme étant un assemblage de diatomées et nanoeucaryotes, ce qui a été étendu au Ra( $\lambda$ ) 19.

La classe de Ra( $\lambda$ ) *phénologie* n°7 regroupe les Ra( $\lambda$ ) n°12 et 34, avec le Ra( $\lambda$ ) 34 labellisé diatomées dominantes. Le Ra( $\lambda$ ) 12 a donc été labellisé diatomées dominantes.

La classe de Ra( $\lambda$ ) *phénologie* n°8 regroupe les Ra( $\lambda$ ) n°13 et 88. Le Ra( $\lambda$ ) 88 a été labellisé comme étant lié à la présence de l'assemblage *Synechococcus* et *Prochlorococcus*, avec une validation de 62%. Par extension, le Ra( $\lambda$ ) 13 a lui aussi été labellisé comme étant un assemblage de *Synechococcus* et *Prochlorococcus*.

La classe de Ra( $\lambda$ ) *phénologie* n°9 regroupe les Ra( $\lambda$ ) n°14, 17 et 100. Pris une à une, ces classes de Ra( $\lambda$ ) ne peuvent pas être labellisées grâce aux données *in-situ*, mais leur couplage donne 7 mesures coïncidentes qui permet de labelliser la classe *phénologie* n°9 comme étant un assemblage de diatomées et de nanoeucaryotes avec une validation de 100%. Le Ra( $\lambda$ ) 14 ne possède pas de mesures coïncidentes mais a été labellisé comme ces

TABLE 4.6 – Informations de labellisation concernant les changements des labellisations existantes de classes de Ra( $\lambda$ ) par rapport aux labellisations de Alvain *et al.* (2008) et Ben Mustapha *et al.* (2014). L'ancien label est noté : Nano (Nanoeucaryotes dominant), SLC (*Synechococcus-like* dominant), Diat (Diatomées dominant). Pour le reste de la légende, se référer à la Table 4.4, page 122.

Classe "SOM"	Classe "Phénologie"	Nombre de matchups	Ancien label	Labellisation	% de labellisation	Score de validation
3	3	6	SLC	<i>Prochlorococcus</i> non dominant	100	84,66
4*	4	3	Nano	Diatomées et Nanoeucaryotes	70	57,33
5*	5	2	Nano	Diatomées et Nanoeucaryotes	100	71,5
6*	5	2	Nano	Diatomées et Nanoeucaryotes	100	71,5
24	3	5	SLC	<i>Prochlorococcus</i> non dominant	100	72,33
27*	5	4	Nano	Diatomées et Nanoeucaryotes	100	71,5
36	24	6	SLC	<i>Prochlorococcus</i> non dominant	100	70
37*	4	2	Nano	Diatomées et Nanoeucaryotes	70	57,33
44	28	3	Diat	Dinoflagellés dominant	58,33	52,66
69	39	5	SLC	<i>Prochlorococcus</i> non dominant	100	74

Ra( $\lambda$ ) jumeaux.

La classe de Ra( $\lambda$ ) *phénologie* n°17 regroupe les Ra( $\lambda$ ) n°25, 67 et 70. Comme l'exemple au dessus, ces classes ne peuvent pas être labellisées une à une, mais en couplant leurs informations des mesures coïncidentes. Les Ra( $\lambda$ ) 25 et 67 ont été labellisés comme étant du *Prochlorococcus* non dominant (avec une validation de 75%), ce qui a été étendu au Ra( $\lambda$ ) 70.

Pour la suite des résultats, nous allons donc considérer trois versions de PHYSAT :

**v2014** Labellisation des anomalies de luminance brute grâce aux données *in-situ* formant des enveloppes statistiques, dont les distributions spatiales ont été projetées sur les données de classes d'anomalies de luminance issues de la carte auto-organisatrice (Alvain *et al.* (2008); Ben Mustapha *et al.* (2014));

**v2017** Labellisation des classes d'anomalies de luminance à partir des données *in-situ* de pigments classées en assemblages (ce Chapitre);

**vMixte** Une version mixte, en gardant en priorité la labellisation v2017, mais en conservant la labellisation v2014 pour les classes de Ra( $\lambda$ ) n'ayant pas pu être labellisées par manque de données *in-situ* coïncidentes selon les critères stricts de la nouvelle approche de calibration / validation.

### 4.2.3 Spectres des classes de $Ra(\lambda)$ labellisées

La Figure 4.9 présente les spectres de référence des classes de  $Ra(\lambda)$  qui ont été labellisées en fonction des 10 assemblages phytoplanctoniques déterminés *in-situ* (i.e., groupes dominants ou assemblages mixtes, définis à la page 118.). Sur cette Figure, les spectres sont associés aux versions v2008 (spectres noirs) et v2017 (spectres rouges, pointillés). L'ensemble des spectres représente la version vMixte.

Les nanoeucaryotes dominants sont caractérisés par des anomalies de luminance relative négatives (en dessous de 1), avec une chute plus importante à 412 et 443 nm, ce qui avait déjà été observé par Alvain *et al.* (2005). Cela leur donne une forme légèrement concave, égale ou inférieure à 1.

Les *Prochlorococcus spp.* sont associés à des spectres de  $Ra(\lambda)$  légèrement plus haut mais inférieurs à la moyenne (inférieurs à 1), avec une forme plate (hormis pour un spectre qui connaît un léger décrochage positif à 412nm).

Les *Synechococcus spp.-like* ont une forme de spectres assez similaire à ceux des *Prochlorococcus spp.*, avec une moyenne légèrement supérieure à 1 ce qui donne un spectre plat avec une légère pente (positive ou négative) entre 412 et 443 nm. Cette observation était faite également par Alvain *et al.* (2005), bien que la majorité des spectres présentaient une

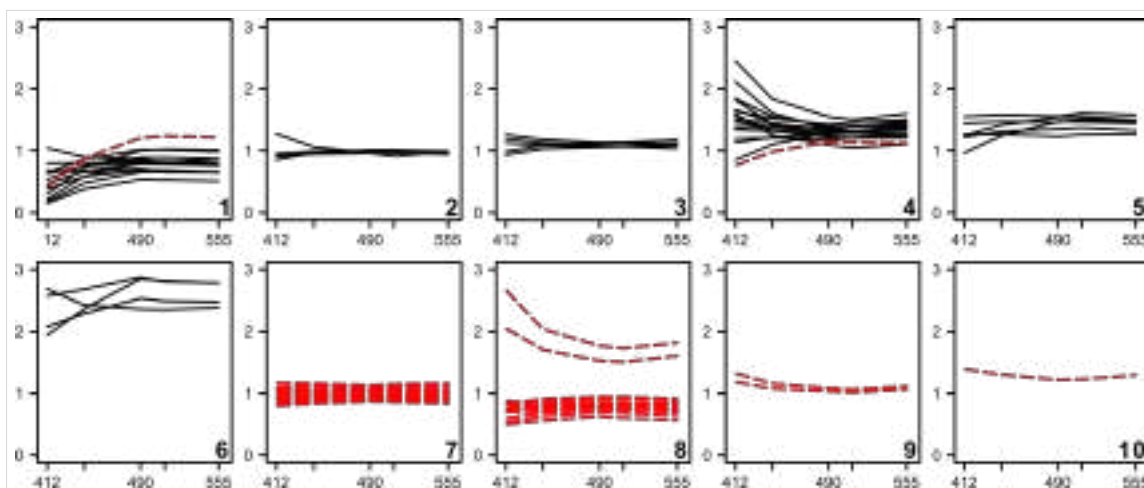


FIGURE 4.9 – Spectres de références des classes de  $Ra(\lambda)$  labellisés comme 1) Nanoeucaryotes dominants, 2) *Prochlorococcus* dominants, 3) *Synechococcus-like* dominants SLC, 4) Diatomées dominantes, 5) Phaeocystis-like dominants, 6) Coccolithophoridés dominants, 7) *Prochlorococcus* non dominants, 8) Diatomées et Nanoeucaryotes, 9) *Synechococcus* et *Prochlorococcus* et 10) Dinoflagellés. Cette labellisation a été effectuée par projection de la calibration de Alvain *et al.* (2008) sur la carte auto-organisatrice des  $Ra(\lambda)$  - v2014 (spectres noirs), ou empiriquement grâce aux mesures coïncidentes avec des données *in-situ* issues de cette thèse - v2017 (spectres pointillés rouges).

penne négative entre 412 et 443 nm.

Les diatomées sont associées à des valeurs d'anomalies de luminance au dessus de la moyenne (supérieures à 1), et ayant une forme souvent convexe due à la chute de l'anomalie de luminance entre 412 et 490 nm.

Les *Phaeocystis spp.-like*, qui n'ont pas été labellisé dans cette étude mais à partir d'une expertise de terrain, ont une forme plate, au dessus de la moyenne de 1.

Les coccolithophoridés, dans la même situation de labellisation que les *Phaeocystis spp.-like*, ont une anomalie de luminance très forte, ce qui s'explique par la forte réfraction de la lumière par leur coquille calcaire. Veuillez noter que les critères de sélection lors du calcul des anomalies de luminance tendent à mettre de côté la plupart des spectres associés aux coccolithophoridés en période de sénescence (forte réfraction de la lumière par leurs coquilles) ; leur distribution dans PHYSAT est donc sous-estimée.

Les *Prochlorococcus spp.* non dominants ont une forme de spectre plat dont l'anomalie de luminance est peu variable entre les spectres et oscille entre 0,9 et 1,1.

L'assemblage de diatomées et de nanoeucaryotes présente deux types de spectres : deux spectres présentent une forte anomalie aux faibles longueurs d'onde et ont une forme convexe, comme il l'a été observé pour les diatomées ; et les autres spectres sont légèrement concave, avec une augmentation de l'anomalie de luminance entre 412 et 490nm, puis une diminution.

L'assemblage de *Synechococcus* et de *Prochlorococcus* est associé empiriquement à deux spectres dont les formes font penser à celles des *Synechococcus-like* dominants et des *Prochlorococcus* dominants. En revanche, une légère convexité est observée ici, avec une remontée légère à 555nm.

Les dinoflagellés sont représentés par un spectre dont la forme est légèrement convexe avec une anomalie de luminance autour de 1,2. Nous rappelons que ce spectre a été conservé bien que les mesures coïncidentes n'étaient pas assez nombreuses.

#### 4.2.4 Validations moyennes des associations classes de $Ra(\lambda)$ et assemblages

Les scores de validation présentés ci-dessus tiennent compte soit des classes de  $Ra(\lambda)$  issues de la carte auto-organisatrice, soit des classes de  $Ra(\lambda)$  *phénologie* prises une à une. Il est également intéressant de voir les scores de validation par groupe phytoplanctonique. Pour cela, toutes les classes de  $Ra(\lambda)$  qui ont été labellisées (soit en v2014 (Alvain *et al.*, 2008; Ben Mustapha *et al.*, 2014), soit dans cette étude en v2017) ont été regroupées en

TABLE 4.7 – Nombre de classes de Ra( $\lambda$ ) associées à chaque groupe phytoplanctonique déterminés empiriquement *in-situ* soit dans la version de Alvain *et al.* (2008); Ben Mustapha *et al.* (2014) (v2014), soit dans cette étude (v2017), et en couplant les deux versions (vMixte).

Description du groupe	v2014	v2017	vMixte
Pas de label	26	68	21
<i>Prochlorococcus</i> non dominant	0	14	14
Assemblage diatomées et nanoeucaryotes	0	12	12
Assemblage <i>Synechococcus-like</i> et <i>Prochlorococcus</i>	0	2	2
Dinoflagellés	0	1	1
Nanoeucaryotes dominants	21	2	13
<i>Prochlorococcus</i> dominant	11	0	4
<i>Synechococcus-like</i> dominant	13	0	4
Diatomées dominant	20	1	19
<i>Phaeocystis</i> dominant (Alvain <i>et al.</i> , 2008)	5	0	5
Coccolithophoridés (Alvain <i>et al.</i> , 2008)	4	0	4

fonction de leur label afin de calculer le score de validation par groupe phytoplanctonique. Le nombre de classe de Ra( $\lambda$ ) labellisées pour chaque groupe est montré en Table 4.7. Veuillez noter que les groupes *Phaeocystis-like* et Coccolithophoridés sont présents dans cette Table mais ne peuvent pas être comptabilisé dans les scores de validation puisque ces groupes ont été déterminés à partir d'une expertise (i.e., pas à partir de données *in-situ* de pigments biomarqueurs, comme ici).

Ainsi, pour chaque groupe phytoplanctonique associé empiriquement à une ou plusieurs classes de Ra( $\lambda$ ), les scores de validation sont présentés en Table 4.8. Ces scores de validation tiennent compte de l'ensemble des informations *in-situ* disponibles pour toutes les classes de Ra( $\lambda$ ) associées empiriquement à un groupe phytoplanctonique et par méthode de labellisation (v2017 et vMixte). Il faut noter que les scores de validation sont dépendants des données *in-situ* et sont ainsi difficilement comparables d'une version de calibration à une autre. À titre d'indication, les scores de validation des précédentes versions de PHYSAT sont disponibles en Tables 2.2 (au Chapitre 2, page 43).

### Validation de la version v2017

Le meilleur score de validation est obtenu pour le groupe des diatomées dominantes, avec 92% de bonnes identifications dans les données de validation *in-situ*, qui sont confondus avec les nanoeucaryotes dominants dans 5% de ces données et avec l'assemblage Nanoeucaryotes-diatomées dans 3%. Veuillez noter que ce bon score est à nuancer car une seule classe de Ra( $\lambda$ ) a été labellisées comme diatomées dominantes dans cette version de labellisation.

Ensuite, les groupes nanoeucaryotes dominants, *Prochlorococcus*, mixte de diatomées et de nanoeucaryotes, ont respectivement des scores de validation de 81, 76, et 75 % de bonnes identifications. Les scores sont un peu moins bons pour l'assemblage de *Synechococcus-like* et *Prochlorococcus spp.* et pour les dinoflagellés<sup>6</sup> avec respectivement 62 et 52% de bonnes identifications.

### Validation de la version Mixte

Le meilleur score de validation est obtenu pour le groupe des *Synechococcus-like* dominant, avec 80% de bonnes identifications dans les données de validation *in-situ*, qui sont confondus avec les *Prochlorococcus spp.* dans 20% de ces données.

Ensuite, les groupes *Prochlorococcus* dominants, non dominants, et l'assemblage de diatomées et de nanoeucaryotes ont respectivement des scores de validation de 76, 75 et 75,6% de bonnes identifications. Les scores sont ensuite un peu moins bon pour les Nanoeucaryotes dominants (67% de bonnes identifications), les diatomées dominantes (63% de bonnes identifications) et des scores d'environ 50% sont observés pour l'assemblage de *Synechococcus-like* avec des *Prochlorococcus spp.* et pour les dinoflagellés<sup>6</sup>.

### Comparaison de l'exercice de validation de deux versions

Les scores de validation de la version v2017 (issues de la labellisation *in-situ*) sont toujours égaux ou supérieurs à ceux de la version Mixte. Ceci s'explique notamment par le fait que la version Mixte comporte des classes de  $Ra(\lambda)$  qui ont été labellisées *via* la projection des labels de Alvain *et al.* (2008) sur la carte auto-organisatrice de Ben Mustapha *et al.* (2014). En effet, cette projection a probablement entraîné une sur-labellisation des classes de  $Ra(\lambda)$ , ce qui a d'ailleurs été montré ici, puisque 10 classes de  $Ra(\lambda)$  ont été labellisées différemment dans cette nouvelle version à partir de la base de données *in-situ* disponible ; ces 10 classes représentant 13,5% du total des classes labellisés pour cette version couplée de Alvain *et al.* (2008) et Ben Mustapha *et al.* (2014). Cette version reste intéressante pour une visualisation de l'ensemble des connaissances à ce jour, surtout d'un point de vue "visuel" (cartes).

Les scores de validation les plus significativement améliorés concernent les nanoeucaryotes et les diatomées dominants. Cependant, dans la version 2017, ces deux groupes ont été associés respectivement à 2 et 1 classes de  $Ra(\lambda)$ , ce qui contribue à l'augmentation significative (mais artificielle) de la validation. En effet, plus le nombre de classes de  $Ra(\lambda)$  labellisés est grand, plus la validation tend à diminuer (ce qui correspond au fait que les erreurs d'identification ont plus de chance d'être trouvées si la labellisation est exhaustive). Cela avait déjà été observé dans l'étude de Ben Mustapha *et al.* (2014), puisque les deux plus mauvais scores de validation étaient trouvés pour les deux groupes ayant le plus de labels associés (i.e., les diatomées et les nanoeucaryotes dominants).

6. Rappelons que ce label a été gardé car il était le seul à correspondre aux dinoflagellés mais que le nombre de mesures coïncidentes était normalement insuffisant pour une labellisation correcte.



Dans les deux versions, la distribution des erreurs d'identification semble suivre un schéma similaire, par exemple, les diatomées dominantes sont confondues majoritairement avec nanoeucaryotes dominants et l'assemblage de diatomées-nanoeucaryotes dans les deux versions. Les erreurs d'identifications, plus nombreuses dans la version Mixte, sont dues au fait que le nombre de classes de  $Ra(\lambda)$  labellisées est plus grand dans cette version Mixte (voir Table 4.7) par rapport à la version 2017, ce qui tend à diminuer la validation, comme nous l'avons expliqué précédemment.

TABLE 4.8 – Scores de validation moyens des labels par groupe phytoplanctonique défini à partir des données *in-situ*. Pour chaque groupe phytoplanctonique, la validation (cellules bleues) ou la mauvaise identification (autres nombres) est donnée en ligne. Pour cela, tous les  $Ra(\lambda)$  qui ont été labellisés spécifiquement pour un groupe sont recherchés dans les données de validation. Si les  $Ra(\lambda)$  sont bien trouvés en face du groupe auquel il a été associé pendant la calibration, cela compte comme une bonne identification. Au contraire, si un autre groupe que celui labellisé est trouvé en face de ces  $Ra(\lambda)$ , cela compte comme une mauvaise identification. Les scores de validation sont divisés en deux parties : A) les scores de la version 2017 (32 classes labellisées) et B) les scores de la version Mixte (79 classes labellisées).

<b>A) Version 2017</b>								
<i>In-situ</i> → Ra( $\lambda$ ) labellisés ↓	NanoD	ProcD	SLCD	DiatD	Proc-nD	Diat-Nano	SLC-Proc	Dino
NanoD	81	-	-	9	1	9	3	-
DiatD	5	-	-	92	-	3	-	-
Proc-nD	2	17,7	2	-	75	-	3,3	-
Diat-Nano	15,5	-	-	3,5	5	76	-	-
SLC-Proc	-	8,5	8,5	-	21	-	62	-
Dino	-	-	-	47,34	-	-	-	52,66

<b>B) Version Mixte</b>								
<i>In-situ</i> → Ra( $\lambda$ ) labellisés ↓	NanoD	ProcD	SLCD	DiatD	Proc-nD	Diat-Nano	SLC-Proc	Dino
NanoD	67	-	-	9	3	18	3	-
ProcD	-	76	1	-	11	-	12	-
SLCD	-	-	80	-	20	-	-	-
DiatD	19,9	4,4	1,6	63	-	11,1	-	-
Proc-nD	1,9	7,7	2	-	75	9,6	3,8	-
Diat-Nano	-	-	-	3,1	21,3	75,6	-	-
SLC-Proc	-	12,5	12,5	-	25	-	50	-
Dino	-	-	-	47,34	-	-	-	52,66

### 4.2.5 Distribution spatiale des classes de $Ra(\lambda)$ labellisées

Dans cette partie, deux versions de cartes sont présentées (v2017 et vMixte), dans deux paragraphes distincts. En effet, ces deux versions sont à analyser avec prudence, et comportent chacune leurs avantages et inconvénients. Leurs utilisations ultérieures dépendront des applications recherchées (couverture spatiale *versus* précision de l'approche par exemple). À terme, l'augmentation des données de terrain disponibles devrait permettre de ne garder qu'une version du type 2017 (i.e., une calibration *in-situ*) qui offrirait alors une couverture spatiale homogène de qualité et plus importante qu'actuellement dans la v2017.

*Nota bene* À partir de cette Section, nous utiliserons le nom donné aux classes de  $Ra(\lambda)$  lors des labellisations empiriques pour décrire les cartes, les spectres etc., plutôt que d'utiliser les termes "les classes de  $Ra(\lambda)$  associées aux [nom du groupe]"; afin de faciliter la lecture. Notez cependant qu'à la lecture de [nom du groupe], nous faisons références à ces classes de  $Ra(\lambda)$  labellisées.

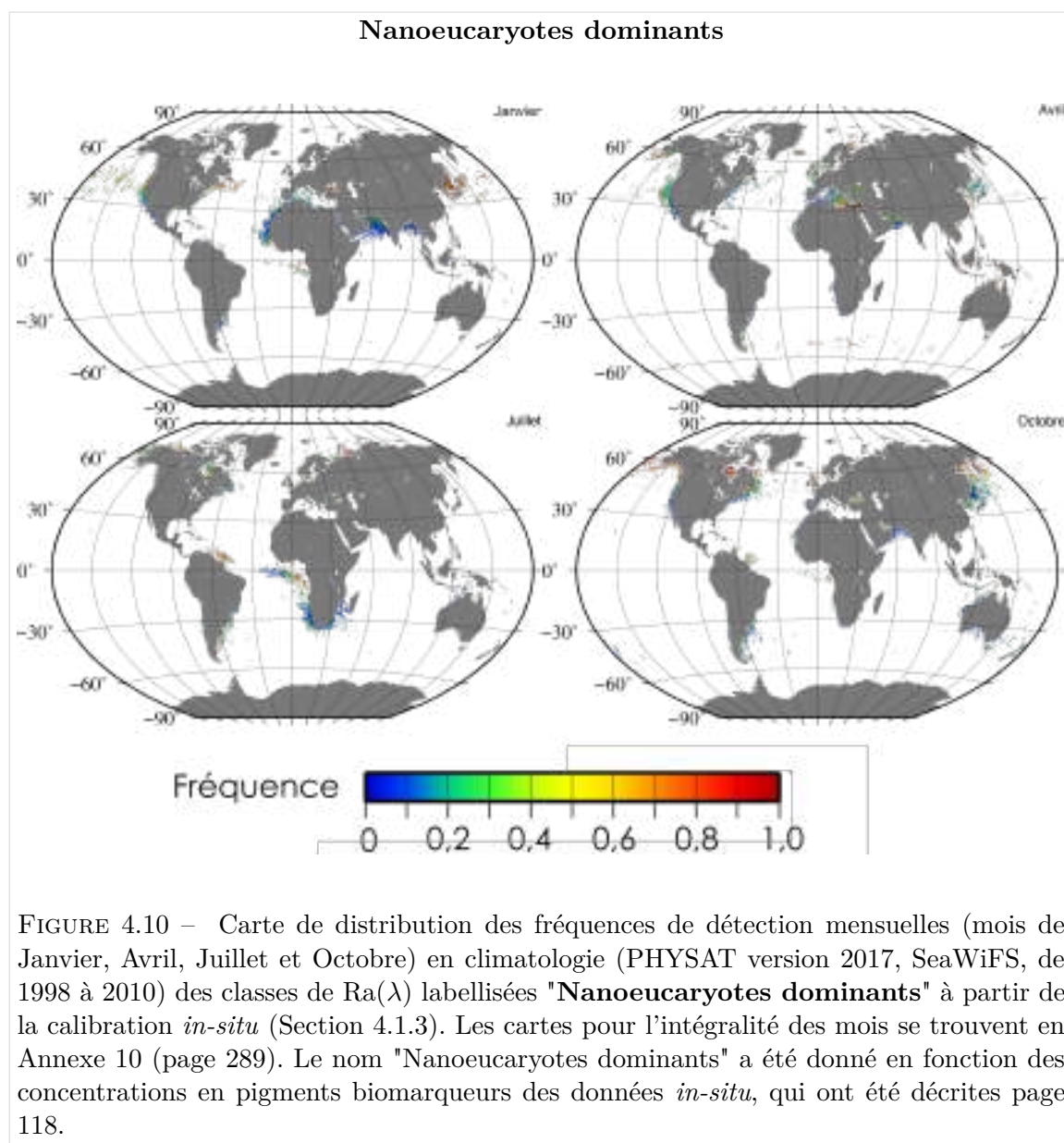
#### **Version 2017 (32 classes de $Ra(\lambda)$ labellisées) : fréquences de distribution des classes de $Ra(\lambda)$ associées aux assemblages phytoplanctoniques**

La Figure 4.10 représente les fréquences de détection des deux classes de  $Ra(\lambda)$  associées aux nanoeucaryotes dominants. Dans cette version, uniquement deux classes de  $Ra(\lambda)$  ont pu être associées empiriquement à ce groupe, ce qui explique une faible couverture spatiale des données. Néanmoins, ces classes ont une distribution plutôt proche des plateaux continentaux, et deux patches importants dans le Pacifique Nord ( $\sim$  Décembre à Mai) et dans l'Atlantique Sud ( $\sim$  Mai à Septembre).

La Figure 4.11 présente les distributions des fréquences de la classe d'anomalies de luminance associée aux diatomées dominantes pendant la calibration. Les diatomées dominantes se trouvent proches des zones côtières. Des patches de plus fortes fréquences sont également visibles dans le Pacifique Nord (presque toute l'année), l'Atlantique Nord ( $\sim$  Avril à Septembre) et dans l'Océan austral ( $\sim$  Décembre à Avril).

La Figure 4.12 présente les distributions des fréquences des 14 classes d'anomalies de luminance associées aux *Prochlorococcus spp.* non dominants pendant la calibration. Les fréquences de ces classes sont importantes toute l'année dans les gyres océaniques de tous les océans, et ne présentent pas de saisonnalité particulière.

*Nota bene* Rappel : ces fréquences ont été calculées sur la base des données labellisées dans cette version, i.e., 32% des classes de  $Ra(\lambda)$  disponibles. Ainsi, ces fréquences peuvent être atténuées dès lors que de nouvelles données *in-situ* seraient disponibles pour la calibration, et que de nouvelles classes de  $Ra(\lambda)$  se trouveraient associées à un label.



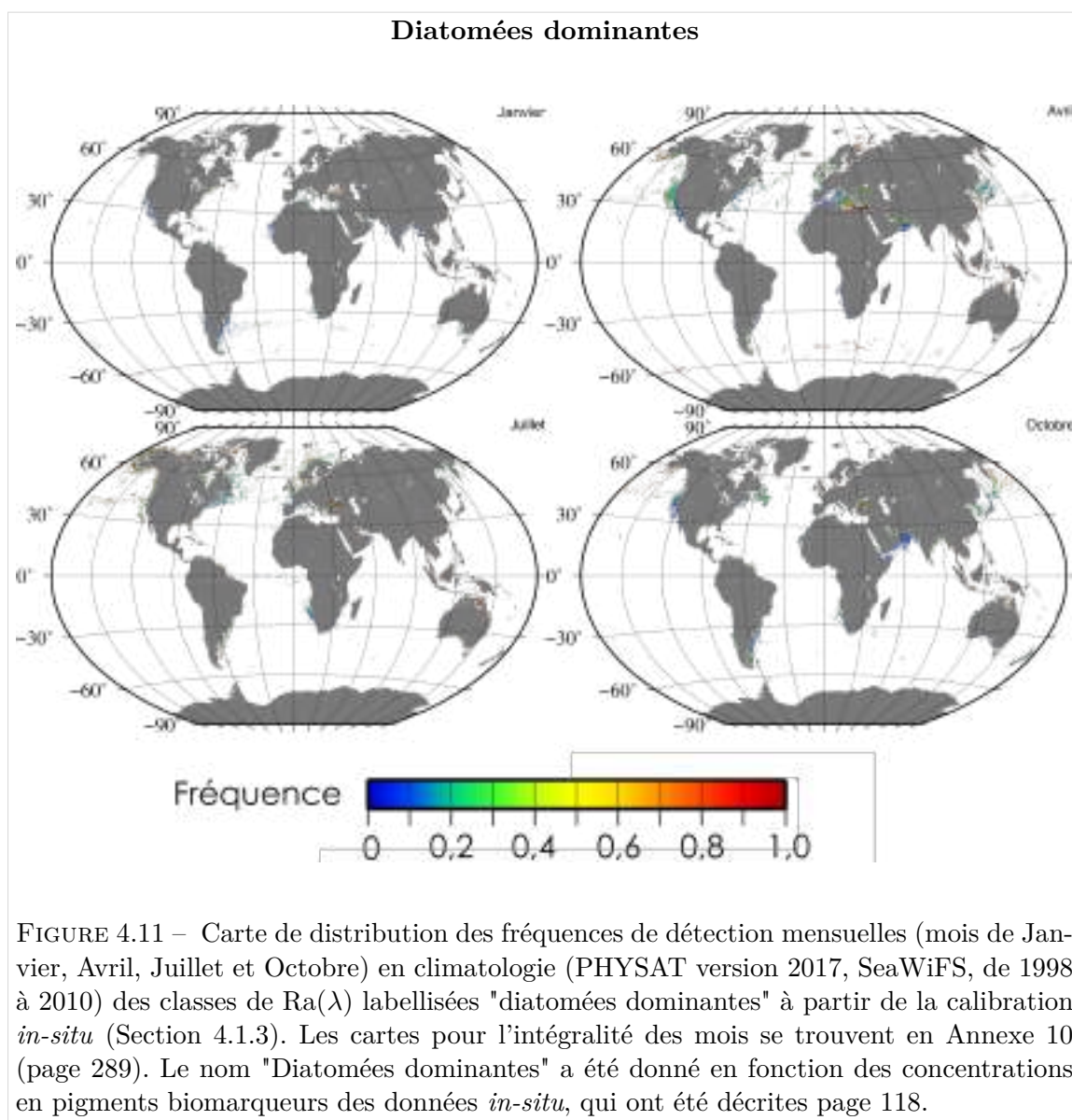
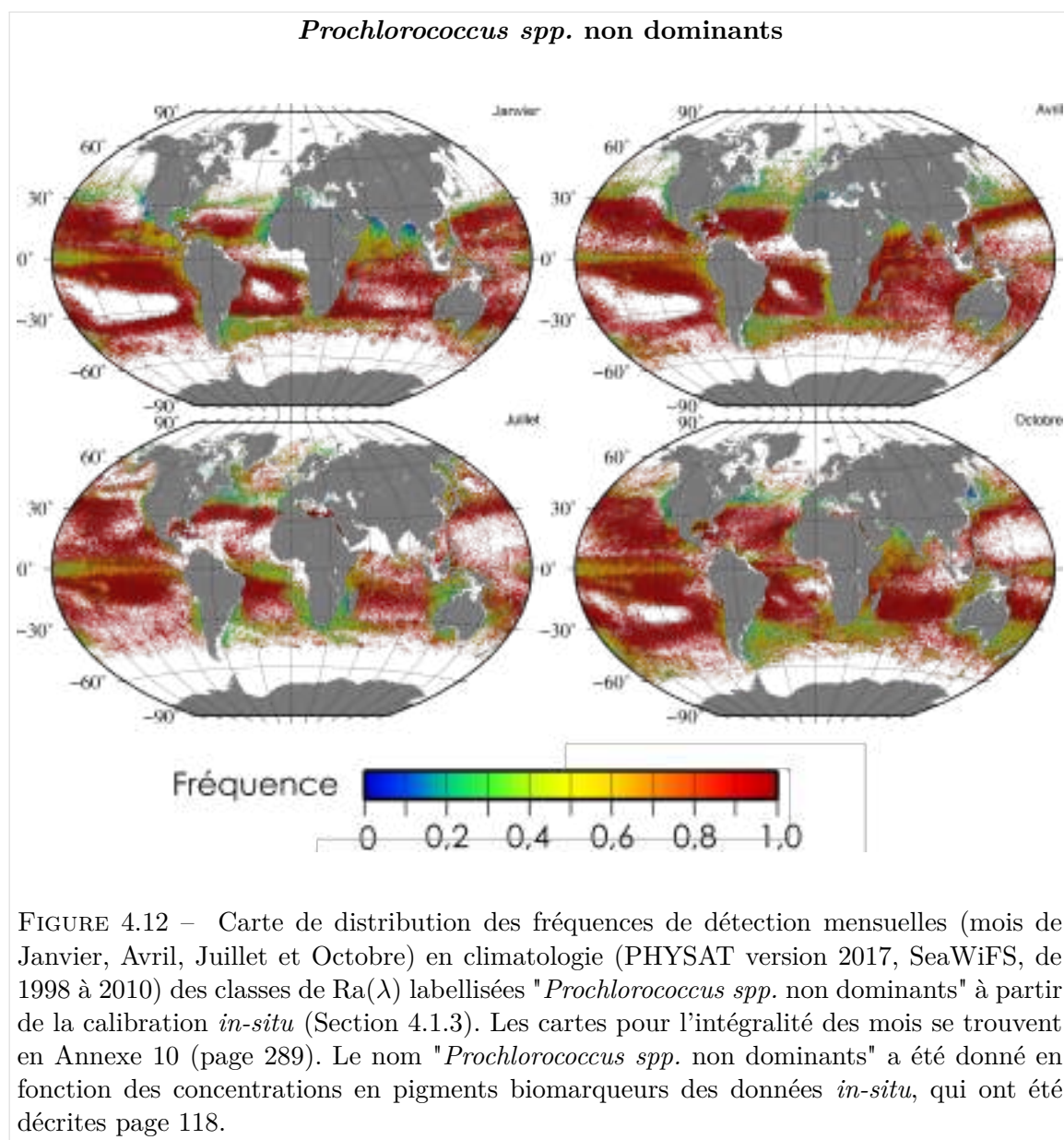
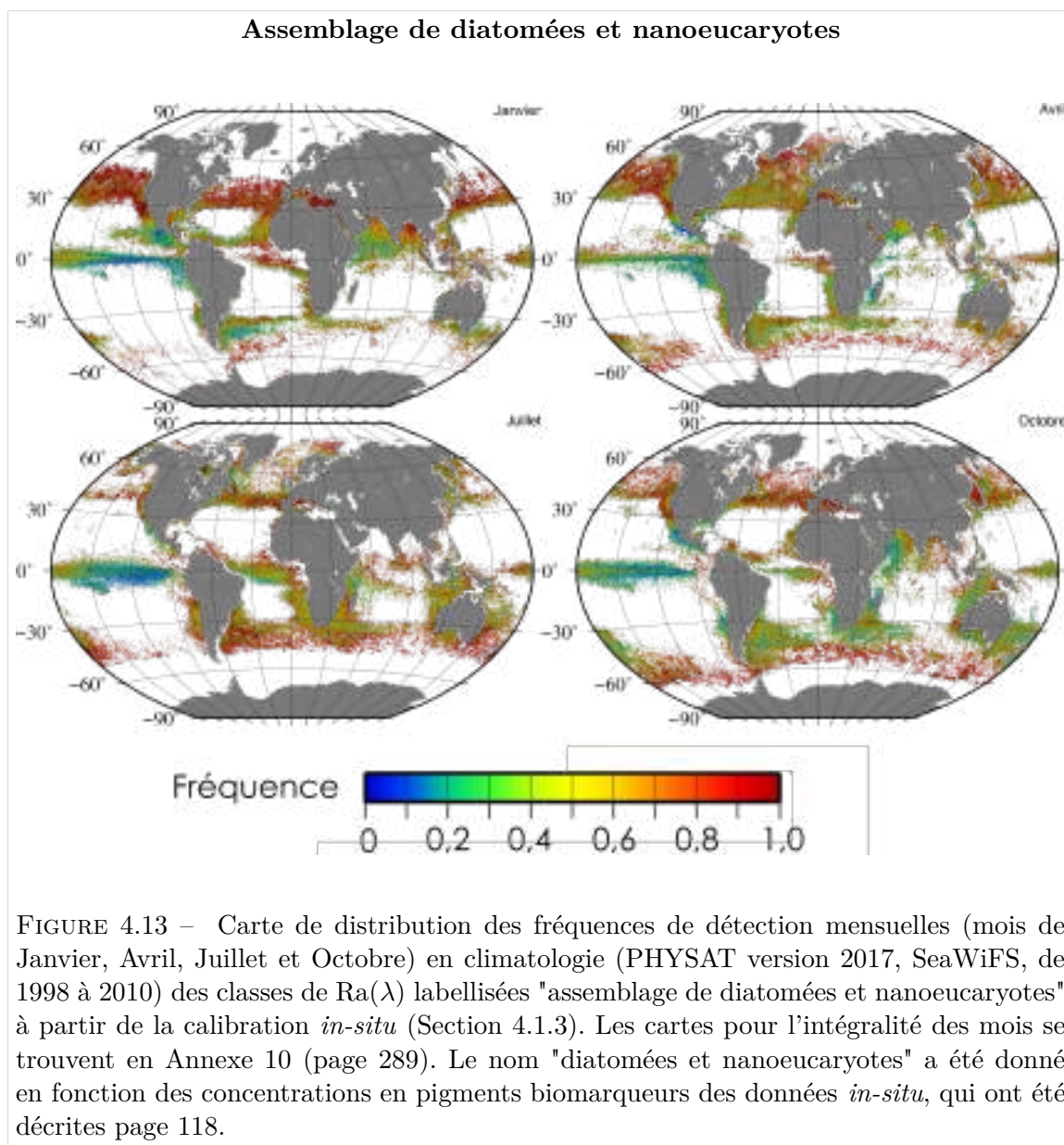


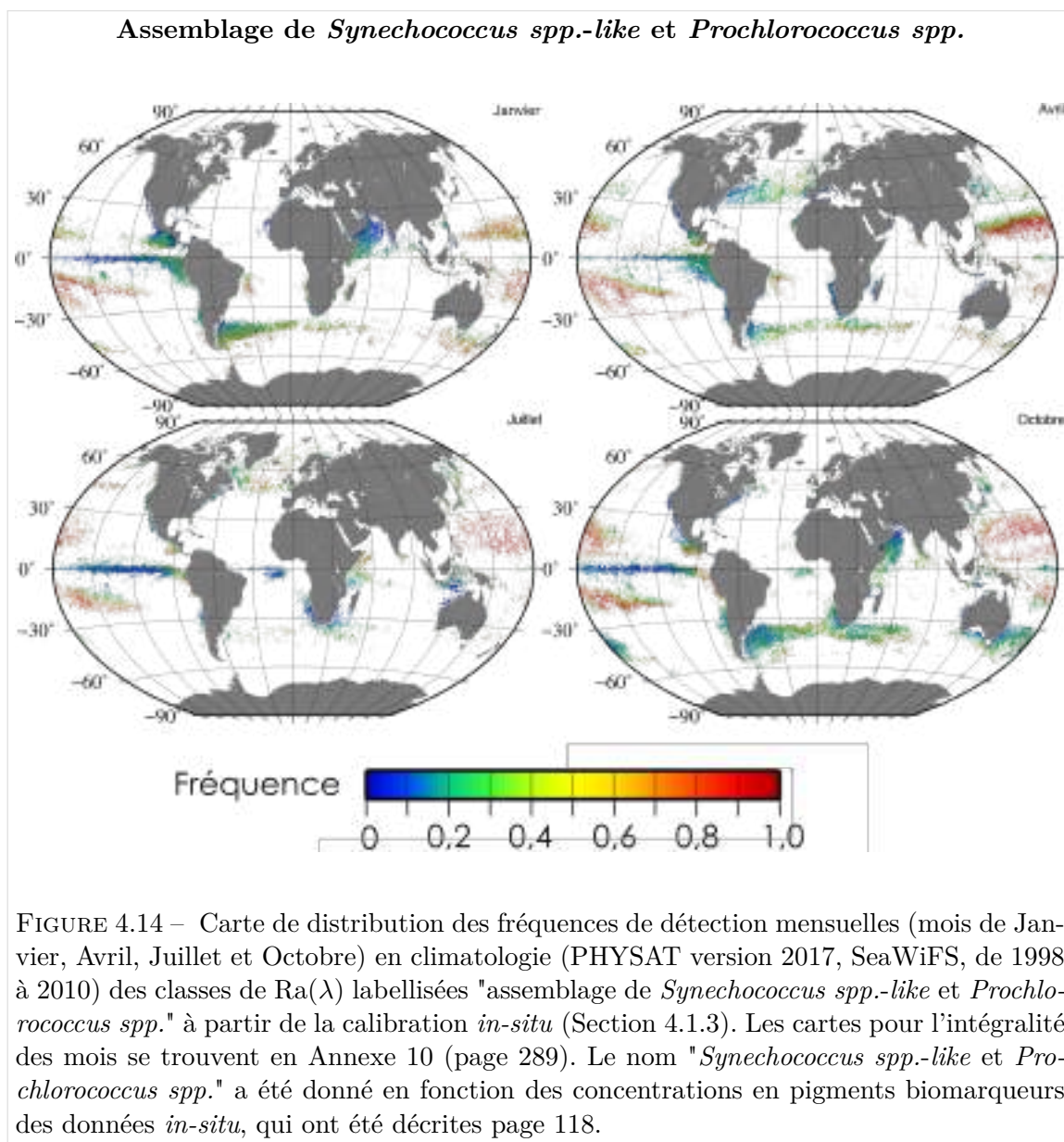
FIGURE 4.11 – Carte de distribution des fréquences de détection mensuelles (mois de Janvier, Avril, Juillet et Octobre) en climatologie (PHYSAT version 2017, SeaWiFS, de 1998 à 2010) des classes de  $Ra(\lambda)$  labellisées "diatomées dominantes" à partir de la calibration *in-situ* (Section 4.1.3). Les cartes pour l'intégralité des mois se trouvent en Annexe 10 (page 289). Le nom "Diatomées dominantes" a été donné en fonction des concentrations en pigments biomarqueurs des données *in-situ*, qui ont été décrites page 118.





La Figure 4.13 présente les distributions des fréquences des 12 classes d'anomalies de luminance associées à l'assemblage de diatomées et de nanoeucaryotes pendant la calibration. Ces classes de  $Ra(\lambda)$  ont également une forte fréquence, et sont situées autour des gyres océaniques, dans les zones de transition. Cet assemblage est présent toute l'année dans les hautes latitudes de l'hémisphère Nord et a une saisonnalité plus marquée dans les basses latitudes de l'hémisphère Sud.

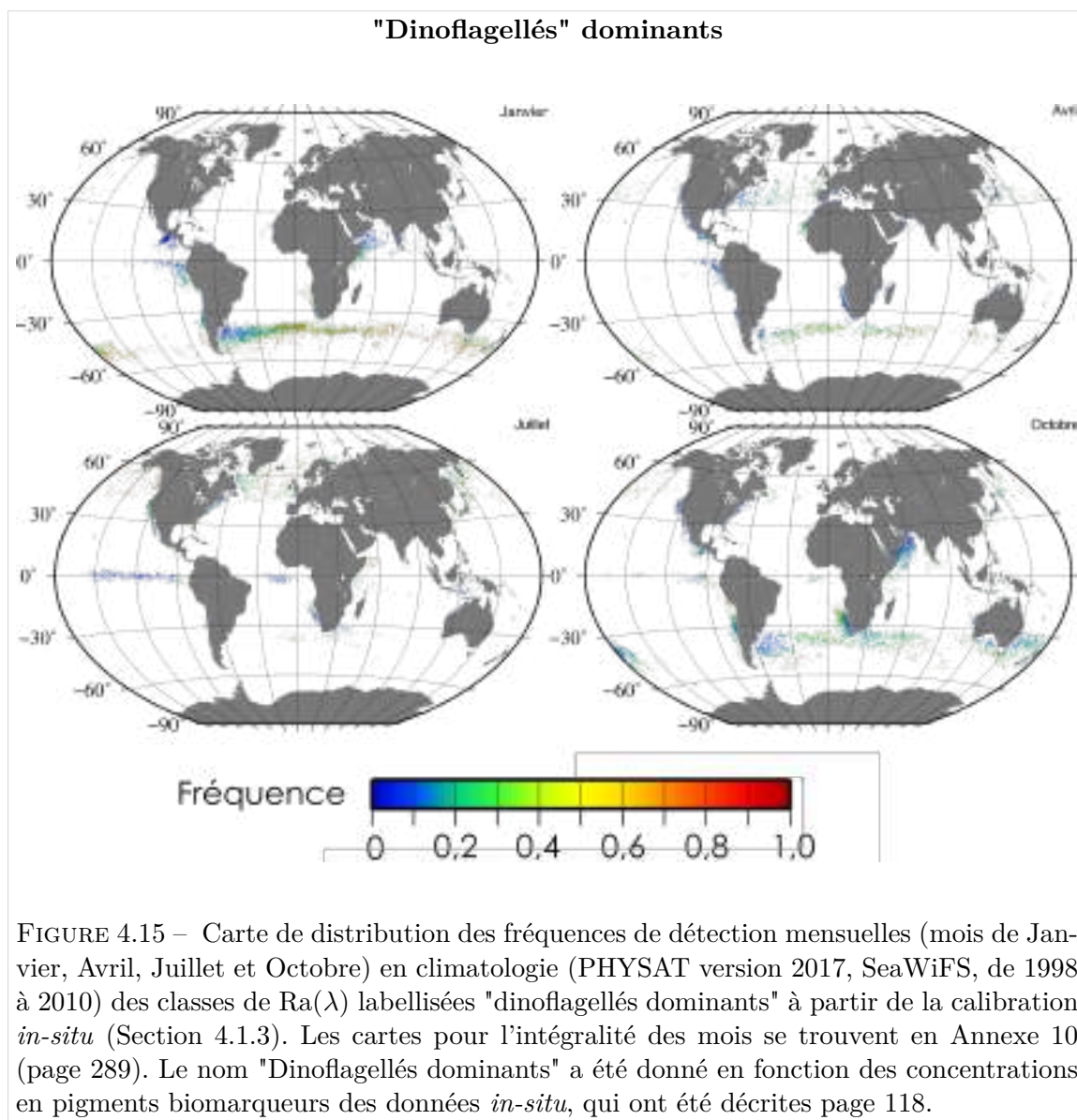
La Figure 4.14 présente les distributions des fréquences des deux classes d'anomalies de luminance associées à l'assemblage de *Synechococcus spp.-like* et *Prochlorococcus spp.* pendant la calibration. Les fréquences les plus importantes des classes de  $Ra(\lambda)$  associées à



cet assemblage sont trouvées dans les gyres océaniques du Pacifique Nord et Sud, ainsi que dans la zone de transition du Front Polaire Antarctique, bien que les fréquences soient plus faibles. La saisonnalité de ces classes de  $Ra(\lambda)$  est peu marquée au niveau des fréquences, mais elle est marquée au niveau de l'étendue spatiale des fréquences de détection, les patches de hautes fréquences ayant une couverture spatiale plus importantes à partir du mois d'Avril.

La Figure 4.15 présente les distributions des fréquences de la classe d'anomalies de luminance associée aux "dinoflagellés" dominants pendant la calibration. Nous rappelons que cette classe a été labellisée avec un nombre de mesures coïncidentes insuffisant, et que la validation est faible ( $\sim 50\%$  de bonnes identifications). Cette classe de  $Ra(\lambda)$  asso-





ciée aux "dinoflagellés" est distribuée principalement dans l'Océan Austral de Novembre à Avril, et quelques patches sont proches des côtes et en Atlantique Nord. Ce signal ressemble fortement à un signal "diatomées dominants" (e.g., Smetacek *et al.* (1990); Tremblay *et al.* (2002); Smetacek *et al.* (2002)), avec lequel les dinoflagellés dominants ont été confondus dans 47,34% des cas lors de la validation. Nous discuterons de ce signal dans la Section suivante (Discussion).

#### **Version 2017 (32 classes de $Ra(\lambda)$ labellisées) : distribution des groupes phytoplanctoniques les plus fréquents**

Les cartes de distributions (v2017) des groupes phytoplanctoniques les plus fréquents sont présentées en Figure 4.16. Sur ces cartes, un pixel est attribué à un assemblage si cet assem-

blage est le plus fréquent par rapport aux autres assemblages pour le mois considéré entre 1998 et 2010. Ces cartes sont dominées spatialement par la présence de deux assemblages (*Prochlorococcus spp.* non dominants et assemblage de diatomées et nanoeucaryotes), ce qui est expliqué par le fait qu'ils représentent plus de 80% des classes de  $Ra(\lambda)$  labellisées dans cette version.

*Nota bene* Nous rappelons que les cartes présentées en Figure 4.16, montrant le groupe le plus fréquent par pixel, sont établies sur la base de 32% de l'intégralité des données disponibles. Ainsi, un pixel considéré comme dominé dans le mois par un groupe particulier, pourrait très bien, in fine, être associé à un autre groupe phytoplanctonique dès lors que plus de données in-situ seraient disponibles, permettant de labelliser plus de classes de  $Ra(\lambda)$ . Ces cartes sont donc à analyser avec prudence (en considérant qu'une petite partie de l'information est représentée ici), comme étant un état des lieux de la connaissance à ce jour et la démonstration du potentiel à venir si suffisamment de mesures in-situ sont acquises. C'est pourquoi nous avons préféré présenter les cartes de fréquences auparavant (Paragraphe et Figures précédents).


Les nanoeucaryotes dominants sont rarement les plus fréquents et uniquement quelques patches peuvent être observés au niveau des hautes latitudes et proches des zones côtières.

Les diatomées dominantes sont plus fréquentes dans les basses latitudes (Océan austral) et quelques patches apparaissent dans les zones côtières, où elles sont rapidement mélangées avec les nanoeucaryotes dès que l'on s'éloigne des côtes.

Les *Prochlorococcus spp.* non dominants sont plus fréquents au niveau de tous les océans ouverts entre 30°Nord et -40°Sud, hormis dans les gyres océaniques Pacifique Nord Ouest et Pacifique Sud, où c'est l'assemblage de *Synechococcus spp.-like* et *Prochlorococcus spp.* qui est le plus fréquent.

Enfin, l'assemblage de diatomées et nanoeucaryotes est le plus fréquent au niveau des hautes latitudes (au dessus de 30°Nord), et au niveau du Front Polaire Antarctique, où quelques patches de "dinoflagellés" sont observés entre Décembre et Février.

Cette version v2017 présente l'avantage d'être issues d'une calibration empirique, et d'avoir été validée *via* des données *in-situ*. Cependant, le fait de n'avoir labellisé que 32% des classes de  $Ra(\lambda)$  disponible fausse la vision globale de la répartition de ces assemblages. C'est pour cette raison que nous avons proposé de présenter une version Mixte dans ce manuscrit, bien que certains des labels de cette version n'aient pas pu être validés.

 **Résumé : Répartition spatiale des classes de Ra( $\lambda$ ) associées empiriquement aux assemblages (v2017)**

Dans la version 2017, qui ne comprend que la labellisation à partir des mesures *in-situ* de concentration pigmentaires, les cartes en climatologie montrent une fréquence importante des assemblages "Nanoeucaryotes et diatomées" et "*Synechococcus spp.-like* et *Prochlorococcus spp.*". Ce sont deux assemblages qui ont été associés avec un grand nombre de classes d'anomalies de luminance (respectivement 12 et 14 classes), alors que les autres assemblages n'ont été associés qu'à 1 ou 2 classes de Ra( $\lambda$ ). Nous rappelons que les cartes présentées (fréquences et assemblage le plus fréquent) ne sont établies que sur la base de 32% des données de Ra( $\lambda$ ) disponibles. Ainsi, les fréquences et distributions de fréquences en seront modifiées dès l'ajout de nouvelles données. C'est ce que nous montrons dans la partie suivante (vMixte).

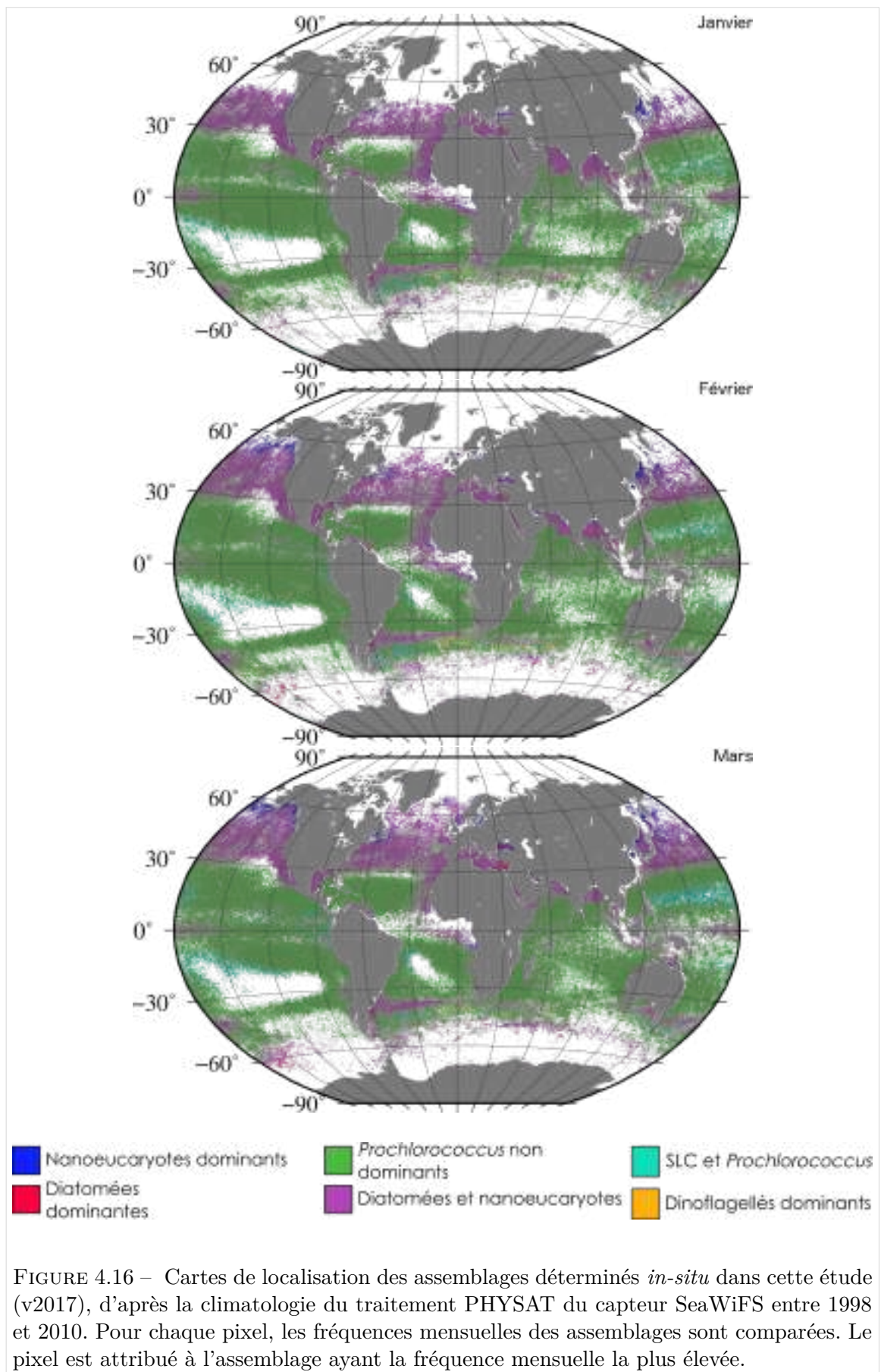


FIGURE 4.16 – Cartes de localisation des assemblages déterminés *in-situ* dans cette étude (v2017), d’après la climatologie du traitement PHYSAT du capteur SeaWiFS entre 1998 et 2010. Pour chaque pixel, les fréquences mensuelles des assemblages sont comparées. Le pixel est attribué à l’assemblage ayant la fréquence mensuelle la plus élevée.

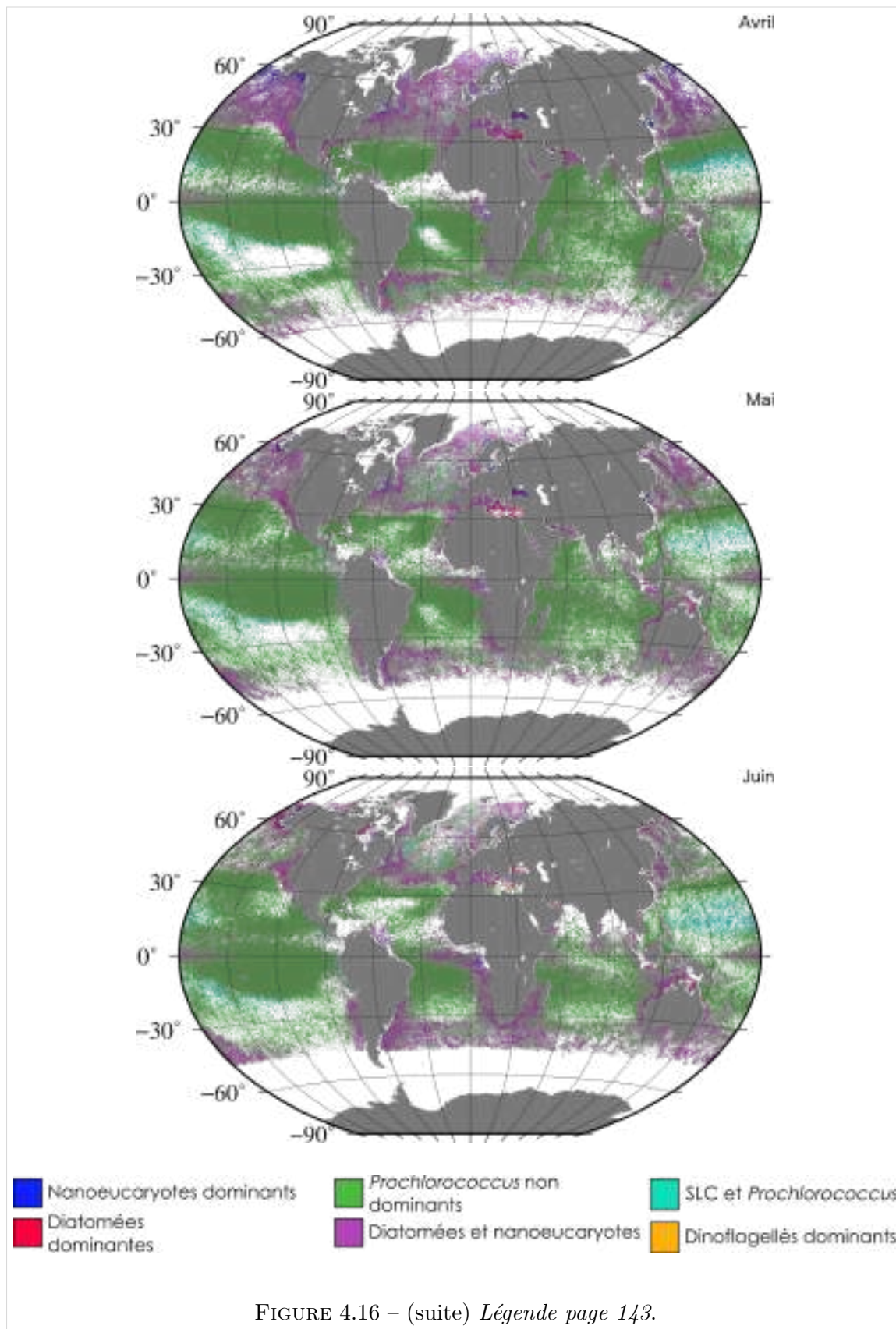


FIGURE 4.16 – (suite) Légende page 143.

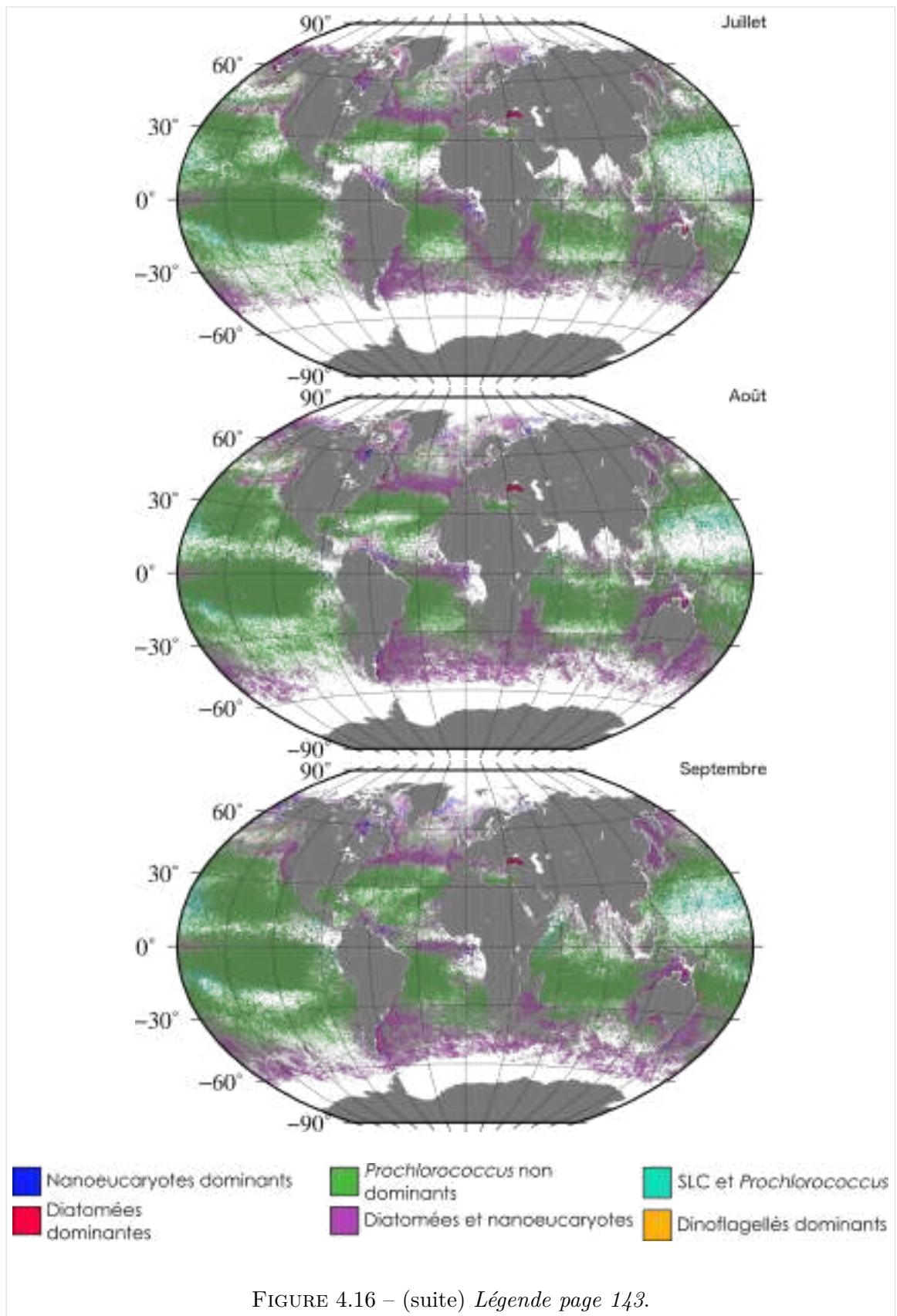


FIGURE 4.16 – (suite) Légende page 143.

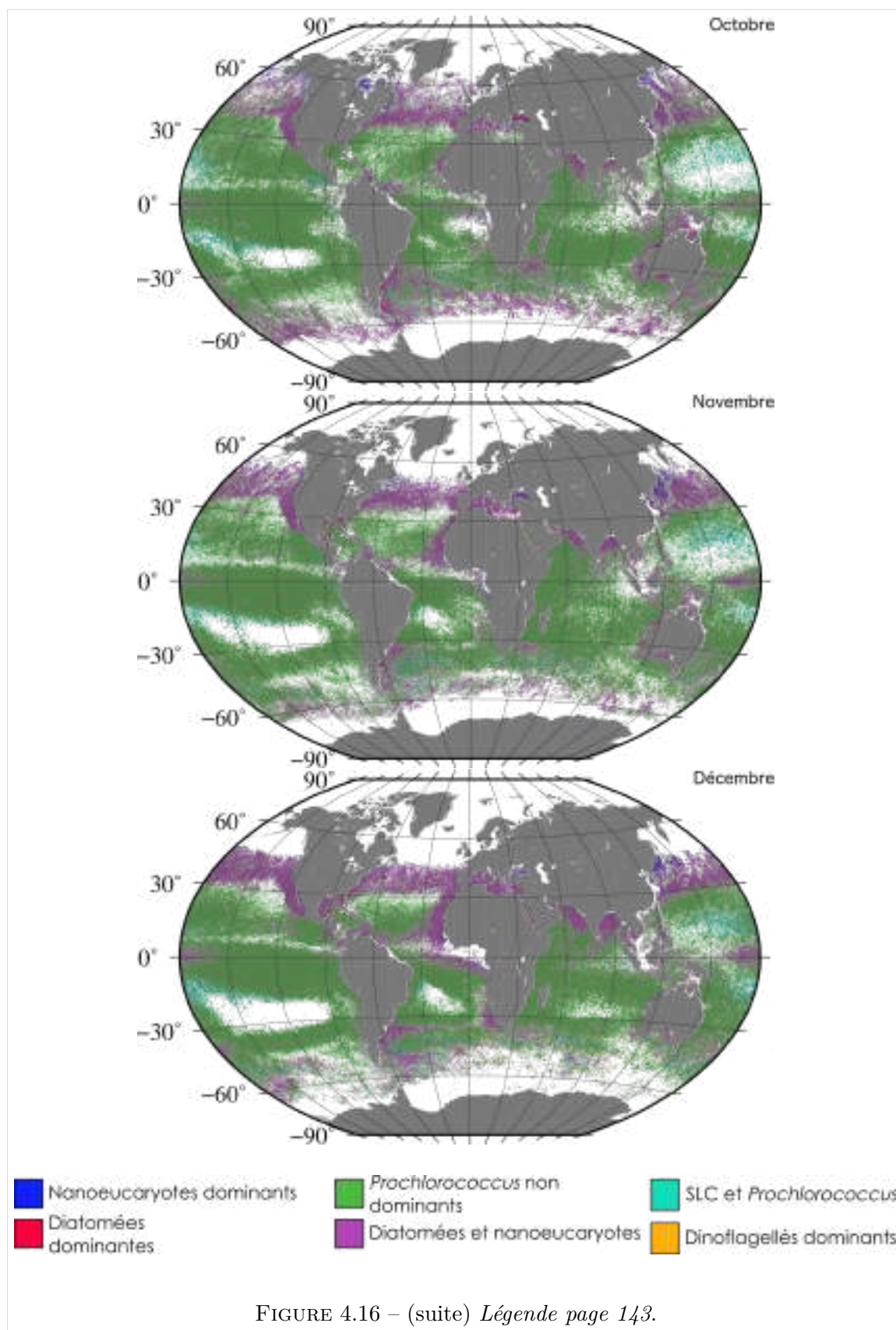


FIGURE 4.16 – (suite) Légende page 143.

### Version mixte (79 classes de $Ra(\lambda)$ labellisées) : distribution des groupes phytoplanctoniques les plus fréquents

*Nota bene* Tout comme la version 2017, nous rappelons que les cartes montrées en Figure 4.17, montrant le groupe le plus fréquent par mois, sont établies sur la base de 79% de l'intégralité de la base des types d'anomalies de luminances, et peuvent donc être modifiées à l'ajout de nouvelles données *in-situ* ou de nouvelles labellisations. Les cartes de fréquences de ces groupes, ayant servi à construire les cartes des groupes les plus fréquents, sont disponibles en Annexe 11 (page 297).

La Figure 4.17 présente les distributions des groupes phytoplanctoniques les plus fréquents. Ainsi, sur ces cartes, un pixel est attribué à un assemblage si cet assemblage est le plus fréquent par rapport aux autres assemblages pour le mois considéré entre 1998 et 2010, en se basant sur les fréquences de chaque groupe dont les cartes sont données en Annexe 11 (page 297).

Pour cette version Mixte, nous rappelons que les labels déterminés empiriquement (v2017) sont gardés en priorité. Les autres labels ont été donnés par projection de la labellisation empirique de Alvain *et al.* (2008) sur la carte auto-organisatrice de Ben Mustapha *et al.* (2014). Ce second set de labels, bien que moins rigoureux pour les raisons évoquées plus haut (notamment, la sur-labellisation due à la projection) peuvent être utilisés en combinaison de la nouvelle labellisation pour les raisons suivantes :

- Les données utilisées pour la calibration empirique de Alvain *et al.* (2008) ne représentent que 1,23% des données utilisées dans la calibration de cette thèse (v2017), qui, traitées par une méthode non linéaire, n'a pas eu un impact fort sur les nouveaux groupes phytoplanctoniques déterminés, d'autant plus que ces données sont bien réparties au sein des classes pigmentaires formées à partir des données *in-situ* ;
- Lorsque ces données sont mises de côtés (non utilisées pour la classification des échantillons *in-situ*), la classification "pigmentaire" est identique, i.e., les 10 assemblages sont formés, les proportions des échantillons *in-situ* attribués à ces assemblages ne varient pas. Ces échantillons ont été intégré à l'étude car elles ont permis la labellisation de la classe de  $Ra(\lambda)$  "diatomées dominantes", pour laquelle les mesures coïncidentes n'étaient pas suffisante sans ces anciennes données.

La Table 4.9 présente le pourcentage de pixels attribués à chaque groupe par rapport au nombre de pixels total pour lesquels un groupe a été identifié par mois. Ces données permettent d'obtenir la couverture spatiale d'un groupe phytoplanctonique, et donne également des indices sur la saisonnalité de ce groupe.



TABLE 4.9 – Pourcentage de pixels associés mensuellement à un groupe phytoplanctonique de la version Mixte, par rapport aux nombres total de pixels pour lesquels un groupe fréquent a pu être identifié, avec en colonne : 1) Nanoeucaryotes dominants, 2) *Prochlorococcus spp.* dominants, 3) *Synechococcus spp.-like* dominants, 4) Diatomées dominantes, 5) *Phaeocystis spp.-like*, 6) Coccolithes, 7) *Prochlorococcus spp.* non dominants, 8) assemblage de diatomées et nanoeucaryotes, 9) assemblage de *Synechococcus spp.-like* et *Prochlorococcus spp.*, et 10) Dinoflagellés dominants. Basé sur la climatologie mensuelle (1998-2010) de PHYSAT-SeaWiFS, version Mixte.

Assemblage → Mois ↓	1	2	3	4	5	6	7	8	9	10
Janvier	6,6	5,5	9,9	8,4	2,1	0,2	48,7	17,2	1,1	0,1
Février	8,1	6,1	10,1	6,1	1,5	0,2	50,6	16,0	1,1	0,1
Mars	10,2	6,6	8,6	3,6	1,0	0,2	52,0	15,8	1,8	0,1
Avril	12,2	7,1	6,0	2,2	0,6	0,2	55,0	15,1	1,6	0,1
Mai	11,4	6,3	5,3	2,8	0,7	0,2	55,8	15,9	1,5	0,1
Juin	8,9	6,5	5,0	4,4	1,1	0,2	53,7	18,3	1,8	0,1
Juillet	9,2	7,5	4,2	3,4	1,3	0,2	51,7	21,2	1,2	0,1
Août	9,5	7,7	3,6	2,7	1,0	0,3	52,6	21,3	1,1	0,1
Septembre	9,1	7,6	4,6	2,3	0,8	0,2	54,1	19,8	1,4	0,1
Octobre	7,9	7,8	6,2	2,7	0,7	0,2	58,1	14,9	1,5	0,1
Novembre	6,8	6,8	7,6	6,1	0,9	0,2	56,9	12,8	1,6	0,2
Décembre	5,8	5,2	8,6	10,1	1,5	0,2	51,3	15,7	1,4	0,1
<b>Moyenne</b>	8,8	6,7	6,6	4,6	1,1	0,2	53,4	17,0	1,4	0,1
<b>Écart-type</b>	1,9	0,9	2,2	2,6	0,4	0	2,7	2,6	0,3	0
<b>Variation<sup>a</sup> (%)</b>	21,7	12,7	33,9	56,5	39,5	13,9	5,1	15,5	18	-

a. Pourcentage de variations par rapport à la moyenne

Sur les cartes de la Figure 4.17, le groupe ayant la plus grande couverture spatiale moyenne sont les *Prochlorococcus spp.* non dominants. Ils sont en effet le groupe le plus fréquent pour en moyenne 53,4% des pixels où un groupe fréquent a été déterminé, et cette couverture spatiale est peu variable ( $\pm 2,8\%$ , soit à peine 5% de variations par rapport à la moyenne), ce qui montre également que leur saisonnalité est peu marquée. Ils sont localisés principalement dans les eaux oligotrophes tropicales, mais peuvent aussi apparaître dans des eaux relativement plus mésotrophes (sur les plateaux continentaux par exemple). Les *Prochlorococcus spp.* tendent à devenir dominants autour de ces patches de *Prochlorococcus spp.* non dominants, alors qu'ils sont mélangés avec des *Synechococcus spp.-like* dans le Nord du gyre Sud-Pacifique et dans le gyre du Pacifique Nord. Les *Synechococcus spp.-like* dominent quand-à eux la zone du Front Polaire Antarctique, côté Pacifique (de Novembre à Mars) et le centre des gyres océaniques (toute l'année).

L'assemblage de diatomées et nanoeucaryotes est le second à dominer la couverture spatiale, avec en moyenne 17% des pixels associés à cet assemblage. En revanche, l'assemblage présente une saisonnalité plus marquée : souvent dans les hautes latitudes, au-delà de 30°Nord entre Novembre et Mars, ils disparaissent peu à peu ensuite pour laisser place à des patches de nanoeucaryotes dominants et de diatomées dominantes. L'assemblage est également identifié autour du gyre océanique Atlantique Sud entre Juin et Septembre.

Les diatomées dominantes représentent le groupe phytoplanctonique à la saisonnalité la plus marquée. En effet, un très grand patch de diatomées apparaît entre Novembre et Mars dans l'Océan Austral, et une saisonnalité aussi marquée existe pour les zones côtières, et en Atlantique Nord (Mai à Juillet).

Les nanoeucaryotes dominants sont principalement identifiés dans l'hémisphère Nord et proche des côtes en début d'année, puis dans les zones côtières de l'hémisphère Sud à partir de Juin, leur période de plus grande couverture spatiale étant entre Mars et Mai.

Les dinoflagellés, comme nous l'avons vu précédemment (paragraphe concernant la v2017), sont distribués dans l'Océan austral, mais ils ne sont pas assez fréquents pour dominer le signal, ce qui explique leur très faible couverture spatiale sur les cartes de distributions des groupes les plus fréquents. En effet, pour les périodes et les régions considérées, ce sont les diatomées dominantes qui ont un signal plus fréquent.

Enfin, les distributions des coccolithophoridés et des *Phaeocystis spp.-like* restent inchangées depuis la version de Alvain *et al.* (2008). Les *Phaeocystis spp.-like* sont présents dans l'Océan austral entre Novembre et Mars, dans les latitudes plus faibles que les patches de diatomées dominantes ; et les coccolithophoridés présentent des patches sur les zones côtières Nord-Est américaine, en Mer du Nord.

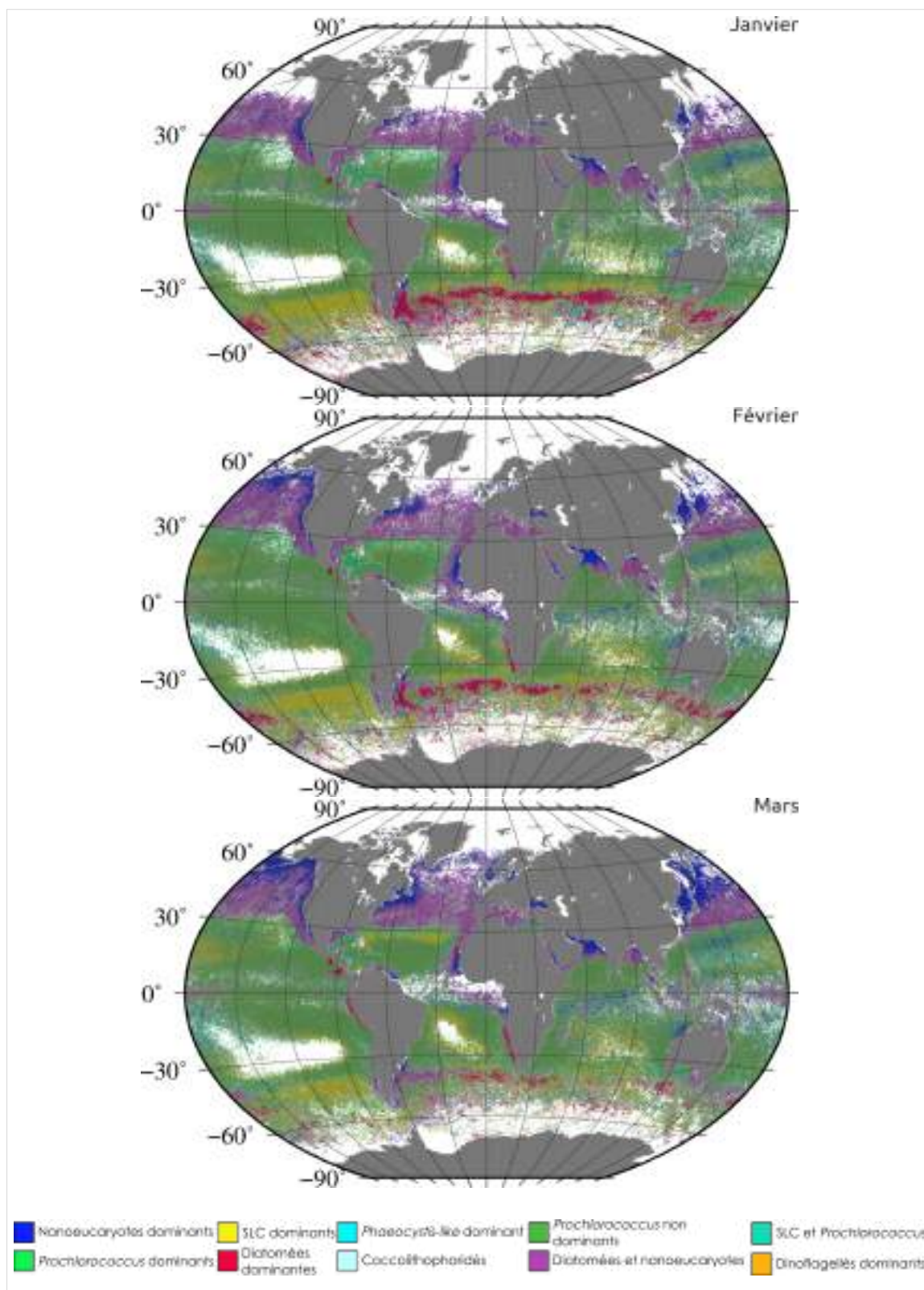
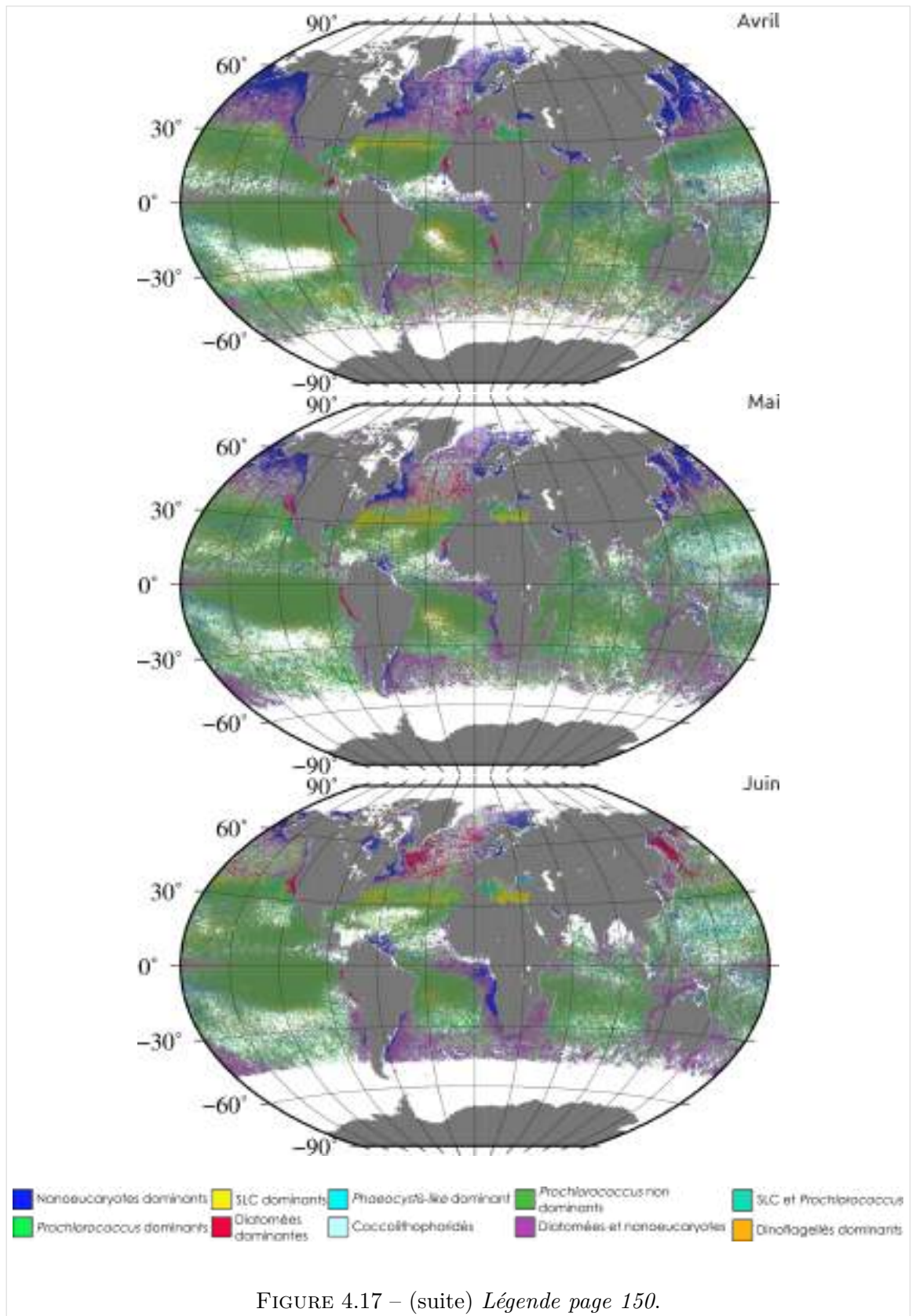
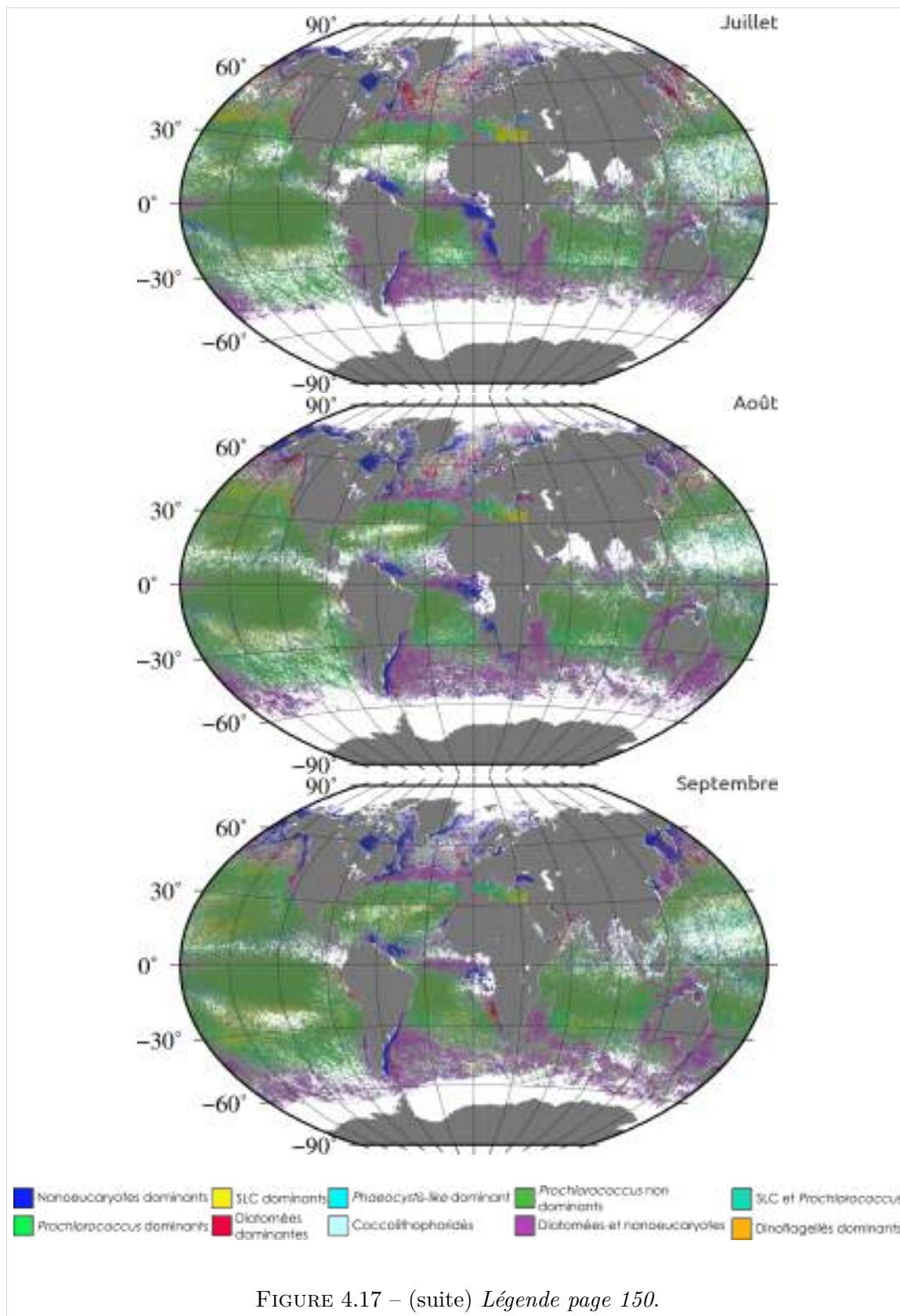
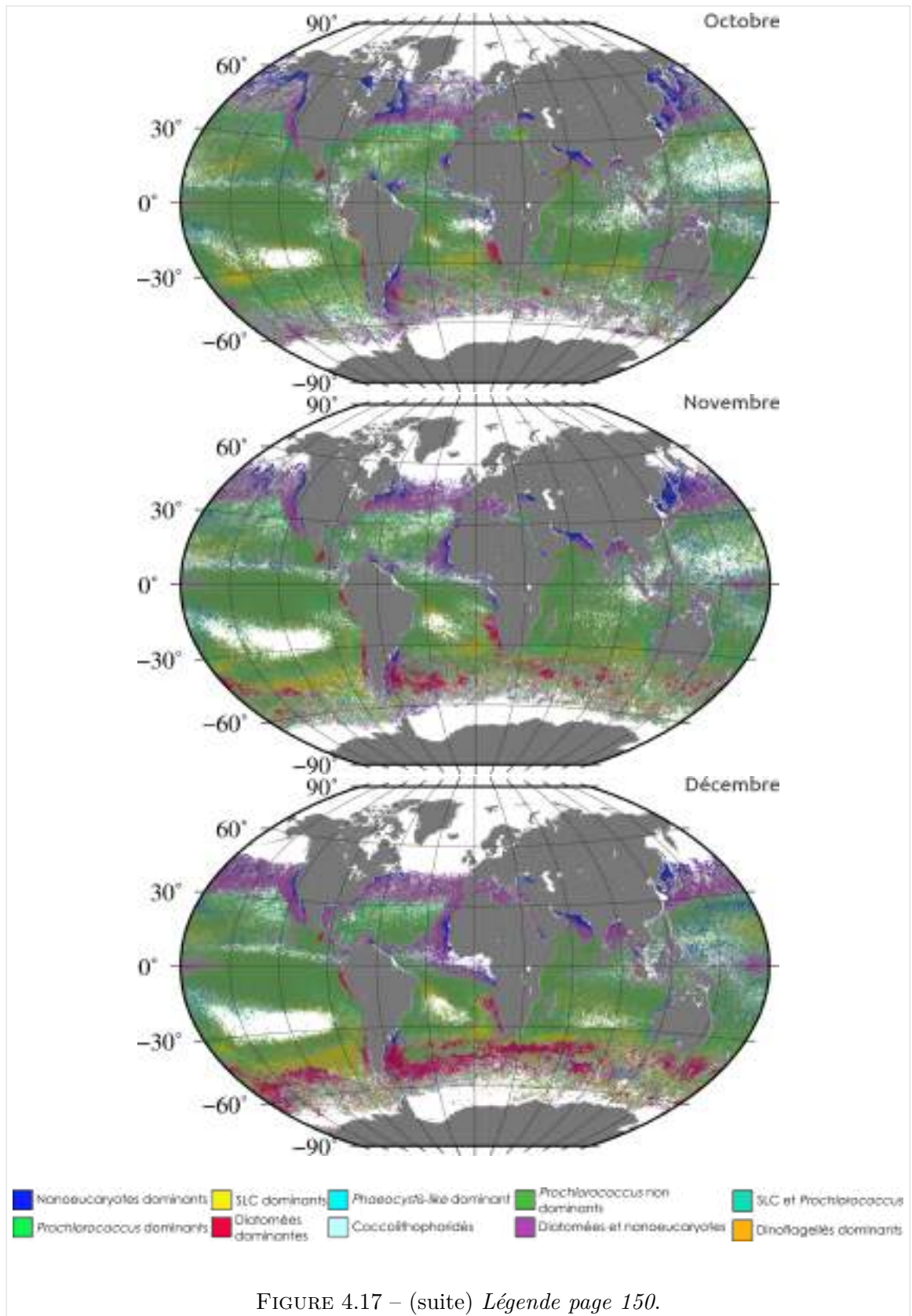


FIGURE 4.17 – Cartes de localisation des assemblages déterminés *in-situ* dans cette étude (vMixte), ainsi que dans Alvain *et al.* (2008) (Groupes *Phaeocystis-like* et Coccolithophoridés), d'après la climatologie du traitement PHYSAT du capteur SeaWiFS entre 1998 et 2010. Pour chaque pixel, les fréquences mensuelles des assemblages sont comparées. Le pixel est attribué à l'assemblage ayant la fréquence mensuelle la plus élevée.







*Nota bene* Veuillez noter que les coccolithophoridés sont identifiés grâce à la forte réfraction de leur coquille calcaire, qu'ils abandonnent lorsque la cellule meurt. Cela donne une couleur "bleue laiteuse" à l'océan, ce qui a fait d'eux le premier groupe phytoplanctonique détecté depuis l'espace. Les coccolithes induisent une anomalie de luminance très forte (entre 3 et 6 à chaque longueurs d'onde, Alvain et al. (2008)). Cependant, PHYSAT sous-estime probablement la distribution de ce groupe, en accord avec Alvain et al. (2008).

La version Mixte a pour avantage une couverture plus importante des données, ce qui donne une meilleure représentativité relative des groupes, notamment pour les cartes de distributions des groupes les plus fréquents mensuellement ; nous pouvons d'ailleurs observer que des pixels ayant été le plus fréquemment associés à un groupe phytoplanctonique dans la version 2017, sont associés à un autre groupe dans la version Mixte. Ceci rend l'utilisation possible pour des études recherchant une vision d'ensemble de la distribution comme la modélisation. Ces produits présentent donc un intérêt, mais sont à améliorer au fil du temps. Ainsi, cette comparaison des deux versions souligne l'importance des données *in-situ*. En effet, nous voyons que ces données, qu'elles soient nouvelles ou simplement analysées différemment, permettent parfois de changer des interprétations qui avaient été faites auparavant.

## 4.3 Discussion

### 4.3.1 Classification des concentrations pigmentaires *in-situ*

Les données de **pigments biomarqueurs** ont été classées grâce à un couplage d'un algorithme de Kohonen (carte auto-organisatrice Kohonen (2013)) et d'une CAH. La méthode utilisée (SOM couplée à une CAH) est une méthode non linéaire qui existe parmi d'autres. Cette méthode permet notamment de s'affranchir des notions de connaissances de la distribution, qui doit être standard (Normale) dans la plupart des autres méthodes statistiques. À partir de 2444 échantillons *in-situ*, provenant de plusieurs campagnes entre 1998 et 2010, 10 classes pigmentaires ont été distinguées. Les profils pigmentaires moyens de ces classes sont statistiquement différents (testé *via* une analyse de variance). Des noms génériques ont été donnés aux 10 classes formées pour plus de facilité dans la lecture.

Ces noms ont été donnés à titre indicatif. Ils ont été définis d'après les observations des auteurs des campagnes et à partir des écarts entre le profil pigmentaire déterminé et un profil de dominance (en comparant les concentrations relatives de références obtenues dans la classification aux seuils de dominance définis par Alvain *et al.* (2005) (Table 4.1, page 97)). Les concentrations en pigments ne permettent pas de définir précisément les proportions de tel ou tel groupe phytoplanctonique au sein d'un assemblage. En effet, le ratio de la concentration en pigments ne reflète pas une abondance.

Bien que quelques noms des assemblages semblent correspondre aux observations effec-

tuées sur le terrain (notamment pour les grosses cellules phytoplanctoniques visibles en microscopie), nous sommes conscients que l'interprétation des profils pigmentaires pourrait être différente pour un autre utilisateur de ces données. En revanche, bien que le nom donné reste à débattre, l'ensemble des échantillons que compose les 10 assemblages sont statistiquement fiables. En effet, après plusieurs essais de classifications (aussi bien au niveau de la carte auto-organisatrice que de la CAH), il reste très peu d'échantillons pour lesquels la place au sein des groupes formés est variable (environ 4% des échantillons soit une centaine d'échantillons ont une place instable dans la classification).

En comparant les assemblages définis ici avec ceux définis par Alvain *et al.* (2005), nous pouvons noter que, à partir des mêmes pigments biomarqueurs, nous ne pouvons définir que ces mêmes groupes phytoplanctoniques (i.e., les noms sont identiques), à la différence que des nuances ont été ajoutées (i.e., les assemblages). Cette limitation aux informations de concentrations en pigments biomarqueurs peut, par exemple, suggérer que le pigment biomarqueur HF19 reflète une présence de nanoeucaryotes (ou Prymnesiophycées selon Jeffrey *et al.* (1997)), qui peuvent être dominants selon les seuils sur les ratio pigmentaires (Alvain *et al.*, 2005), mais il n'est pas possible d'aller plus loin que cette définition (le groupe des nanoeucaryotes présentent une grande diversité d'espèces phytoplanctoniques). Pour une définition plus précise des groupes phytoplanctoniques et assemblages présents, une base de données plus exhaustive serait nécessaire, par exemple des données issues de comptages cellulaires (e.g., du type *Continuous Plankton Recorder*), ou de **cytométrie en flux**.

Ces échantillons *in-situ*, auxquels un nom a été attribué, ont été mis en relation avec leur donnée satellite correspondante (mesures coïncidentes), ce qui a permis de réaliser une calibration. Cette calibration a permis de labelliser empiriquement 32% des classes d'anomalies de luminance (i.e., de donner un nom grâce à l'information *in-situ*), pour lesquelles il a ensuite été possible d'analyser les cartes de distributions de fréquences.

### 4.3.2 Procédure(s) de calibration et de validation

Après avoir effectué les mesures coïncidentes entre les données *in-situ*, groupées en assemblages (profils pigmentaires), et les classes d'anomalies de luminance SeaWiFS, des calibrations ont été effectuées afin de labelliser empiriquement les classes de  $R_a(\lambda)$ . Pour rappel, la procédure de calibration et de validation a été effectuée 400 fois afin de prendre en compte différentes situations de sous-set de données et d'observer l'influence de la séparation en 80-20% du jeu de données sur la calibration.

La répétition de la calibration ne montre pas une grande variabilité. En effet, hormis pour



quelques rares classes de  $Ra(\lambda)$ <sup>7</sup>, les calibrations distinctes donnent souvent 100% de résultats identiques (et en moyenne, 92,81% de calibration sur 400 procédures), c'est-à-dire qu'une classe de  $Ra(\lambda)$ , lorsqu'elle peut être labellisée (nombre de mesures coïncidentes suffisant), l'est qu'importe la séparation entre les données de calibration et de validation. Ceci peut-être expliqué par le fait que les données de matchups sont peu nombreuses (123 mesures coïncidentes) et que leur classification est stable (voir Section ci-dessus).

En revanche, la procédure de répétition est intéressante pour la validation. En effet, si par exemple une mesure coïncidente est différente des autres en termes de groupe phytoplanc-tonique pour une classe de  $Ra(\lambda)$  qui possède 5 mesures coïncidentes, alors la validation pourra être de 0% si cette mesure différente est isolée dans les données de validation ou de 100% si cette mesure est isolée dans les données de calibration. Puisque ce point particulier à une chance sur 5 de se trouver dans les données de validation, nous devrions observer un score moyen de validation de 80% avec un pourcentage de calibration de 100%. La répétition des procédures de validation a permis de juger la robustesse de la validation face à la séparation aléatoire des 20 % de données servant à la validation. En effet, si l'on considère l'exemple ci-dessus, ne réaliser qu'une labellisation tend à sur-estimer ou sous-estimer les scores de validation, voire d'éliminer un label lorsque sa validation montre un score faible lors d'une seule procédure. Dans ce sens, une procédure répétitive permet d'augmenter les chances de trouver un label gagnant pour une classe de  $Ra(\lambda)$ .

*In-situ*, à partir de toutes les données, nous avons défini dix assemblages de phytoplancton sur la base de leur composition pigmentaire relative (description page 118). Sur ces dix assemblages *in-situ*, uniquement six ont été associés à une ou plusieurs classes de  $Ra(\lambda)$ . Les quatre autres assemblages *in-situ* ne présentaient pas assez de mesures coïncidentes et n'ont donc pas pu être associés à un signal spécifique (ces assemblages sont les assemblages n°1, 3, 7 et 9 décrits page 118).

Les pourcentages de validation de la version 2017 de la calibration s'étendent de 50% de bonnes identifications (pour les "dinoflagellés" dominants, mais qui, pour rappel, a été labellisé avec peu de mesures coïncidentes) à 92% de bonnes identifications pour les diatomées (mais pour lesquelles une seule classe de  $Ra(\lambda)$  a été labellisée). L'ensemble des scores de validation est correct. En considérant la version Mixte (cette thèse couplée à la labellisation de Alvain *et al.* (2008)), les scores de validation diminuent, mais restent acceptables (de 50 à 80% de bonnes identifications). Étant donné que la labellisation Mixte prend en compte la projection des labels de Alvain *et al.* (2008), définis par les enveloppes statistiques, sur la carte auto-organisatrice de Ben Mustapha (2013), il n'est pas étonnant de voir les scores de validation diminuer. En effet, cette projection des labels

---

7. Par exemple, les "dinoflagellés" n'ont été associés à la classe de  $Ra(\lambda)$  *som* n°44 que dans 58% des calibrations, soit 230 procédures, et dont le score de validation du label "dinoflagellés" est en moyenne de 53%. La validation montre une confusion avec les diatomées dominantes, sans que ce groupe ne soit associé à cette classe de  $Ra(\lambda)$ .

a probablement entraîné une sur-labellisation des classes de  $Ra(\lambda)$ , qui n'aurait pas été possible avec les données actuelles (en effet, nous n'avons pu labelliser que 32% de la diversité des classes de  $Ra(\lambda)$ ). Ceci n'est pas vérifiable puisque les données *in-situ* dont nous disposons ne le permettent pas.

Lorsque l'identification est mauvaise, il existe des hypothèses pouvant expliquer ces erreurs :

- Le groupe des Nanoecaryotes dominants est principalement confondu avec l'assemblage de diatomées et de nanoecaryotes (18% de confusion avec ce groupe). La dominance étant basée sur un seuil fixe sur les concentrations en pigments biomarqueurs, nous pouvons imaginer que certaines des données *in-situ* considérées comme l'assemblage de diatomées et de nanoecaryotes sont à la limite de la dominance des nanoecaryotes. En effet, pour 26,6% des mesures coïncidentes qui ont été considérées comme des diatomées et des nanoecaryotes mélangés, les nanoecaryotes ont une concentration en pigments biomarqueurs en moyenne à moins de 10% de la dominance (soit une concentration relative d'environ 0,12 en HF19 alors que la dominance est atteinte à 0,14).
- Les *Prochlorococcus* dominants sont confondus majoritairement avec l'assemblage de *Synechococcus-like* et de *Prochlorococcus* (12% des mauvaises identifications) ainsi que les *Prochlorococcus* non dominants (11% des mauvaises identifications) ; pour des raisons similaires à celles évoquées au-dessus. Dans la version de 2008 et de 2014, une part non négligeable des *Prochlorococcus* dominants étaient confondus avec les Nanoecaryotes dominants (respectivement 14 et 19,35% de mauvaises identifications pour les deux versions), ce qui n'est pas retrouvé ici d'après les données utilisées pour la validation.
- Les diatomées dominantes sont confondues le plus souvent avec les nanoecaryotes dominants (19,9% des mauvaises identifications). C'est aussi avec ce groupe que les diatomées étaient le plus confondues dans les versions antérieures de PHYSAT, sauf que les proportions de confusions étaient différentes (16,67% dans la version 2008 et 24% en 2014).
- Les *Prochlorococcus spp.* non dominants sont confondus le plus souvent avec l'assemblage de diatomées et de nanoecaryotes (9,6% de mauvaises identifications) et les *Prochlorococcus spp.* dominants dans 7,7% des mauvaises identifications. L'erreur commise entre les *Prochlorococcus spp.* dominants et non dominants s'explique aussi par les seuils sur les pigments. En revanche, elle est plus difficile à expliquer pour la confusion avec l'assemblage de diatomées et de nanoecaryotes puisqu'ils ne font pas partie des mêmes catégories de phytoplancton (taille, pigments, environnements préférentiels).

- De même, l'assemblage de diatomées et de nanoeucaryotes est lui aussi confondu avec du *Prochlorococcus spp.* non dominant (21,3% de mauvaises identifications). Il semblerait donc, d'un côté comme de l'autre que ces deux assemblages soient confondus soit parce que les données *in-situ* peuvent avoir été mal classées (cas d'échantillons instables dans la classification pigmentaire), soit parce que dans certaines situations, les spectres d'anomalies de luminance de ces deux groupes peuvent être assez ressemblants en fonction de la combinaison de facteurs qui induisent l'anomalies (Alvain *et al.* (2012), discuté ensuite).
- L'assemblage de *Synechococcus spp.-like* et de *Prochlorococcus spp.* est confondu, dans des proportions assez similaires, avec tous les autres assemblages qui contiennent ces groupes (qu'ils soient dominants, non dominant ou mélangé), avec une majorité des confusions pour les *Prochlorococcus spp.* non dominants. Comme pour les précédents points, cela peut également s'expliquer par des seuils fixes de dominance, alors que nous essayons d'établir une classification plus flexible des groupes phytoplanctoniques.
- Les dinoflagellés, qui ont été labellisés pour une seule classe de  $Ra(\lambda)$ , et avec uniquement 3 mesures coïncidentes sont confondus avec les diatomées dominantes.

Ces observations sont similaires, bien que les scores soient légèrement différents, entre les deux versions présentées ici (v2017 et vMixte).

Nous tenons à rappeler que les scores de validation dépendent des données *in-situ*. Ainsi, nous pouvons considérer que les scores de validation trouvés sont acceptables au regard du nombre de mesures coïncidentes mais surtout du nombre de groupes déterminés *in-situ*. En effet, plus le nombre de groupes *in-situ* à associer avec les classes de  $Ra(\lambda)$  est grand, plus on s'attend à diminuer les scores de validation. Cependant, les scores de validation présentés ici restent dans les mêmes ordres de grandeurs que ceux trouvés pour les précédentes validations (2008 et 2014).

De plus, les confusions trouvées pour chacun des assemblages phytoplanctoniques ont un sens écologique. En effet, les petites cellules sont confondues entre-elles (i.e., les *Prochlorococcus* et les *Synechococcus-like*), et les plus grandes cellules également (diatomées, nanoeucaryotes et dinoflagellés). La taille des cellules joue en effet un rôle sur les environnements préférentiels du phytoplancton (Margalef, 1997), ce qui induit que ces cellules de tailles similaires peuvent se trouver dans des environnements similaires. En ajoutant à cela le côté empirique de la méthode, il est facile de comprendre que ces confusions peuvent exister lors de l'association des  $Ra(\lambda)$  et des groupes phytoplanctoniques. Il aurait été plus inquiétant de confondre en grande proportion des diatomées dominantes avec des *Synechococcus-like* dominants par exemple, bien qu'il existe des situations naturelles où

ces deux groupes dominant des masses d'eau juxtaposées (dans le cas de l'efflorescence de diatomées dans l'Océan Austral par exemple).

### 4.3.3 Distribution spatiale des assemblages *via* la distribution des classes de $Ra(\lambda)$ associées

Sur les 100 classes de  $Ra(\lambda)$  disponibles à l'échelle globale pour le capteur SeaWiFS (54 classes si l'on considère leur regroupement *via* leur fréquence d'apparition, leur localisation et leur spectre), 79 ont trouvé un label dans la version Mixte et 32 dans la version de cette thèse, contre 74 dans la dernière version (Ben Mustapha *et al.*, 2014). Dix assemblages sont disponibles grâce à l'analyse empirique des mesures coïncidentes dans la version Mixte.

Les distributions spatiales des classes de  $Ra(\lambda)$  labellisées sont cohérentes avec la littérature. En effet, nous observons les principales grandes structures qui sont :

- L'absence de détection d'une efflorescence dans la zone ***HNLC*** (pour *High Nutrient, Low Chlorophyll*) du Pacifique Sud, ce qui connu (e.g., Sverdrup (1953); Martin et Fitzwater (1988); Moore *et al.* (2001))
- Les petites cellules de types *Prochlorococcus spp.* et *Synechococcus spp.-like* dans les zones oligotrophes (e.g., Armbrust (2009); Longhurst (2010))
- Les plus grosses cellules au niveau des zones côtières, d'upwelling et d'autres zones qui sont riches en nutriments (Longhurst, 2010)
- Les cas intermédiaires qui se situent dans les zones intermédiaires mésotrophes (ou les zones précédentes, sous des conditions de nutriments plus faibles à certaines saisons).

Les petites cellules du types *Prochlorococcus spp.* (dominants ou non), *Synechococcus spp.-like*, et l'assemblage de ces deux groupes, s'organisent au niveau des gyres océaniques du globe; montrant un gradient spatial de *Synechococcus spp.-like* dominants jusqu'au *Prochlorococcus spp.* dominants, en passant par des mélanges de ces deux groupes, lorsque l'on s'éloigne du centre des gyres. Ces gyres océaniques, aux centres des océans, représentent une surface importante du total des océans, et sont connus pour leur oligotrophie, i.e., des quantités de nutriments faibles par rapport à d'autres zones du globe; alors que d'autres facteurs sont suffisants pour une efflorescence (e.g., lumière, température, Longhurst (2010)). Les petites cellules phytoplanctoniques sont plus efficaces dans l'incorporation des nutriments (Margalef, 1978), alors que la quantité de ceux-ci apparait comme une limitation pour d'autres groupes (de plus grande taille notamment<sup>8</sup>). Les *Prochlorococcus spp.* sont considérés comme le type phytoplanctonique le plus abondant dans les océans

8. D'autres facteurs entre en compte, comme des besoins spécifiques en éléments particuliers en fonction du groupe phytoplanctonique, e.g., la silice.

(Armbrust, 2009), et qui possède la plus grande couverture spatiale de la surface de l'eau, ce qui est en accord avec nos observations empiriques issues de la labellisation des Ra( $\lambda$ ).

Les diatomées dominantes, et les "dinoflagellés" se trouvent dans les zones potentiellement riches en nutriments, qui sont les zones côtières, les zones d'upwelling et sont particulièrement présentes dans l'Océan Austral. Les diatomées dominantes présentent la plus forte saisonnalité si l'on considère leur étendue spatiale et leur temps de présence (i.e., le moment et l'endroit où elles sont détectées actuellement par la méthode PHYSAT).

Les diatomées possèdent, en effet, cette caractéristique de se développer rapidement lorsque les conditions le permettent. L'augmentation soudaine de la *biomasse* des diatomées est liée à l'augmentation de la photopériode, et de la quantité de lumière, mais également à la quantité de fer dans l'eau (Sunda et Huntsman, 2004; Denman, 2008). L'efflorescence de diatomées dans l'Océan Austral est également un phénomène connu (e.g., Tremblay *et al.* (2002)), et débute au mois d'Octobre dans la zone de convergence Antarctique (sous 40°Sud). Cette efflorescence peut potentiellement avoir un impact important sur le cycle du carbone *via* la *pompe biologique*, puisqu'il est estimé qu'environ la moitié de cette biomasse prendra la voie de la sédimentation (Smetacek *et al.*, 2012). Conjointement à ces diatomées, la présence de petites cellules est détectée dans notre étude (*Prochlorococcus spp.*, *Synechococcus spp.-like* et assemblage), ce qui avait déjà été constaté dans PHYSAT (Alvain *et al.*, 2008), mais aussi par d'autres études (e.g., Zubkov *et al.* (2000); Smetacek *et al.* (2002)).

Les "dinoflagellés" suivent une distribution spatio-temporelle très proche de celle des diatomées. Nous avons convenu de retenir ce label, bien que les critères de labellisation ne soient pas remplis (nombre de mesures coïncidentes). Ce groupe a été confondu, pour la moitié du temps, avec le groupe des diatomées dominantes lors de l'exercice de validation *in-situ*. Les "dinoflagellés" sont rarement les plus fréquents (moins de 0,1% de la couverture spatiale chaque mois). Il est possible que le signal que nous avons identifié soit effectivement lié aux dinoflagellés, mais il est également possible qu'il ne le soit pas. Si l'on considère la validation, ce signal pourrait en effet se rapporter aux diatomées (ou un autre groupe non identifié à ce stade). De nouvelles données pourraient nous confirmer ou infirmer l'existence de ce signal spécifique. En attendant de nouvelles données, la labellisation de ce signal sera supprimée.

Les nanoeucaryotes dominants occupent plutôt les hautes latitudes de l'hémisphère Nord. Si l'on compare le signal détecté dans cette thèse par rapport aux études PHYSAT précédentes Alvain *et al.* (2008); Ben Mustapha *et al.* (2014), nous constatons une forte diminution de la couverture spatiale de ce groupe, qui est plutôt remplacé par un signal de présence de l'assemblage diatomées et nanoeucaryotes. Alvain *et al.* (2008) ont montré, lors de leur exercice de validation, que les diatomées étaient confondues dans 40% du

temps avec des nanoeucaryotes, et ces confusions sont trouvées principalement en Atlantique Nord. Dans cette étude, nous apportons un élément de réponse à ces nombreuses mauvaises identifications en ajoutant l'assemblage des diatomées et des nanoeucaryotes, ce qui augmente légèrement les scores de validation.

L'assemblage des diatomées et nanoeucaryotes forme une transition entre la distribution des nanoeucaryotes, des diatomées et des autres groupes de petite taille. Il est en effet identifié autour des gyres océaniques et dans la zone de transition entre l'Océan Austral et les autres océans ; mais également dans les hautes latitudes de l'hémisphère Nord de Octobre à Février. En Mars, il semble y avoir une transition entre la présence de cet assemblage et l'apparition des nanoeucaryotes dominants et diatomées dominantes, avec les diatomées qui dominent le signal au centre de l'Atlantique Nord et les nanoeucaryotes qui restent dominants près des côtes.

Les distributions spatiales des groupes dominants et des assemblages caractérisés par des anomalies de luminance dans cette étude sont cohérents avec les distributions connues, soit par analyses synoptiques *via* algorithmes satellites ou des modélisations (e.g., Alvain *et al.* (2008); Brewin *et al.* (2010); Sadeghi *et al.* (2012)), par observations *in-situ* (e.g., Partensky *et al.* (1999); Zubkov *et al.* (2000); Dandonneau *et al.* (2004); Zwiarglmaier *et al.* (2007)), ou par les connaissances théoriques ou expérimentales liées aux préférences de distribution de ces *taxa* et assemblages (e.g., Margalef (1978); Partensky *et al.* (1999); Armbrust (2009); Longhurst (2010)).

#### 4.3.4 Spectres des classes de $Ra(\lambda)$ associées aux assemblages

Les spectres référents des classes d'anomalies de luminance labellisées dans cette étude (Figure 4.9) sont cohérents en termes de forme et d'amplitude avec les spectres qui avaient déjà été labellisés auparavant (soit par Alvain *et al.* (2008), ou par la projection sur les classes de spectres d'anomalies de luminance de Ben Mustapha *et al.* (2014), mais également avec les spectres attendus selon la théorie (Alvain *et al.*, 2012).

En effet, Alvain *et al.* (2012) ont montré que les diatomées sont associées à des fortes valeurs de  $Ra(\lambda)$ , notamment à cause d'un fort coefficient de rétrodiffusion. Nos résultats montrent également que ce type de spectre à forte anomalie de luminance est identifié pour les diatomées (résultats de v2017 et vMixte).

Au contraire, selon Alvain *et al.* (2012), les petites cellules ont une absorption phyto-planctonique forte, qui est compensée par des faibles valeurs d'absorption de la CDOM et de rétrodiffusion, ce qui induit une anomalie proche de 1 ; ce qui est également le cas de nos observations des spectres de  $Ra(\lambda)$  associés à ces groupes (*Prochlorococcus spp.*, *Synechococcus spp.-like* et l'assemblage des deux groupes).

Les cas intermédiaires (assemblages qui comportent des nanoeucaryotes, dominants ou non), selon la théorie, devraient présenter des anomalies inférieures à 1, qui sont dues à une forte absorption du CDOM, non compensée par la rétrodiffusion, qui est modérée pour ce groupe (Alvain *et al.*, 2012). Cette observation est retrouvée pour les nanoeucaryotes dominants, et pour la plupart des spectres associés à l'assemblage nanoeucaryotes-diatomées. En effet, pour cet assemblage, deux types de spectres sont identifiés : anomalies inférieures à 1 (plat), et anomalies supérieures à 1 (concave). Nous faisons l'hypothèse que la présence de ces deux types de spectres est liée à la proportion des diatomées et des nanoeucaryotes dans cet assemblage. Les deux spectres présentant des fortes anomalies de luminance comporteraient une proportion significativement plus grande de diatomées par rapport aux autres spectres (sans que les diatomées ne soient dominantes). Dans les mesures coïncidentes, nous avons, en effet, observé que les deux spectres avec une forte anomalie de luminance présentaient une concentration relative de Fucoxanthine plus importante (sans atteindre une concentration de dominance). Cependant, nous rappelons que la concentration pigmentaire n'est pas le reflet de la biomasse. Pour confirmer ou infirmer cette hypothèse, des mesures d'abondances spécifiques seraient nécessaires.

Les diatomées présentent plutôt un spectre ayant une forte amplitude, et dont la forme peut être convexe ou concave, en lien avec les anomalies de luminances à 412 et 443 nm. Dans cette étude précise (v2017), nous n'avons caractérisé qu'un seul spectre "diatomées dominant", mais qui reste dans les gammes de valeurs des précédents spectres labellisés par Alvain *et al.* (2008) projetés sur la SOM de Ben Mustapha *et al.* (2014), mais dont 4 types de spectres auraient été évincé par la méthode des enveloppes de Alvain *et al.* (2005) (dont le spectre labellisé dans la v2017). Les diatomées présentent la plus grande variabilité de spectres observée dans la labellisation de PHYSAT (les plus fortes variations de formes et d'amplitudes).

Bien que les "petites cellules", i.e., les *Prochlorococcus spp.* (dominants ou non), les *Synechococcus spp.-like* et l'assemblage de ces deux groupes, soient caractérisés par des spectres d'anomalies de luminance qui semblent proches en termes de caractéristiques spectrales (i.e., spectres plats, autour de la moyenne), ces spectres montrent une cohérence dans leur structure spatiale (succession spatiale des groupes dans leur environnement oligotrophe). Ils présentent des scores de validation corrects (du même ordre de grandeur, ou meilleur que Alvain *et al.* (2008)). Il est souvent remarqué que les auteurs utilisant la méthode PHYSAT pour diverses applications, fusionnent la distribution de ces deux groupes (e.g., Kostadinov *et al.* (2017)), notamment dans le but de pouvoir comparer des méthodes de détection des PFTs entre-elles. Bien qu'il reste des études complémentaires à réaliser pour déterminer en quelles mesures ces spectres sont spécifiques à la présence de ces groupes, quelques éléments, mesurés en laboratoire, montrent que ces deux groupes présentent des éléments spectraux spécifiques, comme leur absorption (plus forte pour les *Synechococcus spp.-like*, et présentant un pic à ~550nm, qui n'est pas retrouvé pour les *Prochlorococcus*

*spp.*, Morel *et al.* (1993); Stramski *et al.* (2001)), bien que la rétrodiffusion ne montre pas de différences significatives (Morel *et al.*, 1993; Stramski *et al.*, 2001).

Nous pouvons noter que l'approche de classification par la SOM, couplée à une classification supplémentaire tenant compte de la similarité spatio-temporelle de détection, est ici fondamentale. En effet, bien que les spectres labellisés ne se recoupent pas (car ils sont issus des classifications), ils peuvent parfois être proches dans leurs caractéristiques spectrales, i.e., forme et amplitude. En suivant une approche du type des enveloppes de Alvain *et al.* (2008), nous constatons qu'il aurait été difficile ici de définir ce même type d'enveloppes, sans définir de conditions statistiques strictes afin de ne pas prendre en compte les spectres spectralement similaires et ainsi éviter les confusions. Les enveloppes auraient donc été une approche réductrice, et n'auraient pas permis de prendre en compte une grande diversité de spectres d'anomalies de luminance dans la caractérisation des assemblages phytoplanctoniques.

Des hypothèses peuvent expliquer la variation du signal d'anomalie de luminance au sein d'un groupe, dans son amplitude et/ou sa forme (liste basée sur les éléments et références données aux Chapitres 1, 5). Ces éléments ont un impact sur les IOPs mesurées, et donc potentiellement sur le signal d'anomalie de luminance :

- La concentration et la composition en pigments peuvent varier en fonction de nombreux facteurs structuraux mais aussi en acclimatation à une modification de l'environnement lumineux (Bricaud et Morel, 1986; Johnsen *et al.*, 1994; Whitmire *et al.*, 2010).
- Le stade de croissance des cellules peut induire différentes tailles cellulaires, différentes morphologies et différents arrangements inter-cellulaires (cellules libres, cellules formant des chaînes, voire des agrégats) (Boss *et al.*, 2009; Moutier *et al.*, 2017).
- La concentration des matières dissoutes associées varie en fonction du stade de croissance de la population phytoplanctonique puisque, par exemple, la concentration en CDOM est plus importante en fin d'efflorescence (Bricaud *et al.*, 1981; Siegel *et al.*, 2005).
- Enfin, les assemblages décrits pourraient être associés à des communautés phytoplanctoniques, qui, bien que non dominantes, ont un impact fort sur le signal d'anomalies de luminance. Ceci n'a pas pu être évalué ici, puisque les données pigmentaires ne permettent pas de réaliser ce type d'observations. Des données d'abondance seraient plus adéquates pour évaluer ce point.

Des pistes de recherches sont données au Chapitre 6 (page 223) afin d'évaluer l'importance



de ces hypothèses dans le cadre de PHYSAT.

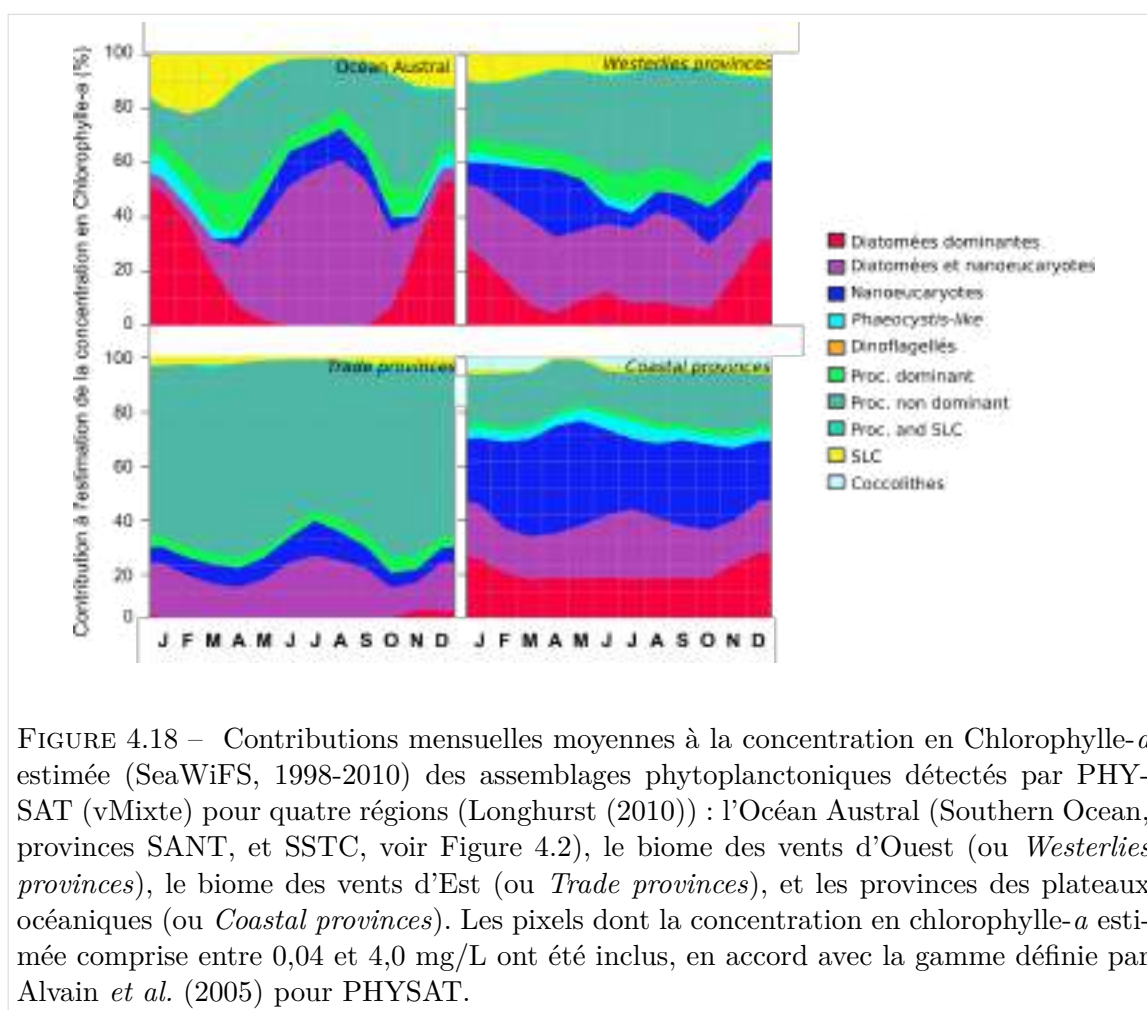
#### 4.3.5 Contribution des assemblages à la concentration en chlorophylle-*a* estimée

La Figure 4.18 présente les contributions des assemblages de la version Mixte à la concentration en Chlorophylle-*a* estimée par satellite pour quatre zones écologiques (Longhurst, 2010). Pour réaliser ces graphiques, les concentrations en chlorophylle-*a* estimées pour des pixels attribués à un groupe ont été collectées de manière journalière, puis moyennées par mois de 1998 à 2010, pour les pixels dont la concentration estimée correspond à la gamme d'applicabilité de PHYSAT (0,04 à 4,0 mg/L, Alvain *et al.* (2005)), à partir des données SeaWiFS et de l'algorithme d'estimation de la concentration en Chlorophylle-*a* correspondant (OCx, O'Reilly *et al.* (1998)).

Quatre régions écologiques ont été définies selon les biomes de Longhurst (Longhurst, 2010), afin d'observer l'adéquation de nos observations avec les caractéristiques phytoplanctoniques de ces biomes : l'Océan Austral, le biome des *vents d'Ouest*, le biome des *vents d'Est*, et le biome des plateaux continentaux (aussi appelé "provinces côtières").

L'Océan Austral est la région où les diatomées dominantes montrent leur plus forte contribution à la concentration en Chlorophylle-*a* estimée, qui atteint en moyenne plus de 50% de la concentration en Chlorophylle-*a* totale en Décembre et Janvier, où les diatomées peuvent représenter plus de 90% des cellules (Longhurst, 2010). Le reste de l'année, l'assemblage des nanoeucaryotes et des diatomées est le contributeur majeur à la concentration en Chlorophylle-*a* (avec une contribution de 30 à 60%). La plupart du temps dans cette région, ce sont donc les pico- et nano-cellules phytoplanctoniques qui sont les contributeurs majeurs à la concentration en Chlorophylle-*a* estimée, ce qui est en accord avec, par exemple, Helbling *et al.* (1991). Des facteurs environnementaux, comme l'advection de fer (Coale *et al.*, 1996; Gall *et al.*, 2001), une lumière suffisante (Nelson et Smith, 1991) et un apport de silice Boyd *et al.* (1999), permettent la croissance des grandes cellules, comme les diatomées. Dans l'Océan Austral, les diatomées sont compétitives pour la ressource et deviennent dominantes (Cavender-Bares *et al.*, 1999; Coale *et al.*, 1996; Gall *et al.*, 2001; Bathmann *et al.*, 1997; Laubscher *et al.*, 1993).

Le biome des vents d'Est, ou *Trade provinces*, caractérise les zones océaniques des faibles latitudes, qui se trouvent entre les gyres subtropicaux de tous les océans. Dans ce biome, la concentration en Chlorophylle-*a* est principalement liée à la présence des petites cellules selon nos observations (et surtout des *Prochlorococcus spp.*, qui contribuent à ~ 60% à la concentration en Chlorophylle-*a* totale), ce qui est en accord avec la définition de ce biome (Longhurst, 2010). L'assemblage de diatomées et de nanoeucaryotes présente un cycle saisonnier (présent principalement dans les zones de transition) et contribue entre 10 et 22% à la concentration en Chlorophylle-*a* estimée totale, alors que le centre des gyres



océaniques est dominé par les *Prochlorococcus spp.* et les *Synechococcus spp.-like* (e.g., Dandonneau *et al.* (2004)). Les efflorescences de cette zone sont liées principalement à des intensifications de remontée des courants profonds dans les zones de front de courants (upwelling, e.g., Longhurst (2010); Polovina *et al.* (2017)).

Le biome des vents d'Ouest (ou *Westerlies provinces*) est un cas intermédiaire. En effet, les diatomées dominantes sont les seules à présenter un cycle saisonnier, contribuant à 30% de la concentration en Chlorophylle-*a* estimée en Janvier, alors cette contribution diminue jusqu'à 5% en Avril et Octobre. L'assemblage des nanoeucaryotes et des diatomées suit un schéma similaire, avec une contribution entre 20 et 30% à la concentration en Chlorophylle-*a* estimée. Les *Prochlorococcus spp.* présentent également une contribution importante, mais sans cycle saisonnier, d'environ 25-30% de contribution à la Chlorophylle-*a*. Ducklow et Harris (1993) avaient fait cette observation dans l'Atlantique Nord, montrant que les diatomées n'étaient pas les principales responsables des concentrations en Chlorophylle-*a* observées, et montraient une forte contribution des nano- et pico-phytoplancton.

Enfin, la province côtière, qui tient compte des plateaux océaniques du globe, ne présente pas de grande variation des contributions à la concentration en Chlorophylle-*a* estimée, où seuls les diatomées et l'assemblage diatomées et nanoeucaryotes ont un cycle saisonnier. Cependant, ces estimations de contributions ne prennent en compte que la gamme d'application de PHYSAT (0,04 à 4,0 mg/L de Chlorophylle-*a* estimée), dans les eaux du Cas I (Morel et Prieur, 1977). Dans les eaux du Cas II, les diatomées et les nanoeucaryotes (et leur mélange) devraient avoir une contribution plus importante à la concentration en Chlorophylle-*a* (Armbrust, 2009). Dans ce cas, PHYSAT sous estime probablement la distribution des groupes phytoplanctoniques, et leur contribution à la concentration en Chlorophylle-*a* estimée.

*Nota bene* Ces observations des contributions des assemblages à la concentration en Chlorophylle-*a* estimées par l'algorithme de O'Reilly *et al.* (1998) donnent une vision simplifiée de l'impact de chaque groupe sur la concentration en Chlorophylle-*a* sur quelques zones du globe où ils ont des comportements différents. Cependant, il faut considérer que la concentration en Chlorophylle-*a*, telle qu'estimée par satellite, est souvent sur ou sous-estimée. Alvain *et al.* (2006) ont montré que l'algorithme OCx sous-estimait la concentration en chlorophylle-*a* en présence de diatomées, alors qu'il sur-estimait cette concentration en présence de *Synechococcus spp.-like*.

## 4.4 Conclusion

Dans cette étude, nous avons utilisé l'ensemble des données *in-situ* qui étaient à notre disposition, afin de trouver des correspondances avec les classes de  $Ra(\lambda)$ . Grâce à cette manière de procéder, nous pouvons valoriser un plus grand nombre de données *in-situ*. Bien

que les mesures coïncidentes ne représentent qu'une petite partie de l'ensemble (environ 5% de la totalité des données *in-situ*), disposer initialement de grandes bases de données permet de définir les communautés phytoplanctoniques plus précisément qu'en considérant uniquement les mesures coïncidentes.

Sur les dix assemblages définis *in-situ*, six ont pu être associés empiriquement avec la présence d'une ou de plusieurs classes d'anomalies de luminance spécifiques. Ces spectres d'anomalies de luminance sont en accord avec les spectres précédemment trouvés par Alvain *et al.* (2008), mais également avec la théorie (Alvain *et al.*, 2012). Les scores de validation de ces associations sont du même ordre de grandeur que ceux précédemment trouvés pour la méthode PHYSAT.

Cette étude a permis de montrer qu'une analyse fine de la base de données *in-situ* permet de trouver des signaux spécifiques d'anomalie de luminance pour des assemblages phytoplanctoniques au-delà de la dominance. Cette étude reste cependant à compléter par d'autres mesures de terrains. Des pistes de recherches sont données au Chapitre 6.



# Potentiel d'information porté par les anomalies de luminance : étude en Mer du Nord

---

## Problématique de ce Chapitre

D'après le Chapitre précédent, nous avons vu que les développements potentiels issus des résultats théoriques (Alvain *et al.*, 2012) peuvent être mis en application en traitant différemment l'information *in-situ* disponible à l'échelle globale. À ce stade de l'étude, nous nous sommes donc posé la question suivante :

**L'information satellite étant désormais traitée de façon à représenter au maximum la variabilité des anomalies de luminance, que se passe-t-il si on dispose par ailleurs d'informations *in-situ* détaillées ? Peut-on trouver des relations empiriques grâce aux mesures coïncidentes ? Comment traiter au mieux les données *in-situ* pour potentialiser les liens entre les deux types d'informations ?**

Pour tenter d'apporter une réponse à ces questions, nous avons choisi le jeu de données CPR (*Continuous Plankton Recorder*) en Mer du Nord, plus particulièrement l'abondance des espèces de diatomées, afin d'évaluer les relations entre la présence de diatomées spécifiques et des classes de  $Ra(\lambda)$  au-delà des cas de dominance. Cela a nécessité une adaptation des méthodes, aussi bien du côté satellite que *in-situ*.

*Cette étude a fait l'objet d'une publication dans Frontiers in Marine Sciences, acceptée le 20 novembre 2017, en Annexe 12 (page 309).*

## Introduction

Après avoir montré le potentiel de détection des **assemblages** phytoplanctoniques à l'échelle globale dans le Chapitre 4 (page 93), nous souhaitons évaluer jusqu'à quel niveau de description l'information portée par les  $Ra(\lambda)$  peut aller. Pour répondre le plus exhaustivement possible à cette interrogation, la base de données idéale devrait porter sur une grande échelle spatiale et avec un échantillonnage complet du phytoplancton à haute fréquence pendant au moins un an, de façon à avoir une bonne caractérisation des situations phytoplanctoniques *in-situ*, mais également d'accroître les chances d'obtenir des mesures coïncidentes (ciel clair par exemple). Malheureusement, une telle base de données n'existe pas encore à l'heure actuelle. Cependant, il existe des bases de données qui se rapprochent de cet idéal dans le but de tester le potentiel de l'utilisation des  $Ra(\lambda)$ . C'est, par exemple, le cas de la base de données du *Continuous Plankton Recorder*, ou CPR (SAHFOS, 2015), qui a collecté des données d'abondance phytoplanctonique sur la Mer du Nord depuis plusieurs décennies. Ce jeu de données est adéquat pour un premier essai de **labellisation** des anomalies de luminance en Mer du Nord. En revanche, les données de type "abondance", comme celles issues du CPR, n'ont pas encore été utilisées pour la labellisation de PHYSAT. Il faut donc, en premier lieu, développer une méthode permettant de les intégrer à la procédure de labellisation, c'est-à-dire en extraire une information "mise en forme" pour être mise en parallèle avec les anomalies de luminance. En effet, l'objectif est d'extraire une information pertinente de la base de données *in-situ* pour cette étude, c'est-à-dire organiser sous forme d'assemblages les diatomées présentes dans la zone sur la période considérée.

### 5.1 Matériels et méthodes

La partie *Matériels et méthodes* sera divisée en 4 parties, comme dans le Chapitre précédent, en suivant le schéma classique, qui a été introduit au Chapitre 2 (Figure 2.1, page 28) : analyse des données *in-situ*, analyse des mesures satellites, détermination des mesures coïncidentes et mise en relation des deux types d'informations.

#### 5.1.1 Caractérisation des assemblages CPR *in-situ* : étape 1

Le jeu de données CPR est inégalé en termes de nombre d'échantillons, de couverture spatio-temporelle et de représentation de l'abondance du microphytoplancton. Plus précisément, l'analyse des filtres CPR recouvrent quatre catégories de phytoplancton : les dinoflagellés, les diatomées, les silicoflagellés et les prymnésiphytes. De plus amples

informations sur le CPR (la machine et les méthodes de détermination pour le phytoplancton) sont données en Annexe 13 (page 321), afin de ne pas alourdir ce Chapitre. En particulier, le jeu de données CPR enregistre l'abondance de 72 espèces de diatomées entre 1998 et 2010, qui représentent 40 à 90% de l'abondance du microphytoplancton, en été (Juin à Août) et au printemps (Mars à Juin) respectivement (Leterme *et al.*, 2006). Il comporte 5977 échantillons sur la période qui nous intéresse, dont la distribution spatiale est présentée en Figure 5.1.

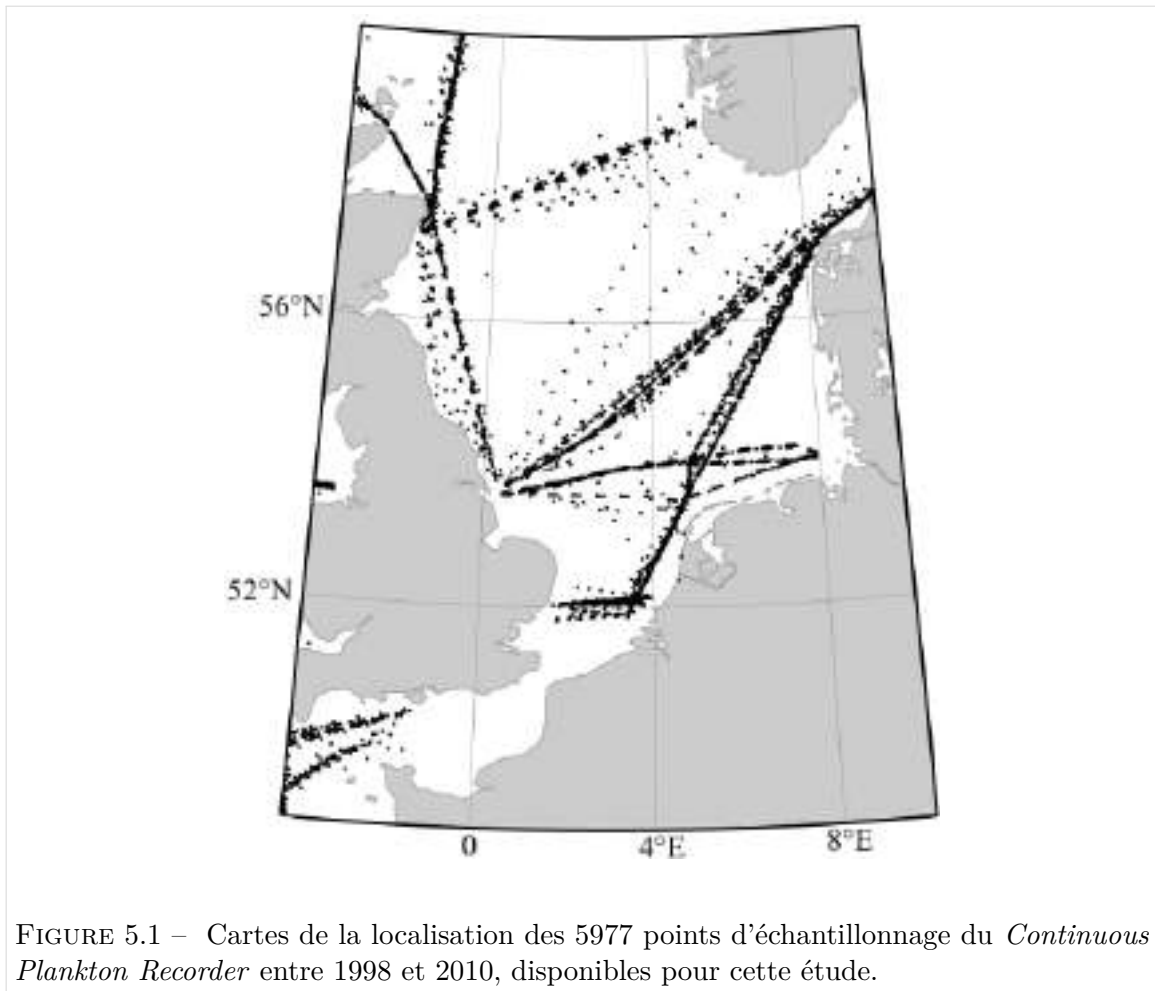


FIGURE 5.1 – Cartes de la localisation des 5977 points d'échantillonnage du *Continuous Plankton Recorder* entre 1998 et 2010, disponibles pour cette étude.

Bien que les diatomées ne soient pas le seul groupe phytoplanctonique présent en Mer du Nord, et qu'elles peuvent être associées à d'autres types de phytoplancton (notamment d'autres tailles comme le pico- et le nano-phytoplancton), s'intéresser à ce groupe taxonomique représente une première étape dans la description des communautés phytoplanctoniques. Ainsi, la définition des assemblages de diatomées, en tant qu'indicateur, peut nous donner une première réponse à la question : est-ce que des anomalies de luminance spécifiques peuvent-être associées à la présence dans les eaux de surface d'assemblages particuliers (ici, les diatomées), comme l'a suggérée l'étude théorique de Alvain *et al.* (2012) ?



Pour répondre à cette question, nous allons tout d’abord analyser la composition des données CPR *in-situ*, sans *a priori* sur une possible future association avec les  $Ra(\lambda)$ . Pour les données d’abondance, la recherche de structures dans les données (i.e., existe-il des cas similaires et récurrents de composition phytoplanctonique dans les données CPR ?) passe par une classification des échantillons suivant leur composition (Section 5.1.1.1). Ensuite, nous avons choisi d’utiliser la méthode Indicator Value de Dufrêne et Legendre (1997) pour déterminer le nombre d’assemblages récurrents dans ces données et caractériser les espèces *indicatrices* de ces assemblages (Section 5.1.1.2). Ces deux étapes sont réalisées dans le but d’obtenir une information écologique condensée, adéquate pour être associée avec les classes de  $Ra(\lambda)$  dans l’étape 3, i.e., l’étape de labellisation des anomalies.

#### 5.1.1.1 Structuration des données CPR

L’abondance des diatomées *via* le CPR enregistre l’occurrence de 72 espèces, dont la liste est donnée en Annexe 14 (page 323). De plus, nous disposons également de l’abondance des groupes phytoplanctoniques suivant : dinoflagellés, prymnésiofycées et silicoflagellés. En revanche, pour ces trois groupes, l’information n’est pas aussi détaillée que pour les diatomées car l’abondance n’est pas disponible au niveau spécifique<sup>1</sup>. Ces espèces peuvent former des assemblages présents dans des zones géographiques et des saisonnalités similaires, en accord avec leurs préférences pour le milieu physique, chimique et/ou biologique (Sverdrup, 1953; Platt et Jassby, 1976).

Pour identifier cette structuration dans les données CPR, nous avons utilisé plusieurs types de classifications hiérarchiques (lien moyen, lien complet et lien simple). Le meilleur résultat est donné pour la classification hiérarchique ascendante de type *UPGMA* (*Unweighted Pair Group Method with Arithmetic mean*, Jain et Dubes (1988)) qui est un algorithme qui réduit progressivement la taille du jeu de données en groupant successivement deux éléments dont la distance est la plus faible. En pratique, il faut calculer une distance entre les échantillons ; ici c’est la distance de Bray et Curtis qui a été utilisée car c’est une distance qui est conseillée pour le regroupement de l’abondance d’espèces (Bray et Curtis, 1957; Legendre et Legendre, 2012), et qui donne également le meilleur résultat (voir le paragraphe Justification du choix, ci-dessous).

---

1. Ces groupes ne sont pas identifiés jusqu’à l’espèce ou au *taxon* par les analystes du CPR, sauf dans le cas de quelques rares campagnes plus détaillées. Les dinoflagellés peuvent être comptés jusqu’au niveau spécifique pour les dinoflagellés à coquilles dures

### Calcul de la distance de dissimilarité de Bray-Curtis

La "distance" de Bray-Curtis<sup>2</sup> est utilisée pour évaluer la dissimilarité entre deux échantillons, en termes d'abondance d'espèces présentes dans chacun de ces échantillons. Sa valeur est comprise entre 0 (les échantillons ont une composition identique) et 1 (les échantillons sont tout à fait dissemblables).

La distance de Bray-Curtis est calculée via l'équation 5.1 sur la matrice d'abondance (abondance de chaque espèce à travers les échantillons), en prenant les échantillons "*i*" et "*j*" par paires :

$$dBC_{ij} = 1 - \frac{2 * \sum_{s=1}^n \min(N_{si}, N_{sj})}{S_i + S_j} \quad (5.1)$$

Où *i* et *j* sont les deux échantillons comparés pour lesquels *n* espèces sont communes ;  $N_{si}$  est d'abondance de l'espèce *s* dans l'échantillon *i* et  $N_{sj}$  l'abondance de l'espèce *s* dans l'échantillon *j* ;  $\min(N_{si}, N_{sj})$  correspond au minimum obtenu entre  $N_{si}$  et  $N_{sj}$  sur les échantillons *i* et *j* et  $S_i$  et  $S_j$  sont les sommes respectives des abondances des *n* espèces présentes dans les deux sites *i* et *j*.

Une matrice de dissimilarité est obtenue. Dans le cas des données CPR, cette matrice a une dimension de 5977 lignes par 5977 colonnes. C'est cette matrice de dissimilarité qui a été utilisée pour classer les échantillons.

Cependant, le calcul de la dissimilarité de Bray-Curtis pose un problème dans les situations extrêmes pour le regroupement (valeurs de dissimilarité de 0 et de 1). En effet, les situations extrêmes vont avoir des dissimilarités avec les autres échantillons proches de la valeur 1, ce qui va avoir pour conséquence de réduire les différences observables dans les autres échantillons. Pour pallier ceci, la dissimilarité de Bray-Curtis a été calculée pour les espèces qui ne sont pas considérées comme rares<sup>3</sup>. Cela correspond aux 11 espèces de diatomées suivantes (dans l'ordre alphabétique) : *Chaetoceros sp.1*, *Chaetoceros sp.2*, *Coscinodiscus sp.1*, *Coscinodiscus wailesii*, *Odontella sinensis*, *Proboscia alata*, *Pseudonitzschia delicatissima*, *Pseudonitzschia seriata*, *Rhizosolenia imbricata*, *Thalassiosira spp.1*, *Thalassionema nitzschioides*

Grâce à la sélection sur les espèces pour le calcul de la distance, 25% des paires d'échantillons ont une dissimilarité inférieure à 0,5 ; et la médiane de la distribution des mesures

2. Le terme "distance" est communément utilisé mais il faudrait plutôt utilisé le terme de "dissemblance". En effet, pour être considérée comme une métrique de distance au sens strict du terme, la métrique doit satisfaire l'*inégalité triangulaire*, ce qui n'est pas le cas de la métrique de Bray-Curtis. En revanche, cette métrique de dissimilarité est particulièrement utilisée en écologie sur les données d'abondance (Legendre et Legendre, 2012).

3. Le terme rare est utilisée pour les espèces qui sont présentes dans moins de 5% des échantillons, peu importe leurs abondances.

de dissimilarité est située à 0,82; c'est à dire que 50% des paires d'échantillons ont une dissimilarité inférieure à 0,82 et 50% des paires d'échantillons ont une dissimilarité supérieure à 0,82; pour une dissimilarité moyenne entre les paires d'échantillon de 0,74. Enfin, 2,06% des paires d'échantillons sont tout à fait identiques ( $dBC = 0$ ) et 3,9% des paires d'échantillons sont tout à fait dissemblables ( $dBC = 1$ ).

### **Classification des échantillons par l'algorithme UPGMA**

Une fois les distances de Bray-Curtis calculées, la matrice de distance est ensuite utilisée pour regrouper les échantillons en termes de composition au moyen d'une classification hiérarchique ascendante UPGMA (Jain et Dubes, 1988). La méthode UPGMA regroupe les échantillons en ordre de similarité (donc à partir des dissimilarités de Bray-Curtis les plus faibles). D'abord, l'algorithme groupe deux échantillons, puis il peut soit (i) grouper deux autres échantillons ou (ii) attacher un échantillon à un groupe pré-formé, et ceci en fonction des dissimilarités jusqu'à former un arbre enraciné (appelé aussi dendrogramme) où tous les échantillons sont groupés en fonction de leur similarité. À chaque étape de la construction du dendrogramme, la distance de Bray-Curtis est recalculée : lorsqu'un groupe d'échantillon est formé, alors la dissimilarité de Bray-Curtis entre ce groupe et les autres échantillons est moyennée (c'est le regroupement par "lien moyen").

**Exemple de construction d'un dendrogramme *via* l'UPGMA :** Soit la matrice d'abondance de 4 espèces (A, B, C et D) sur 10 échantillons :

A	10	8	9	10	1	0	4	2	3	0
B	1	2	1	0	8	6	10	1	0	3
C	2	1	0	4	1	2	1	10	9	9
D	6	8	7	5	8	4	8	9	6	7

La matrice de dissimilarité de Bray-Curtis est donc :

Échantillons	1	2	3	4	5	6	7	8	9
2	0,16								
3	0,11	0,11							
4	0,11	0,26	0,22						
5	0,51	0,35	0,49	0,62					
6	0,55	0,55	0,66	0,61	0,27				
7	0,43	0,29	0,40	0,52	0,12	0,37			
8	0,46	0,41	0,49	0,46	0,45	0,59	0,47		
9	0,41	0,46	0,49	0,35	0,56	0,60	0,51	0,15	
10	0,53	0,47	0,56	0,53	0,41	0,42	0,48	0,17	0,19

Les plus petites dissimilarités sont trouvées pour les échantillons n°2-3 (0,11) et 1-4 (0,11).

Les deux premières étapes regroupent donc les échantillons 2 et 3 ensemble et 1 et 4 ensemble.

La distance recalculée entre ces deux premiers ensemble et le reste des échantillons est donc :

Échantillons	5	6	7	8	9	10
1-4	0,565	0,58	0,475	0,46	0,38	0,53
2-3	0,42	0,605	0,345	0,45	0,475	0,515

Ce qui donne la matrice de distance suivante :

Échantillons	1-4	2-3	5	6	7	8	9
2-3	0,16						
5	0,565	0,42					
6	0,58	0,605	0,27				
7	0,475	0,345	0,12	0,37			
8	0,46	0,45	0,45	0,59	0,47		
9	0,38	0,475	0,56	0,60	0,51	0,15	
10	0,53	0,515	0,41	0,42	0,48	0,17	0,19

Sur cette matrice moyennée, la plus petite dissimilarité est maintenant entre les échantillons n°5 et 7. Ce sont ces deux échantillons qui seront groupés à la prochaine étape de la construction du dendrogramme.

### Justification du choix de l'UPGMA et de la distance de Bray-Curtis

Le choix de l'algorithme de classification et de la métrique de Bray-Curtis pour effectuer cette classification a été décidé au moyen des *corrélations cophénétiques*, comme conseillé par Legendre et Legendre (2012). Les principales distances conseillées pour le calcul de similarités (ou de dissimilarités) à partir de données d'abondance sont (Legendre et Legendre, 2012) : la dissimilarité de Bray-Curtis, la distance du  $\text{Khi}^2$ , distance de Gower

TABLE 5.1 – Table des corrélations cophénétiques (Legendre et Legendre, 2012) pour les métriques de distances ou de dissimilarité (en ligne) et les algorithmes de regroupement (en colonne). La valeur de corrélation la plus forte est mise en évidence en gras. Les distances ont été calculées sur les données d'abondance CPR, les corrélations cophénétiques ont été réalisées entre la distance calculée sur les données CPR et les distances qui résultent de la classification choisie.

Algorithme → ↓ Distance	Lien simple	Lien moyen	Lien complet
Bray-Curtis	0,53	<b>0,98</b>	0,89
Khi <sup>2</sup>	0,58	0,83	0,76
Gower	0,95	0,95	0,88
Euclidienne	0,96	0,96	0,91
Mahalanobis	0,92	0,89	0,86

(symétrique, quantitatif, Gower (1971)), le coefficient de Jaccard, la distance euclidienne, la distance de Mahalanobis. Les algorithmes de regroupement testés sont : le lien moyen (UPGMA, ci-dessus), le lien simple<sup>4</sup>, le lien complet<sup>5</sup>. Les méthodes de type "*k-means*" ont été exclues car un nombre de groupes doit être déterminé en amont, et ceci n'est pas applicable à notre problématique. Les résultats des *corrélations cophénétiques* sont présentés en Table 5.1. La meilleure corrélation, i.e., la distance et l'algorithme de regroupement qui respectent le plus la structuration initiale des données, est trouvée pour une dissimilarité de Bray-Curtis et un algorithme de regroupement moyen.

À ce stade de l'étude, nous avons donc obtenu une classification des données CPR, c'est-à-dire que nous avons établi le lien entre les échantillons CPR en terme de composition similaires en diatomées. Il faut maintenant définir le nombre de groupes idéal (c'est-à-dire le nombre d'assemblages), ainsi que ce qui caractérise ce groupe d'échantillons en termes de composition. Ceci peut-être réalisé à partir de la méthode d'Indicator Value développée par Dufrière et Legendre (1997).

### 5.1.1.2 La méthode Indicator Value de Dufrière et Legendre (1997)

Rappelons que l'objectif ici est de caractériser la composition des regroupements effectués sur les données CPR de façon à définir quelles espèces ou taxa sont indicateurs (défini ensuite) de chacun des groupes de données CPR.

4. La plus petite distance (ou dissimilarité) entre deux groupes formés, ou entre un groupe et un nouvel élément à grouper, est conservée.

5. Le lien complet est aussi appelé lien "maximum". Dans cet algorithme de regroupement, la plus grande distance (ou dissimilarité) entre deux groupes formés, ou entre un groupe et un nouvel élément à grouper, est conservée.

## Métriques de l'IndVal et calculs

La méthode Indicator Value (ou IndVal) de Dufrêne et Legendre (1997) est une approche qui permet d'identifier, dans un groupe d'échantillons, la ou les espèces qui sont *indicatrices* de ces échantillons. Elle a été appliquée avec succès dans un grand nombre de disciplines et pour diverses thématiques (liste non exhaustive) : écologie terrestre (Chen *et al.*, 2010), écologie des océans (Darnis *et al.*, 2008), médecine (Meadow *et al.*, 2014), biochimie (Schröder *et al.*, 2015), changement climatique (Connor *et al.*, 2013), pollutions des systèmes (Nahmani *et al.*, 2006).

Cette méthode se déroule en deux étapes : les échantillons sont groupés par similarité de composition (étape ci-dessus, Section 5.1.1.1), puis la métrique "Indicator Value" est calculée en fonction de cette classification, ce qui permet aux espèces indicatrices des groupes d'être identifiées. Cette méthode a été choisie pour cette étude car elle est particulièrement efficace avec des données d'abondance. En effet, elle utilise l'abondance et l'occurrence des espèces<sup>6</sup>. Elle est basée sur une approche asymétrique simple qui combine deux critères : la *fidélité* et la *spécificité* d'une espèce pour un groupe d'échantillons déterminé auparavant.

La fidélité d'une espèce est à son maximum (i.e., prend la valeur de 1,0) lorsque cette espèce est présente uniquement dans un groupe d'échantillon (les autres groupes d'échantillons sont donc exempts de celle-ci) et la spécificité est à son maximum (i.e., prend la valeur de 1,0) lorsque cette espèce est présente dans tous les échantillons de son groupe d'échantillon. Cela se traduit par les Équations 5.2 et 5.3 :

**Fidélité :**

$$A_{sg} = \frac{N_{\text{individu}_{sg}}}{N_{\text{individu}_s}} \quad (5.2)$$

**Spécificité :**

$$B_{sg} = \frac{N_{\text{echantillons}_{sg}}}{N_{\text{echantillons}_g}} \quad (5.3)$$

Où  $A_{sg}$  est la mesure de fidélité avec  $N_{\text{individu}_{sg}}$  le nombre moyen d'individu de l'espèce  $s$  dans le groupe d'échantillon  $g$  et  $N_{\text{individu}_s}$  est la somme des nombres moyens des individus de l'espèce  $s$  dans tous les groupes de sites définis en amont.

et  $B_{sg}$  est la mesure de spécificité où  $N_{\text{echantillons}_{sg}}$  est le nombre d'échantillons dans le groupe d'échantillon  $g$  où l'espèce  $s$  est présente et  $N_{\text{echantillons}_g}$  est le nombre d'échantillon total dans le groupe d'échantillon  $g$ .

<sup>6</sup> Ici, le terme "espèce" est utilisé, mais la méthode IndVal peut aussi être appliquée sur des groupements d'espèces.

Ainsi, la mesure de l’Indicator Value est définie comme (Équation 5.4) :

$$\text{IndVal}_{sg} = A_{sg} * B_{sg} \quad (5.4)$$

Où  $\text{IndVal}_{sg}$  est la valeur indicatrice de l’espèce  $s$  dans le groupe d’échantillon  $g$  et  $A_{sg}$  est la mesure de fidélité (Équation 5.2)  $B_{sg}$  est la mesure de spécificité (Équation 5.3). Notons que la valeur de  $\text{IndVal}_{sg}$  peut être multipliée par 100 pour avoir des valeurs en pourcentage.

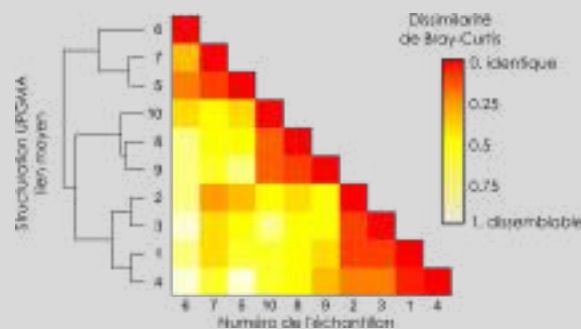
Ainsi, nous obtenons une valeur indicatrice pour chaque espèce  $s$  et pour chaque groupe d’échantillons  $g$ . La valeur indicatrice de l’espèce  $s$  est ensuite définie comme la valeur maximale à travers des groupes d’échantillons  $g$  des  $\text{IndVal}$  de l’espèce  $s$  (Équation 5.5, Dufrêne et Legendre (1997)) :

$$\text{IndVal}_s = \max(\text{IndVal}_{sg}) \quad (5.5)$$

**Exemple de calcul de l'IndVal :** Soit la matrice d'abondance de 4 espèces (A, B, C et D) sur 10 échantillons :

	1	2	3	4	5	6	7	8	9	10
A	10	8	9	10	1	0	4	2	3	0
B	1	2	1	0	8	6	10	1	0	3
C	2	1	0	4	1	2	1	10	9	9
D	6	8	7	5	8	4	8	9	6	7

Ce qui donne une structuration, selon les dissimilarités de Bray-Curtis, suivante :



Où la matrice colorée représente les dissimilarités de Bray-Curtis (blanc étant tout à fait dissemblable et rouge tout à fait identique) et le dendrogramme à droite est la structuration des données selon un algorithme UPGMA et un lien moyen.

Trois groupes peuvent être distingués selon le dendrogramme : les échantillons 1 à 4 appartiennent au groupe 1, 5 à 7 au groupe 2, et 8 à 10 au groupe 3. Selon ces groupes définis, la fidélité (Équation 5.2) et la spécificité (Équation 5.3) peuvent être calculées :

<b>Fidélité</b>	1	2	3	<b>Spécificité</b>	1	2	3
A	1,00	0,67	0,67	A	0,74	0,13	0,13
B	0,75	1,0	0,67	B	0,1	0,77	0,13
C	0,75	1,0	1,0	C	0,14	0,11	0,75
D	1,0	1,0	1,0	D	0,32	0,33	0,36

En multipliant ces deux indices (Équation 5.4), nous obtenons la valeur indicatrice pour chaque espèce, dans chaque groupe d'échantillons (l'IndVal<sub>s</sub> de l'équation 5.5 est mis en gras) :

<b>IndVal</b>	1	2	3
A	<b>0,74</b>	0,09	0,09
B	0,07	<b>0,77</b>	0,09
C	0,11	0,11	<b>0,75</b>
D	0,32	0,33	<b>0,36</b>



### Interprétation de l’IndVal

L’interprétation de l’IndVal se fait au travers des trois métriques (i.e., fidélité, spécificité et valeur indicatrice). En effet, une espèce peut être très spécifique, mais avec une abondance faible, ce qui donnera une IndVal faible ; et *a contrario* elle peut être très abondante mais peu spécifique, ce qui donnera également une IndVal faible. Il faut donc aussi analyser les métriques intermédiaires pour comprendre le rôle de l’espèce dans le groupe d’échantillons.

Afin de tester la robustesse du calcul de l’IndVal, Dufrêne et Legendre (1997) proposent le calcul d’une probabilité (*p-value*) qui est obtenue en permutant aléatoirement les échantillons au sein des groupes d’échantillons (Dufrêne et Legendre, 1997; Edgington et Onghena, 2007). En effet, la permutation des échantillons, initialement classés en groupes d’échantillons de compositions similaires, crée une destructuration du jeu de données initial (Dufrêne et Legendre, 1997). Cette probabilité estime les chances d’obtenir une valeur d’IndVal pour une espèce donnée aussi élevée si elle était répartie aléatoirement au sein des groupes d’échantillons. La *p-value* permet donc d’évaluer la robustesse du calcul de l’IndVal pour chaque espèce à travers différentes possibilités d’assemblages qui sont créées artificiellement par permutations des échantillons.

**Exemple d'interprétation de l'IndVal :**

La composition en espèces pour les 10 échantillons a permis de mettre en évidence qu'il existe trois structures récurrentes, c'est-à-dire trois assemblages potentiels. La mesure de fidélité, pour les quatre espèces et les trois assemblages ne permet pas de tirer des conclusions. En effet, il existe peu d'échantillons où les espèces sont absentes, ce qui a pour conséquences d'avoir une fidélité élevée de toutes les espèces à travers les 3 assemblages. Ici, c'est donc la spécificité permet de faire des conclusions :

- L'espèce A est très spécifique à l'assemblage n°1 - ce qui a pour conséquence qu'elle soit indicatrice de cet assemblage, avec une IndVal = 0,74
- L'espèce B est très spécifique à l'assemblage n°2 - ce qui a pour conséquence qu'elle soit indicatrice de cet assemblage, avec une IndVal = 0,77
- L'espèce C est très spécifique à l'assemblage n°3 - ce qui a pour conséquence qu'elle soit indicatrice de cet assemblage, avec une IndVal = 0,75
- En revanche, l'espèce D est très fidèle mais peu spécifique : en effet, sa spécificité est très similaire pour les trois assemblages (avec des valeurs de 0,32 ; 0,33 et 0,36). Sa IndVal maximale est trouvée pour l'assemblage 3, avec une valeur de IndVal de 0,36, mais ce n'est pas significatif : la probabilité d'obtenir une IndVal aussi élevée en permutant les échantillons est forte (ici,  $p=0,79$ ), ce qui était prévisible puisque l'IndVal de cette espèce est quasiment identique pour les trois groupes d'échantillons.

Ici, nous interprétons que les assemblages n°1, 2 et 3 sont respectivement caractérisés par les espèces A, B et C. L'espèce D est une espèce dite "*eurytopique*" : elle a une grande capacité de dispersion, qui est liée à de grandes gammes d'optimum de développement, elle est donc retrouvée, dans des abondances similaires, dans les trois assemblages.

**Choix du nombre de groupes d'échantillons pour le calcul de l'IndVal**

Pour que la méthode Indicator Value donne des résultats probants, il faut, en amont, avoir défini correctement les groupes d'échantillons et le nombre de groupes d'échantillons à analyser. Le nombre idéal de groupes d'échantillons doit avoir à la fois un sens statistique et un sens écologique (Dufrene et Legendre, 1997; Legendre et Legendre, 2012). La méthode IndVal peut permettre d'évaluer à la fois le sens écologique (*via* le résultat de l'IndVal), mais également le sens statistique, grâce au calcul de la probabilité d'obtenir des Indval aussi élevées que celles observées en réalisant des permutations. En pratique, les 5977 échantillons ont été classés grâce à l'algorithme UPGMA et une dissimilarité de Bray-Curtis (Section 5.1.1.1, page 170). Le dendrogramme obtenu (aussi appelé typologie des données) a été divisé en deux groupes d'échantillons, puis l'IndVal (et la probabilité associée) a été calculé. Cette étape est recommencée pour une division en trois groupes d'échantillons et ainsi de suite.

Le nombre de groupes d’échantillons idéal (et donc, le nombre d’assemblages) est obtenu lorsque :

- i le calcul de la valeur de l’IndVal n’augmente plus significativement pour un nombre d’assemblages plus élevé ;
- ii il y a au moins une espèce indicatrice par assemblage ;
- iii les valeurs de l’IndVal dépassent 0,25 ou 25% (Dufrière et Legendre, 1997) ;
- iv les assemblages sont reconnus dans la littérature.



#### Résumé : analyse des données *in-situ*

Les données du *Continuous Plankton Recorder* détaillent l’abondance de 72 espèces de diatomées. Ces espèces peuvent former des assemblages : elles co-occurrent sur un ensemble de sites. Ces assemblages ont été caractérisés grâce à l’utilisation de la méthode IndVal (Dufrière et Legendre, 1997) : les échantillons *Continuous Plankton Recorder* ont été groupés par composition similaire, puis une valeur indicatrice a été calculée pour déterminer les espèces indicatrices dans ces groupes d’échantillons. Un nombre d’assemblages optimal a été déterminé (voir les résultats, Section 5.2.1, à partir de la page 184). Ce travail a été réalisé dans l’idée de proposer une méthode qui pourra être appliquée à d’autres bases de données d’abondance, peu importe leur provenance (e.g., comptages manuels ou automatisés, comptages par *cytométrie en flux*), et pourra ainsi faire partie de la nouvelle chaîne de traitement de l’algorithme PHYSAT.

### 5.1.2 Calcul des $Ra(\lambda)$ dans la région de la Mer du Nord : étape 2

Parallèlement à l’analyse des données *in-situ* détaillée ci-dessus, une étape de traitement des données satellites associées (même période, même région) a été nécessaire, avant la mise en correspondance de ces deux types de données *via* les mesures coïncidentes. Dans cette étude, les niveaux 2 de luminances normalisées SeaWiFS ont été utilisés ( $nLw(\lambda)$ , NASA Goddard Space Flight Center (2014)) entre 1998 et 2004 dans la Mer du Nord et une partie de la Manche (entre  $-4^{\circ}$ Ouest et  $9^{\circ}$ Est et entre  $48$  et  $60^{\circ}$ Nord). Les données satellites de niveau 2 ont une résolution de  $1\text{km}^2$ .

Les données du capteur de couleur de l'océan SeaWiFS ont été sélectionnées car ce sont elles qui possèdent le plus de mesures coïncidentes avec les données CPR sur une période de 7 ans.

Les masques standards ont été appliqués sur les données SeaWiFS pour supprimer l'effet des nuages, de la glace, des terres et du sun-glint. Comme pour la procédure traditionnelle de PHYSAT (Alvain *et al.*, 2005), les pixels ayant une épaisseur optique d'aérosol (mesurée à 865nm) supérieure à 0,15 ont également été exclus afin d'assurer une correction atmosphérique de la meilleure qualité possible.

PHYSAT est une méthode qui a prouvé son efficacité dans les eaux des océans ouverts. En Mer du Nord, la zone est plus complexe en termes de classification optique de l'eau. En effet, il peut rester des cas où le phytoplancton n'est pas le seul facteur qui influence les mesures de couleur de l'océan dans l'algorithme classique Cas I / Cas II (IOCCG *report*, 2000; Vantrepotte *et al.*, 2012). Pour pallier ceci et utiliser le maximum de pixels disponibles, la classification optique proposée par Vantrepotte *et al.* (2012) a été appliquée dans cette zone pour extraire les eaux dont les propriétés optiques sont influencées par le phytoplancton (et leurs matières associées) et masquer les eaux dont les propriétés sont influencées par les matières terrigènes (Vantrepotte *et al.*, 2012).

Les anomalies de luminance sont ensuite calculées pour les pixels sélectionnés par l'équation :

$$Ra(\lambda) = \frac{nLw(\lambda)_{obs}}{nLw_{ref}(\lambda, Chl-a)} \quad (5.6)$$

où  $Ra(\lambda)$  est l'anomalie de luminance pour chaque longueur d'onde, tenant compte du ratio entre le spectre de couleur  $nLw(\lambda)$  observé sur un pixel par rapport au spectre de référence moyen  $nLw_{ref}(\lambda, Chl-a)$  donné pour la concentration en Chlorophylle-*a* estimée sur ce pixel.

Comme précisé dans le Chapitre 2 (page 27), ce genre d'étude nécessite une table de référence  $nLw_{ref}(\lambda, Chl-a)$  spécifique. Cette table de référence a été construite à partir des mêmes sélections (masques et classifications optiques) à partir d'un grand nombre de pixels pour des pas de chlorophylle allant d'une concentration estimées de 0,04 à 4,0  $mg.m^{-3}$ .

Les classifications "**SOM**" et "*Phenology*" ont été appliquées dans les conditions classiques, détaillées au Chapitre 2 (page 27), sur la base de données d'anomalies de luminance spécifique à la Mer du Nord. Ainsi, l'algorithme SOM a permis d'obtenir 100 classes de  $Ra(\lambda)$  ayant des formes et amplitudes différentes et l'algorithme "*Phenology*" a permis

de réduire ce nombre à 85 classes de  $Ra(\lambda)$  ayant des formes, amplitudes et saisonnalités différentes. Ce sont ces classes d'anomalies de luminance qui vont être utilisées pour évaluer les liens potentiels entre leur présence et la détection d'assemblage caractéristiques *in-situ*.



### Résumé : analyse des anomalies de luminance

Une base de données de 85 classes de  $Ra(\lambda)$ , spécifiques à la Manche-Mer du Nord, a été construite à partir des données SeaWiFS de niveau 2 pour les eaux dont les propriétés optiques sont influencées par le phytoplancton. Ces données s'étendent sur la période de 1998 à 2004.

## 5.1.3 Essai de labellisation et validation des labels : étape 3

### 5.1.3.1 Procédure classique de calibration et validation

Après avoir réalisé les étapes 1 (analyse des données *in-situ*) et 2 (analyse des anomalies de luminance en Mer du Nord), il est possible d'observer des correspondances empiriques potentielles entre les données *in-situ* et les données satellites. Pour cela, les données du CPR ont été associées au pixel satellite correspondant (pavé de pixel de  $3 \times 3$  autour du "matchup central", correspondant à une aire de  $9\text{km}^2$ ) avec un décalage de  $\pm 3\text{h}$  entre l'heure de la mesure CPR et le passage du satellite au dessus de ce point, comme conseillé dans Bailey et Werdell (2006) et utilisé dans les versions précédentes de PHYSAT.

Les mesures coïncidentes entre ces deux jeux de données sont au nombre de 450, soit 7% du jeu de données CPR. Elles ont été divisées en deux sous-sets : un sous-set de calibration (80% des mesures coïncidentes, soit 360 mesures) et un sous-set de validation (20% des mesures coïncidentes, soit 90 mesures). Ce tirage aléatoire est stratifié c'est-à-dire que chaque assemblage (décrits dans les résultats) a son propre lot de mesures coïncidentes de calibration et de validation (voir Table 5.5 dans la partie Résultats).

L'étape de calibration permet de labelliser une classe de  $Ra(\lambda)$ , c'est-à-dire de lui associer empiriquement un nom selon l'expertise *in-situ*, si au moins 50% des observations de cette classe sont trouvées pour le même assemblage *in-situ* dans le sous-set de calibration (seuil classique de PHYSAT, Alvain *et al.* (2005), Ben Mustapha *et al.* (2014)).

Ensuite, les  $Ra(\lambda)$  qui remplissent les conditions pour être labellisés ont été validés avec le sous-set de validation, qui est indépendant, c'est-à-dire contient des données non utilisées pour la calibration. Dans la procédure de validation, toutes les classes de  $Ra(\lambda)$  labellisées

pour un assemblage sont considérées. L'occurrence des classes de  $Ra(\lambda)$  labellisées (une ou plusieurs classes) est vérifiée dans le sous-set de validation. Idéalement, dans le sous-set de validation, les classes de  $Ra(\lambda)$  labellisées doivent être associées avec le même assemblage de diatomées que celui qui leur a été attribué durant l'étape de calibration (ceci correspond à une validation de 100%). Le score de validation sera en revanche diminué si une classe de  $Ra(\lambda)$  a été associée à l'assemblage  $x$  durant la calibration, mais correspond à un assemblage  $y$  dans le sous-set de validation. De plus, une erreur d'association est également comptabilisé lorsque qu'une classe de  $Ra(\lambda)$  labellisée n'est pas présente dans le sous-set de validation. Dans ce cas, ce n'est pas vraiment une erreur d'association au sens premier du terme, mais la validation n'est pas applicable.

### 5.1.3.2 Répétitions de la procédure de calibration et de validation

En raison de la variabilité spatio-temporelle qui existe dans les données CPR, et du fait que les mesures coïncidentes ne sont pas toujours également réparties dans le temps et l'espace, nous avons répliqué les procédures de calibration et validation. En effet, cette procédure a été appliquée 1200 fois, et ainsi, 1200 sous-sets de calibration et de validation ont pu être évalués dans le cadre de ces procédures distinctes. Ceci permet de calculer une valeur moyenne de validation et de s'affranchir des effets de la division aléatoire des données en deux sets 80/20%.

Une répétition des procédures de calibration permet d'évaluer la robustesse de la calibration *via* le calcul, *a posteriori*, d'un taux de calibration pour chaque classe de  $Ra(\lambda)$  (concept déjà appliqué en couleur de l'océan, e.g., Craig *et al.* (2012), Bracher *et al.* (2015), Brewin *et al.* (2015)). Cela permet de retenir uniquement les labellisations robustes, c'est-à-dire celles qui ont été trouvées quels que soient les sous-sets de calibration utilisés. Ainsi, pour chaque procédure distincte, des sous-sets de calibration et de validation sont choisis de manière aléatoire, puis une calibration est réalisée et la validation de cette calibration est calculée. Le nombre de répétitions de la procédure ( $n_{procédure} = 1200$ ) a été défini de telle sorte que chaque mesure coïncidente se retrouve au moins une fois aléatoirement dans le sous-set de validation.



### Résumé : associations entre $Ra(\lambda)$ et assemblage *in-situ*

Les classes  $Ra(\lambda)$  disponibles sont labellisées si plus de 50% d’un même assemblage *in-situ* est observé pour cette classe dans l’étape de calibration, basée sur les mesures coïncidentes. Une validation est calculée sur une partie des données qui n’a pas servie pour la calibration. Cette étape est répétée pour calculer un taux de calibration et des scores de validation moyens afin d’assurer la robustesse des associations empiriques conservées au final.

## 5.2 Résultats

### 5.2.1 Espèces indicatrices & assemblages de diatomées

La Figure 5.2 représente la structuration des données CPR à partir de l’algorithme UPGMA et d’une distance de Bray-Curtis. En accord avec cette structuration, la méthode IndVal a été appliquée depuis le niveau de regroupement 1 (L1) jusqu’au niveau 6 (L6). Les résultats des valeurs indicatrices à travers ces niveaux de regroupement sont données en Table 5.2. Dans cette Table, les espèces rares ont été enlevées par soucis de clarté, mais l’intégralité des résultats est listée en Annexe 15 (page 325).

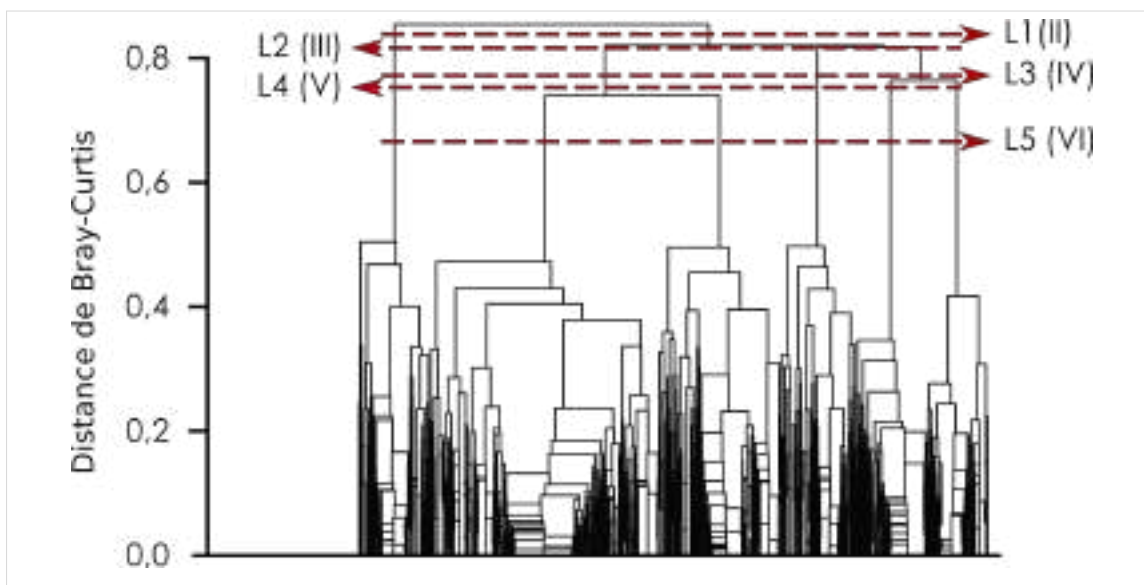


FIGURE 5.2 – Structuration des échantillons CPR obtenue à partir d’un algorithme UPGMA (lien moyen) et d’une dissimilarité de Bray-Curtis, calculés sur les données d’abondance des diatomées. Chaque branche de ce dendrogramme représente un échantillon CPR.  $L_x$  ( $L$  pour *level*) correspond à  $x$  niveaux de regroupement successifs, ce qui donne un nombre d’assemblages indiqué entre parenthèses par un chiffre romain.

TABLE 5.2 – Ce tableau représente une combinaison des résultats de la classification des échantillons CPR et du calcul des valeurs indicatrices de la méthode IndVal. Chaque colonne représente un niveau de regroupement (L+chiffre arabe), pour lequel on délimite donc  $n$  assemblages (chiffre romain entre parenthèses). Les valeurs indicatrices (IndVal, en pourcentage) sont obtenues pour les niveaux successifs allant du premier niveau (L1, deux assemblages) au sixième niveau (L6, sept assemblages). Pour un taxon, à un niveau de regroupement spécifique (i.e., pour une case particulière du tableau), l'IndVal est donné ainsi que le numéro de l'assemblage pour lequel cet IndVal a été maximum (indiqué par le chiffre romain en parenthèses). Les résultats pour les six taxa les plus abondants sont donnés par soucis de clarté, mais une Table complète peut être trouvée en annexe 15. Les valeurs d'Indval maximales pour chaque taxon ont été mises en évidence par une astérisque, les valeurs significatives (selon la procédure de Dufrêne et Legendre (1997)) sont signalées en gras. Une colonne est mise en évidence en bleu : elle correspond au niveau de regroupement et aux valeurs d'IndVal conservés pour la suite de l'étude (voir page 186 pour la justification du choix).

	L1 (II)	L2 (III)	L3 (IV)	L4 (V)	L5 (VI)	L6 (VII)
<i>Chaetoceros sp.1</i>	40,25 (I)	33,9 (II)	<b>27,8 (III)</b>	<b>38,3 (II)</b>	<b>61,6 (II)*</b>	<b>59,3 (II)</b>
<i>Chaetoceros sp.2</i>	<b>43,83 (I)</b>	31,4 (I)	<b>70,2 (I)*</b>	<b>65,1 (I)</b>	<b>60,6 (I)</b>	<b>59,4 (I)</b>
<i>T. nitzschioides</i>	31,93 (I)	30 (II)	<b>83,6 (II)*</b>	<b>81,1 (III)</b>	<b>77,8 (III)</b>	<b>77,1 (III)</b>
<i>P. seriata</i>	27,83 (I)	23,17 (II)	87,19 (IV)*	82,86 (IV)	<b>74,8 (IV)</b>	<b>71,6 (IV)</b>
<i>Thalassiosira sp.</i>	<b>92,4 (II)*</b>	<b>83,9 (III)</b>	<b>76,7 (III)</b>	<b>72,5 (IV)</b>	<b>67,9 (V)</b>	<b>67 (V)</b>
<i>R. imbricata</i>	23,92 (I)	49,57 (I)	44,27 (I)	<b>86,1 (V)*</b>	<b>82,9 (VI)</b>	<b>81,9 (VI)</b>

### Description détaillée du premier niveau de regroupement (deux assemblages)

Au premier niveau de regroupement (L1 sur la Figure 5.2 et première colonne de la Table 5.2), deux espèces principales de diatomées étaient indicatrices chacune d'un assemblage (*Chaetoceros sp2* et *Thalassiosira spp.*). Pour le premier assemblage, *Chaetoceros sp2* est associé le plus souvent à *Rhizosolenia hebetata* et à *Skeletonema costatum* (voir Annexe 15, page 325). Pour le second assemblage, *Thalassiosira spp.* est associé le plus souvent à *Odentella sinensis* et *Proboscia alata* (voir Annexe 15).

### Niveaux de regroupement successifs

En passant des niveaux successifs de regroupement L1 à L5 (Table 5.2), les valeurs d'Ind-Val augmentent pour les 6 espèces les plus représentées dans la région (ou décroissent non significativement). Cela signifie qu'au fil des niveaux de regroupement successifs, les diatomées deviennent de plus en plus caractéristiques d'un groupe d'échantillons particulier. En revanche, ce n'est pas le cas pour l'espèce *Thalassiosira spp.* pour laquelle l'IndVal décroît significativement au cours des niveaux successifs de regroupement. À partir du



niveau de regroupement n°6, l'IndVal stagne pour toutes les espèces, et les assemblages qui caractérisent ces espèces ne sont plus modifiés.

**Le cas particulier de *Thalassiosira spp.***

*Thalassiosira spp.* semble être un taxon "particulier". En effet, sa IndVal décroît significativement pour les niveaux successifs de regroupement 1 à 5, passant de 92,4 à 67,9%. La mesure de spécificité pour cette espèce présente toujours une valeur assez élevée, alors que la mesure de fidélité l'est moins. Cela veut dire qu'elle est toujours présente dans un grand nombre d'échantillons, mais qu'elle est plus abondante dans un groupe de sites particulier. Par exemple, au niveau de regroupement L5, distinguant 6 assemblages, les valeurs de spécificité, de fidélité et d'IndVal sont réparties selon la Table 5.3. Les valeurs de spécificité sont élevées (hormis pour l'assemblage n°4 qui ne comporte aucune occurrence de *Thalassiosira spp.*) puisque la valeur de spécificité minimale est de 25,6 (l'espèce est donc présente au moins dans un quart des échantillons de chaque assemblage). Cependant, la valeur de fidélité est nettement plus élevée pour l'assemblage n°5, ce qui signifie que *Thalassiosira spp.* est beaucoup plus abondante pour cet assemblage.

TABLE 5.3 – Répartition des valeurs de spécificité, de fidélité et d'IndVal du taxon *Thalassiosira spp.* au niveau de regroupement L5, distinguant six assemblages (colonnes). Les valeurs sont données en pourcentage et les valeurs maximales sont signalées par un astérisque.

Assemblage n°	1	2	3	4	5	6
Spécificité	32,7	57,6	45,4	0	85*	25,6
Fidélité	5,8	18,9	21,4	0	97,5*	5,5
IndVal	1,9	10,9	9,6	0	82,9*	1,4

**Le cinquième niveau de regroupement (six assemblages)**

Le cinquième niveau de regroupement, permettant de différencier six assemblages de diatomées, a montré le meilleur compromis entre le nombre d'assemblages, les espèces indicatrices et les valeurs indicatrices de la méthode IndVal. De plus, au-delà de six assemblages, les groupes d'échantillons deviennent trop similaires pour que des caractéristiques particulières s'en dégagent. Ceci est visible sur le dendrogramme de classification des échantillons CPR (Figure 5.2), où les cinq premiers niveaux de regroupement se distinguent à partir d'une distance de Bray-Curtis de 0,85 à 0,72 (ces 6 groupes d'échantillons sont donc fortement dissemblables) alors que le sixième niveau (7 assemblages) n'apparaît qu'à une distance de Bray-Curtis de 0,48.

Par conséquent, **six assemblages ont été retenus pour cette étude**, qui sont chacun caractérisés par une espèce indicatrice principale (avec une forte Valeur Indicatrice) et

TABLE 5.4 – Description des assemblages de diatomées issus de la méthode Indicator Value avec l'espèce indicatrice principale en gras, et un set d'espèces secondaires. Les valeurs entre parenthèses sont les IndVal de chaque espèce pour le groupe de sites considéré.

***Chaetoceros sp.2*** (60,6); *Rhizosolenia styliformis* (4,9); *Corethron hystrix* (2,15)  
***Chaetoceros sp.1*** (61,6); *Bacteriastrum sp.* (1,8); *Leptocylindrus sp.* (1,9)  
***Thalassionema nitzschoïdes*** (78,1); *Odontella sinensis* (16,1); *Fragillaria sp.* (5,3)  
***Pseudonitzschia seriata*** (74,8); *Proboscia alata* (18,7); *Coscinodiscus sp.* (4,4)  
***Thalassiosira spp.*** (67,9); *Skeletonema costatum* (15,6); *Asterionellopsis glacialis* (7,0)  
***Rhizosolenia imbricata*** (82,9); *Rhizosolenia hebetata* (8,7); *Paralia sulcata* (7,4)

un set d'espèces indicatrices secondaires (avec des valeurs indicatrices plus faibles, mais significatives). Ces six assemblages, conservés pour la suite de l'étude, sont décrits dans la Table 5.4.

Au total, 34 espèces ou taxa de diatomées sont caractéristiques de ces six assemblages (IndVal significative), et au moins trois espèces sont caractéristiques de chaque assemblage. Six espèces de diatomées ont leur IndVal qui dépasse le seuil de 25% préconisé par Dufrene et Legendre (1997). Ces six espèces sont aussi les six espèces de diatomées les plus abondantes en Mer du Nord. Ainsi, la classification des échantillons CPR basée sur les 11 diatomées principales a permis de mettre en évidence qu'à partir de six assemblages, les six diatomées les plus dominantes devenaient caractéristiques d'un assemblage.

Une fois cette première étape réalisée, les mesures coïncidentes ont été réalisées afin d'observer si des associations fortes pouvaient être trouvées avec les classes d'anomalies de luminance en Mer du Nord.

## 5.2.2 Procédures successives de calibration

Le nombre de mesures coïncidentes satellites disponibles pour chaque assemblage *in-situ* est présenté en Table 5.5. La procédure de calibration et validation a été réalisée 1200 fois de manière distincte afin de prendre en compte la variabilité naturelle des assemblages de diatomées (voir Section 5.1.3.1, page 182). La Figure 5.3 présente le nombre de fois où les classes de  $Ra(\lambda)$  ont été labellisées parmi les 1200 essais de calibration. Ce nombre de fois où une labellisation des classes de  $Ra(\lambda)$  a été possible, selon les critères définis précédemment, sera appelé "taux de calibration" par la suite.

Selon les critères de labellisation (Section 5.1.3.1), environ 30% des classes de  $Ra(\lambda)$  n'ont jamais été associées à un assemblage de diatomées (par exemple, le  $Ra(\lambda)$  n°7 en Figure 5.3). Ces  $Ra(\lambda)$  sont, à ce stade, soit absents des mesures coïncidentes, soit ils ne sont pas suffisamment présents pour être labellisés (moins de 5 mesures coïncidentes). Ensuite,

TABLE 5.5 – Nombre de mesures coïncidentes pour chaque assemblage, caractérisé par son espèce indicatrice (en gras), et un set d'espèces secondaires accompagnant l'espèce indicatrice déterminées par l'approche IndVal de Dufrêne et Legendre (1997). "M" est le nombre de mesures coïncidentes disponibles pour l'assemblage considéré, donnant un nombre de mesures coïncidentes "Cal" pour la recherche d'association avec les classes de Ra( $\lambda$ ) (80% de "M"), et un nombre "Val" de mesures coïncidentes pour calculer un taux de validation des associations trouvées (20% de "M").

Assemblage	M	Cal	Val
<b><i>Chaetoceros sp.2</i></b> , <i>Rhizosolenia styliiformis</i> , <i>Corethron hystrix</i>	52	41	11
<b><i>Chaetoceros sp.1</i></b> , <i>Bacteriastrum sp.</i> , <i>Leptocylindrus sp.</i>	72	57	15
<b><i>Thalassionema nitzschoïdes</i></b> , <i>Odontella sinensis</i> , <i>Fragillaria sp.</i>	93	74	19
<b><i>Pseudonitzschia seriata</i></b> , <i>Proboscia alata</i> , <i>Coscinodiscus sp.</i>	65	51	14
<b><i>Thalassiosira sp.</i></b> , <i>Skeletonema costatum</i> , <i>Asterionellopsis glacialis</i>	116	92	24
<b><i>Rhizosolenia imbricata</i></b> , <i>Rhizosolenia hebetata</i> , <i>Paralia sulcata</i>	52	41	11

33% des classes de Ra( $\lambda$ ) ont été associées à un seul assemblage de diatomées spécifique pour toutes les procédures de calibration (par exemple, le Ra( $\lambda$ ) n°15 en Figure 5.3).

Enfin, 37% des classes de Ra( $\lambda$ ) ont été associées successivement à différents assemblages au cours des procédures de calibration (par exemple, le Ra( $\lambda$ ) n°25 en Figure 5.3). Ces cas de "labellisations multiples" semblent possibles si l'on considère que les communautés associées aux assemblages de diatomées peuvent être les mêmes dans certains cas (voir la Section Discussion). Par exemple, le Ra( $\lambda$ ) n°57 a été labellisé dans moins de 20% des procédures de calibration et pour la moitié associé à l'assemblage *Thalassionema nitzschoïdes* et l'autre moitié associé à l'assemblage *Thalassiosira spp.*. Pour ce Ra( $\lambda$ ), l'information *in-situ* disponible à ce stade est insuffisante pour savoir s'il est possible de le caractériser en termes de phytoplancton. Ce Ra( $\lambda$ ) est probablement associable à un type phytoplanctonique non identifié pour le moment ou correspond à un type de spectre non spécifique selon l'étude théorique de Alvain *et al.* (2012) (cf. Section Discussion).

La plupart des classes de Ra( $\lambda$ ), pour lesquelles une correspondance a été trouvée, ont été associées en grande majorité à un assemblage de diatomées particulier (par exemple, le Ra( $\lambda$ ) n°50 en Figure 5.3). Ce résultat montre qu'il est indispensable de répéter la procédure de calibration. En effet, ce type de procédure répétitive permet d'apprécier la robustesse des associations trouvées et leurs récurrences. En effet, lorsqu'un Ra( $\lambda$ ) a un fort "taux de calibration", et que cette calibration permet de l'associer très souvent à un assemblage spécifique, la robustesse de l'association peut être considérée comme meilleure par rapport à une procédure de calibration simple.

Un seuil a été fixé pour définir une association "robuste" : si le taux de calibration dépasse

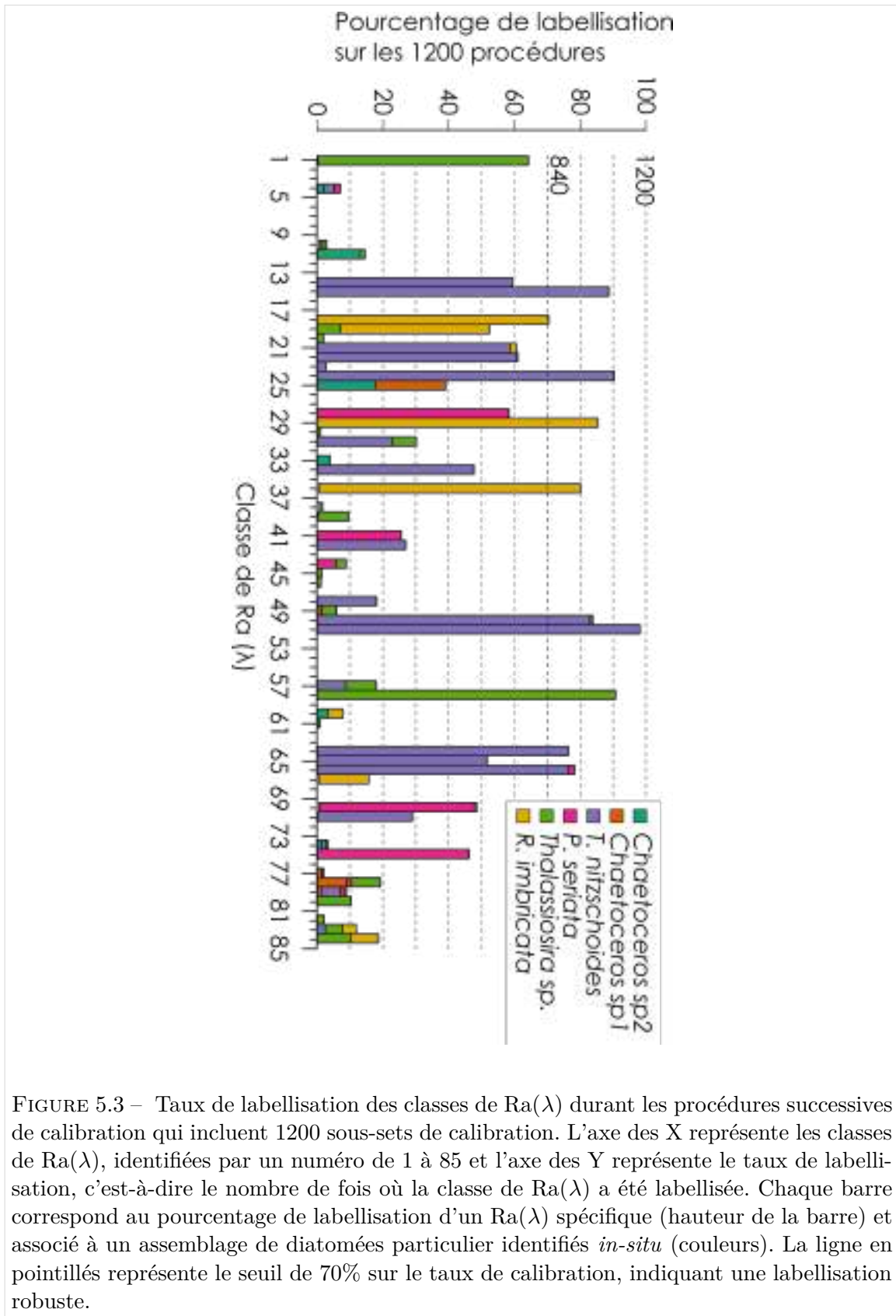


FIGURE 5.3 – Taux de labellisation des classes de  $Ra(\lambda)$  durant les procédures successives de calibration qui incluent 1200 sous-sets de calibration. L'axe des X représente les classes de  $Ra(\lambda)$ , identifiées par un numéro de 1 à 85 et l'axe des Y représente le taux de labellisation, c'est-à-dire le nombre de fois où la classe de  $Ra(\lambda)$  a été labellisée. Chaque barre correspond au pourcentage de labellisation d'un  $Ra(\lambda)$  spécifique (hauteur de la barre) et associé à un assemblage de diatomées particulier identifiés *in-situ* (couleurs). La ligne en pointillés représente le seuil de 70% sur le taux de calibration, indiquant une labellisation robuste.

70% pour une classe de  $Ra(\lambda)$  face à un assemblage spécifique (ce qui correspond à 840 procédures gagnantes dans cette étude), l'association est considérée comme robuste, c'est-à-dire que la labellisation a été trouvée peu importe les données considérées dans le sous-set de calibration.

En considérant ce seuil sur le taux de calibration, la procédure a permis de labelliser 11,7% des 85 classes de  $Ra(\lambda)$  disponibles. Parmi les six assemblages identifiés par l'analyse des données CPR, trois ont pu être associés de manière robuste à dix classes de  $Ra(\lambda)$ . Ainsi, les assemblages caractérisés par *Thalassionema nitzschooides*, *Thalassiosira spp.*, et *Rhizosolenia imbricata* ont été associés avec des classes de  $Ra(\lambda)$  spécifiques plus de 840 fois sur les 1200 procédures, ce qui n'est pas le cas des assemblages caractérisés par les deux taxa de *Chaetoceros spp.* et *Pseudonitzschia seriata*.

### 5.2.3 Spectres d'anomalies de luminance associés aux assemblages de diatomées

La Figure 5.4 présente les spectres référents des classes de  $Ra(\lambda)$  qui ont pu être associés empiriquement à la présence d'un assemblage spécifique de diatomées. Ces spectres de  $Ra(\lambda)$  sont caractérisés par des amplitudes, des formes et des saisonnalités variables. Bien que certains spectres semblent très similaires en termes de forme et d'amplitude, ils n'ont pas du tout la même *phénologie* (voir Figure 5.5), ce qui induit, selon l'hypothèse formulée en Chapitre 3 (page 55), que ce sont des spectres différents dans le cadre de notre étude (qui ne prend désormais plus en compte uniquement la forme et l'amplitude).

L'assemblage caractérisé par *Thalassionema nitzschooides* a été associé avec six classes de  $Ra(\lambda)$  à partir des mesures coïncidentes disponibles, ce qui représente 7,1% de la diversité des classes de  $Ra(\lambda)$  en Mer du Nord (voir Figure 5.4.a pour les spectres référents). Ces  $Ra(\lambda)$  peuvent être divisés en deux sous-groupes en accord avec leurs caractéristiques : forte anomalie à 412 nm et faible anomalie 443 nm.

L'assemblage caractérisé par *Thalassiosira spp.* a été associé à une classe de  $Ra(\lambda)$  (1,2% de la diversité des  $Ra(\lambda)$ , Figure 5.4.b pour le spectre référent). Ce spectre a une forme dite "plate", c'est-à-dire avec des valeurs d'anomalie de luminance similaires pour les cinq longueurs d'ondes.

Enfin, l'assemblage caractérisé par *Rhizosolenia imbricata* a été associé à trois classes de  $Ra(\lambda)$  (3,3% de la diversité totale des  $Ra(\lambda)$  dans la région, Figure 5.4.c pour les spectres référents). Ces spectres sont assez variables autant en forme qu'en amplitude.

À ce stade, dix classes de  $Ra(\lambda)$ , avec leurs amplitudes, formes et saisonnalités spécifiques, ont été associées à la présence *in-situ* de trois assemblages de diatomées différents.

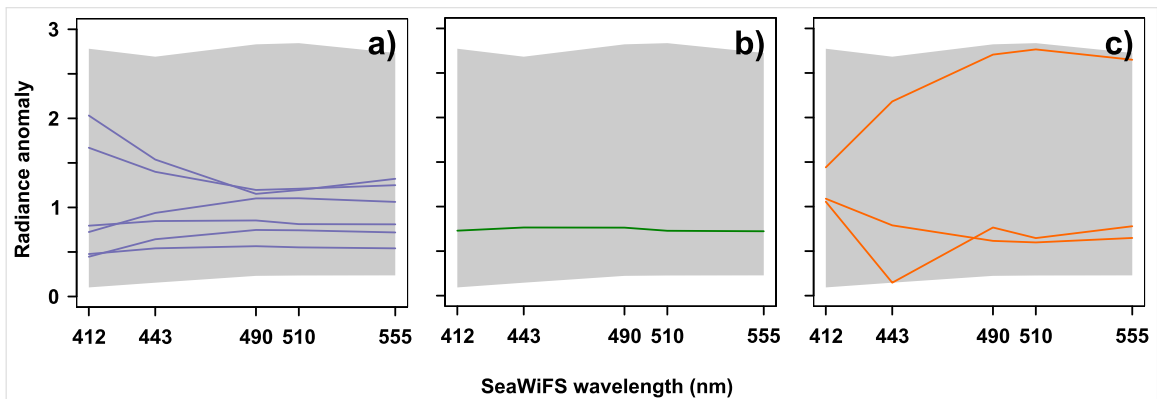


FIGURE 5.4 – Spectres d’anomalies de luminance labellisés (spectres de référence des classes de  $Ra(\lambda)$ ) pour l’assemblage caractérisé par a) *Thalassionema nitzschooides*, b) *Thalassiosira sp.* et c) *Rhizosolenia imbricata*. L’arrière-plan gris montre la diversité des spectres de  $Ra(\lambda)$  (minimum et maximum) de la Mer du Nord. L’axe des X représente les cinq longueurs d’onde SeaWiFS et l’axe des Y représente l’anomalie de luminance, en accord avec l’équation 2.1 au Chapitre 2 (page 27).

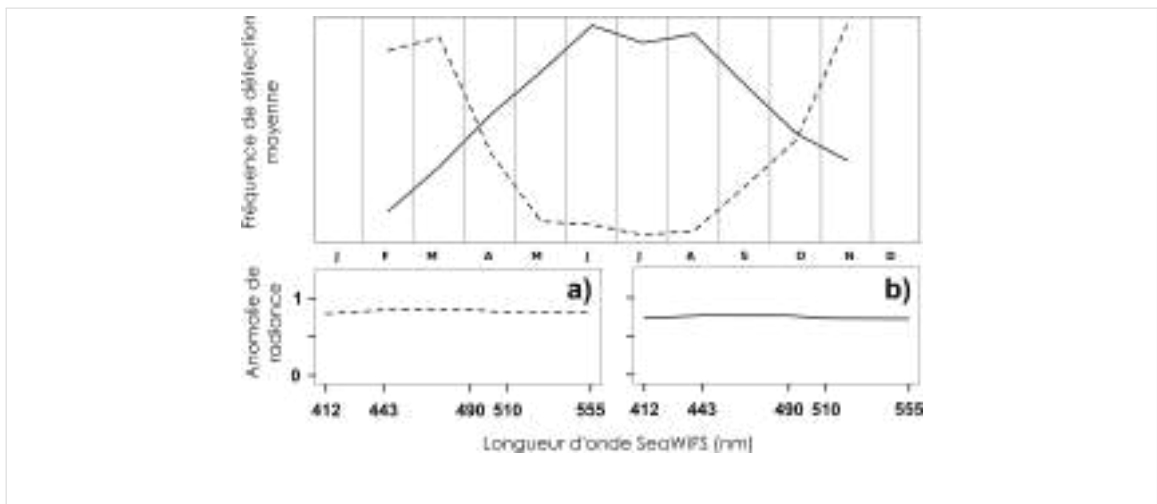


FIGURE 5.5 – Cycles saisonniers moyens de deux classes de  $Ra(\lambda)$ , associés empiriquement aux assemblages caractérisés par a) *Thalassionema nitzschooides* et b) *Thalassiosira spp.*. Ces deux spectres semblent identiques en ce qui concerne leur forme et leur amplitude, mais ils montrent deux saisonnalités différentes. Le  $Ra(\lambda)$  associé à *Thalassionema nitzschooides* (ligne pointillée) montre un pic de fréquence en hiver alors que le  $Ra(\lambda)$  associé à *Thalassiosira spp.* a un maximum de détection estival.

### 5.2.4 Validation

La validation se déroule après chaque étape de calibration au cours des 1200 procédures. Ainsi, 1200 validations intermédiaires sont obtenues, qui peuvent ensuite être moyennées. Dans cette Section, un exemple aléatoire de validation intermédiaire est présenté en détails, puis les scores de validation moyens sont listés.

### Vocabulaire : validation

La validation peut être présentée de quatre manières :

- **Validation intermédiaire** : validation classique qui concerne une procédure spécifique de calibration et validation avec une séparation du jeu de données en 80/20%.
- **Validation intermédiaire *a posteriori*** : validation **après sélection des associations robustes** qui concerne une procédure spécifique de calibration et validation. Une association est considérée comme robuste si le taux de calibration atteint 70% des procédures pour un assemblage spécifique.
- **Validation totale moyenne** : validation classique moyennée sur l'ensemble des procédures de calibration et validation.
- **Validation totale moyenne *a posteriori*** : validation classique moyennée sur l'ensemble des procédures de calibration et validation après sélection des associations robustes.

### Exemple de validation intermédiaire classique et *a posteriori*

Pour un set de calibration (80%) intermédiaire choisi au hasard comme exemple, 19 classes de  $Ra(\lambda)$  ont été associées avec les assemblages de diatomées, c'est-à-dire que pour ces 19 classes, au moins la moitié des mesures coïncidentes étaient caractérisées par un seul et même assemblage. Dans cet exemple précis, les assemblages définis précédemment ont été associés à :

- l'assemblage *Chaetoceros spp2* n'a pas été associé à une classe de  $Ra(\lambda)$
- l'assemblage *Chaetoceros spp1* a été associé à une classe de  $Ra(\lambda)$
- l'assemblage *Thalassionema nitzschoides* a été associé à neuf classes de  $Ra(\lambda)$
- l'assemblage *Pseudonitzschia seriata* a été associé à deux classes de  $Ra(\lambda)$

- l'assemblage *Thalassiosira sp.* a été associé à quatre classes de  $Ra(\lambda)$
- l'assemblage *Rhizosolenia imbricata* a été associé à trois classes de  $Ra(\lambda)$

Pour le set de validation correspondant, les taux de validation par classes de  $Ra(\lambda)$  sont présentés en Table 5.6.

Certaines classes de  $Ra(\lambda)$  peuvent être associées en phase de calibration à un assemblage mais ne peuvent pas être validées puisque la classe de  $Ra(\lambda)$  n'est pas présente dans les données de validation ("NA" dans la Table 5.6). Dans une procédure de calibration spécifique, des scores de validation moyens par assemblages peuvent être calculés en fonction du nombre de classes de  $Ra(\lambda)$  auxquelles il a été associé (dernière ligne de la Table 5.6). L'assemblage (a) *Chaetoceros sp.2* n'a pas de score moyen puisqu'il n'a pas été associé à une classe de  $Ra(\lambda)$  au cours de cette procédure spécifique. Pour les assemblages (c) *Thalassionema nitzschia* et (f) *Rhizosolenia imbricata*, les scores de validation sont assez bon avec respectivement 67,7 et 83,3% de bonnes identifications de la classe de  $Ra(\lambda)$  avec l'assemblage *in-situ* spécifique. En revanche, pour les trois autres assemblages, les scores de validation sont nuls (0% pour (b) *Chaetoceros sp1.* et (d) *Pseudonitzschia seriata*) ou faibles (25% pour (e) *Thalassiosira spp.*).

Cet exemple souligne l'importance de la répétition de la calibration pour ces données particulières. Par exemple, sans répétitions, le  $Ra(\lambda)$  n°1 aurait pu être labellisé assemblage *Thalassiosira spp.* alors que les scores de validation sont à 0%. En effet, pour ce  $Ra(\lambda)$ , qui a été associé dans 63% des 1200 calibrations distinctes à cet assemblage, la validation est toujours à 0%. Ceci s'explique par le fait que 50% des mesures coïncidentes sont associées à cet assemblage *Thalassiosira spp.*. Ainsi, selon la configuration de la séparation entre les données de calibration et de validation, le  $Ra(\lambda)$  n°1 sera soit labellisé mais non validé (0%), ou non labellisé.

Le seuil de 70% sur le "taux de calibration" a été choisi car il empêche ce genre de situation d'être prises en compte dans le choix final des classes de  $Ra(\lambda)$  ayant une association robuste avec un assemblage CPR. Le seuil permet de faire une validation *a posteriori* en testant la validation des associations sélectionnées. En effet, après avoir choisi les associations  $Ra(\lambda)$ -assemblages robustes, les scores de validation peuvent être recalculés uniquement pour ces  $Ra(\lambda)$ . Les résultats sont présentés en Table 5.7. En supprimant les labellisations qui existent pour un petit nombre de procédures de calibration, cette procédure spécifique permet de labelliser 9 classes  $Ra(\lambda)$ , au lieu de précédemment 19 classes. Les scores de validation sont alors nettement améliorés.



TABLE 5.6 – Scores de validation intermédiaires issus d’une calibration aléatoire parmi les 1200 calibrations réalisées. Pour chaque classe de  $Ra(\lambda)$  (ligne), le pourcentage d’occurrence des six assemblages CPR est donné pour les données de validation avec les assemblages caractérisés par a) *Chaetoceros sp2.*, b) *Chaetoceros sp1.*, c) *Thalassionema nitzschoides*, d) *Pseudonitzschia seriata*, e) *Thalassiosira sp.* et f) *Rhizosolenia imbricata*. Les cellules grisées correspondent à l’assemblage auquel la classe de  $Ra(\lambda)$  a été associée. "NA" est inscrit lorsque la validation n’a pas pu être effectuée, car la classe de  $Ra(\lambda)$  était absente des données de validation. Les classes de  $Ra(\lambda)$  suivies d’une astérisque ont été considérées comme robustement labellisées selon le seuil de 70% de taux de calibration (voir Section précédente).

Assemblage → ↓ Classe de $Ra(\lambda)$	a	b	c	d	e	f
Scores de validation ↘						
1	0	50	0	50	0	0
14	16,67	0	33,34	16,67	33,34	0
15*	0	0	75	0	25	0
18*	0	50	0	0	0	50
19	0	0	0	50	0	50
24*	0	0	100	0	0	0
25	100	0	0	0	0	0
29*	0	0	0	0	0	100
34	NA	NA	NA	NA	NA	NA
36*	0	0	0	0	0	100
41	0	0	100	0	0	0
50*	0	0	100	0	0	0
51*	0	0	100	0	0	0
58*	0	0	0	0	100	0
65	16,67	16,67	33,34	16,67	16,67	0
66*	25	0	75	0	0	0
71	25	0	25	25	25	0
75	NA	NA	NA	NA	NA	NA
84	0	0	0	0	0	100
<b>Score moyen →</b>	NA	0	67,7	0	25	83,3

TABLE 5.7 – Scores de validation *a posteriori* issus de la sélection des associations robustes (70% de taux de calibration minimum, soit au moins 840 fois sur 1200 procédures) pour une calibration aléatoire parmi les 1200 calibrations réalisées. Pour chaque classe de  $Ra(\lambda)$  (en ligne), le pourcentage d'occurrence des six assemblages CPR est donné pour les données de validation avec les assemblages caractérisés par a) *Chaetoceros sp2.*, b) *Chaetoceros sp1.*, c) *Thalassionema nitzschoides*, d) *Pseudonitzschia seriata*, e) *Thalassiosira spp.* et f) *Rhizosolenia imbricata* et uniquement pour les associations considérées comme robustes (\*) selon le seuil sur le taux de calibration. Les cellules grisées correspondent à l'assemblage auquel la classe de  $Ra(\lambda)$  a été associée.

Assemblage → ↓ Classe de $Ra(\lambda)$	a	b	c	d	e	f
Scores de validation ↘						
15*	0	0	75	0	25	0
18*	0	50	0	0	0	50
24*	0	0	100	0	0	0
29*	0	0	0	0	0	100
36*	0	0	0	0	0	100
50*	0	0	100	0	0	0
51*	0	0	100	0	0	0
58*	0	0	0	0	100	0
66*	25	0	75	0	0	0
<b>Score moyen <i>a posteriori</i> →</b>	-	-	90	-	100	83,3

### Validation totale moyenne et *a posteriori*

Les scores de validation totaux sont présentés en Figure 5.6 et un résumé est présenté en Table 5.8. La Figure 5.6 montre les moyennes des scores de validation (et l'écart type) pour toutes les classes de  $Ra(\lambda)$  qui ont été labellisées au moins une fois au cours des procédures de calibration. Ainsi, les scores de validation ont une variabilité assez élevée en fonction des différents jeu de données de calibration et de validation.

Lorsque le seuil sur le taux de calibration n'est pas atteint (70% de taux de calibration), i.e., les trois assemblages pour lesquels une association robuste avec les  $Ra(\lambda)$  n'a pas été trouvée, les scores de validation sont aux environs de 20% (voir la colonne "Validation totale" de la Table 5.8).

Pour les trois autres assemblages pour lesquels une (ou plusieurs) association robuste a pu être trouvée, les scores de validation sont présentés de deux manières :

- Premièrement, les scores de validation totaux, qui prennent en compte tous les  $Ra(\lambda)$  labellisés, qu'ils soient robustement labellisés ou non, atteignent 61,8%, 60,6%, and 60,5%, respectivement pour les assemblages caractérisés par *Thalassionema nitzschoides*, *Thalassiosira sp.* et *Rhizosolenia imbricata* (Table 5.8, colonne "Validation totale")
- Deuxièmement, en sélectionnant les  $Ra(\lambda)$  labellisés grâce aux taux de calibration (qui doit être supérieur à 70%), les scores de validation sont augmentés, atteignant 74,7%, 95,2%, and 63,8%, respectivement pour les assemblages caractérisés par *Thalassionema nitzschoides*, *Thalassiosira spp.* et *Rhizosolenia imbricata* (Table 5.8, colonne "Validation sélectionnée")

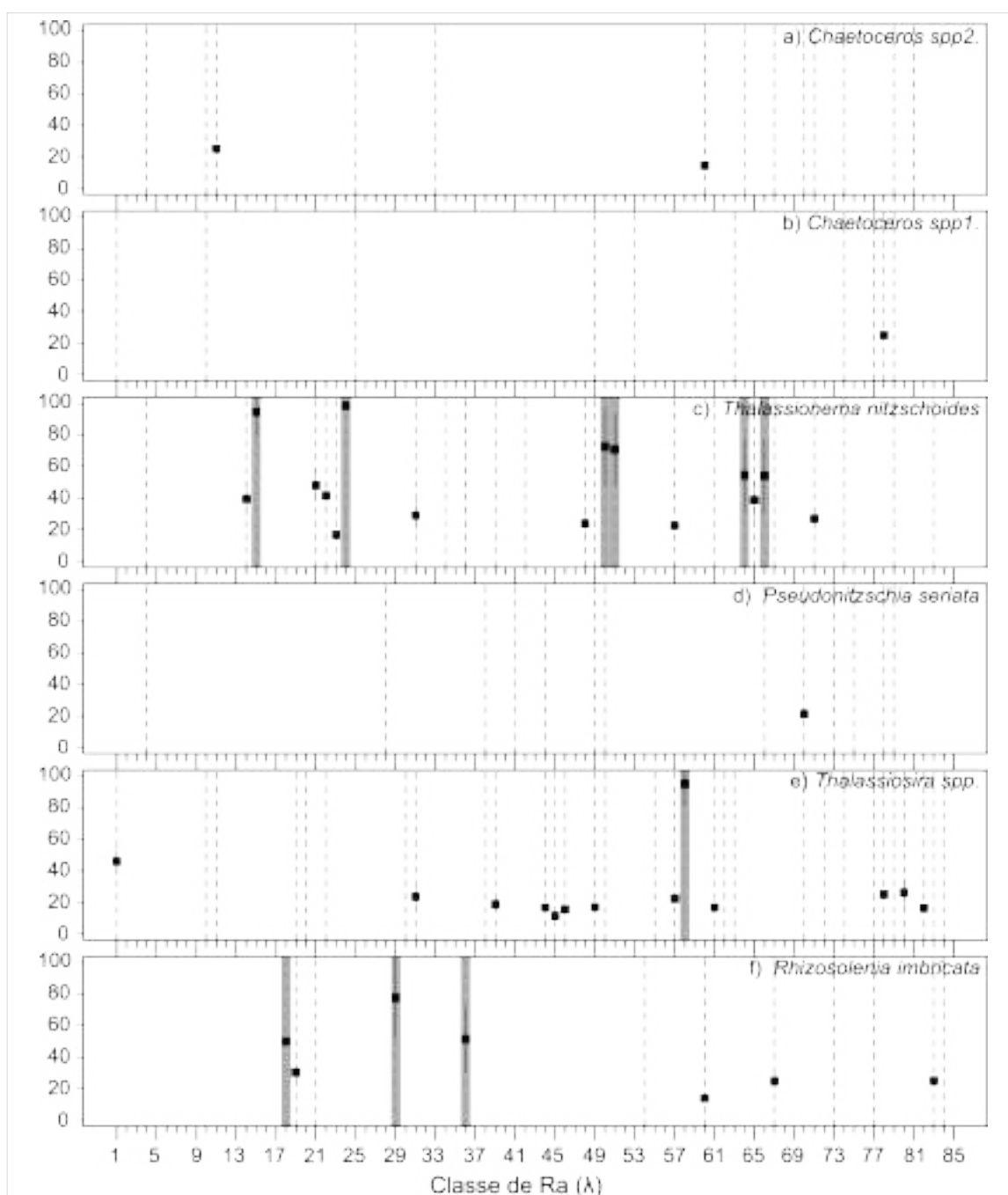


FIGURE 5.6 – Moyenne des scores de validation (axe des Y, en pourcentage, symboles carrés noirs) pour les 1200 procédures de labellisation des classes de  $Ra(\lambda)$  (axe des X), avec l'écart type (lignes noires) pour les six assemblages de diatomées caractérisés par : a) *Chaetoceros* sp2., b) *Chaetoceros* sp1., c) *Thalassionema nitzschoides*, d) *Pseudonitzschia seriata*, e) *Thalassiosira* sp. et f) *Rhizosolenia imbricata*. Une ligne pointillée est dessinée lorsque la classe de  $Ra(\lambda)$  a été labellisée au moins une fois durant les 1200 procédures successives. **Un arrière-plan gris est dessiné lorsque la classe de  $Ra(\lambda)$  a été considérée comme robustement labellisée**, en accord avec le seuil de 70% sur le taux de labellisation (Figure 5.3). Quand l'information *in-situ* n'a pas été pas suffisante pour calibrer et valider l'association de façon robuste, il n'y a que une ligne en pointillés (par exemple,  $Ra(\lambda)$  °4).

TABLE 5.8 – Ce tableau représente les scores de validation de deux manières : (i) "Validation totale" : les scores de validation totaux, qui tiennent compte de tous les  $Ra(\lambda)$  labellisés dans les 1200 procédures successives, i.e., la moyenne de tous les scores de validation présentés dans la Figure 5.6 ; et (ii) "Validation sélectionnée" : les scores de validation des  $Ra(\lambda)$  sélectionnés sur la base du taux de calibration (seuil de 70% sur les taux de calibration = présence de cette calibration dans au moins 840 fois sur les 1200 fois), i.e., la moyenne des scores de validation ayant un arrière-plan gris dans la Figure 5.6. Le symbole  $\emptyset$  est inscrit dans les cas où aucune association n'est trouvée. \* Ne sont inscrits que les  $Ra(\lambda)$  labellisés de manière robuste selon le taux de calibration. Une colonne est mise en évidence en bleu : elle correspond aux scores de validation présentés dans la suite de l'étude.

Nom de l'assemblage	Validation totale (%)	Validation sélectionnée (%)	$Ra(\lambda)$ labellisés*
<i>Chaetoceros sp.2</i>	17,06	$\emptyset$	$\emptyset$
<i>Chaetoceros sp.1</i>	25,0	$\emptyset$	$\emptyset$
<i>Thalassionema nitzschoïdes</i>	61,82	74,73	n°15, 24, 50, 51, 64, 66
<i>Pseudonitzschia seriata</i>	21,25	$\emptyset$	$\emptyset$
<i>Thalassiosira sp.</i>	60,58	95,24	n°58
<i>Rhizosolenia imbricata</i>	60,47	63,82	n°18, 29, 36

### 5.2.5 Distribution spatio-temporelle des anomalies de luminance associées aux assemblages de diatomées

Dans le but de produire des cartes de fréquence de détection de ces classes de  $Ra(\lambda)$  labellisées, les classes de  $Ra(\lambda)$  qui ont été empiriquement associées à des assemblages ont pu être recherchées dans les cartes journalières de notre base de données de  $Ra(\lambda)$ . Des cartes de fréquences mensuelles moyennes des classes de  $Ra(\lambda)$  labellisées sont présentées en Figure 5.7. Pour produire ces cartes, les fréquences d'occurrence des  $Ra(\lambda)$  labellisés ont été moyennées sur chaque pixel et pour chaque mois entre 1998 et 2004. La fréquence des  $Ra(\lambda)$  labellisés, peu importe l'assemblage auquel ils ont été associés, dépassent rarement 0,5. Cela signifie qu'un pixel est régulièrement détecté comme étant associé à un assemblage moins de la moitié d'un mois, ce qui suggère qu'il reste un fort potentiel d'extension de la caractérisation des  $Ra(\lambda)$  en Mer du Nord, autant pour les assemblages décrits ici que pour d'autres types de phytoplancton ; et qu'une grande partie de la base de données reste disponible pour des études futures (voir Discussion).

Cependant, les cartes présentent des patterns spatiaux et temporels intéressants au regard des connaissances passées, notamment pour les six classes de  $Ra(\lambda)$  qui ont été identifiées comme correspondant à l'assemblage *Thalassionema nitzschoides* (Figure 5.7.a). Leur fréquence maximale est observée au printemps (Mars et Avril) et en automne (Septembre et Octobre). Ces classes sont plus fréquemment trouvées dans la partie Sud de la Mer du Nord, dans la Manche et le long de la côte Ouest du Royaume-Uni. En été, ces six classes de  $Ra(\lambda)$  ont une fréquence faible, et rencontrées dans la partie Nord de la Mer du Nord et le long des côtes Sud du Royaume-Uni.

L'unique classe de  $Ra(\lambda)$ , qui a pu être associée avec l'assemblage caractérisé par *Thalassiosira spp.* a une fréquence faible à ce stade, relativement au nombre total de classes de  $Ra(\lambda)$  possibles. Cependant, cette classe présente une fréquence maximale en milieu d'année (Juin et Juillet) et est principalement localisée au centre de la Mer du Nord (Figure 5.7.b).

Les trois classes de  $Ra(\lambda)$  associées à l'assemblage *Rhizosolenia imbricata* présentent une fréquence maximale au printemps (Mars et Avril) qui diminue progressivement les mois suivants pour rester faible le reste de l'année. Ils sont localisés dans la Manche et le long des côtes Sud du Royaume-Uni. En été (Juin et Juillet), un patch de ces  $Ra(\lambda)$ , avec une fréquence assez élevée, apparaît le long des côtes norvégiennes et disparaît les mois suivants (Figure 5.7.c). Ce patch apparaît pour deux des trois classes de  $Ra(\lambda)$  associées à l'assemblage de *Rhizosolenia imbricata*.

FIGURE 5.7 – *pages 200 à 209* - Cartes de climatologie de fréquence de détection des  $Ra(\lambda)$  (données SeaWiFS niveau 2 de 1998 à 2004) qui ont été associés à l’assemblage caractérisé par : (a) *Thalassionema nitzschoides* (pages 200 à 203), (b) *Thalassiosira spp* (pages 204 à 206), (c) *Rhizosolenia imbricata* (pages 207 à 209). Les fréquences sont calculées en divisant le nombre de fois où des  $Ra(\lambda)$  labellisés spécifiquement pour un assemblage sont détectés, sur le nombre de fois où une autre classe de  $Ra(\lambda)$  (labellisés ou non à ce stade) est détectée pour un pixel donné. Pour un assemblage donné, si plus d’un  $Ra(\lambda)$  a été labellisé, leurs fréquences sont ajoutées. Notez que les barres de couleur des fréquences sont différentes pour les trois assemblages. Les pixels sont blancs lorsque moins de trois occurrences sont détectées dans les mois considérés. De Novembre à Février (hiver boréal), l’angle solaire est important et le signal satellite ne peut pas être correctement estimé au dessus de  $52^{\circ}\text{N}$ . Afin de ne pas biaiser le calcul de la fréquence, les résultats ne sont présentés que pour la période allant de Mars à Octobre.

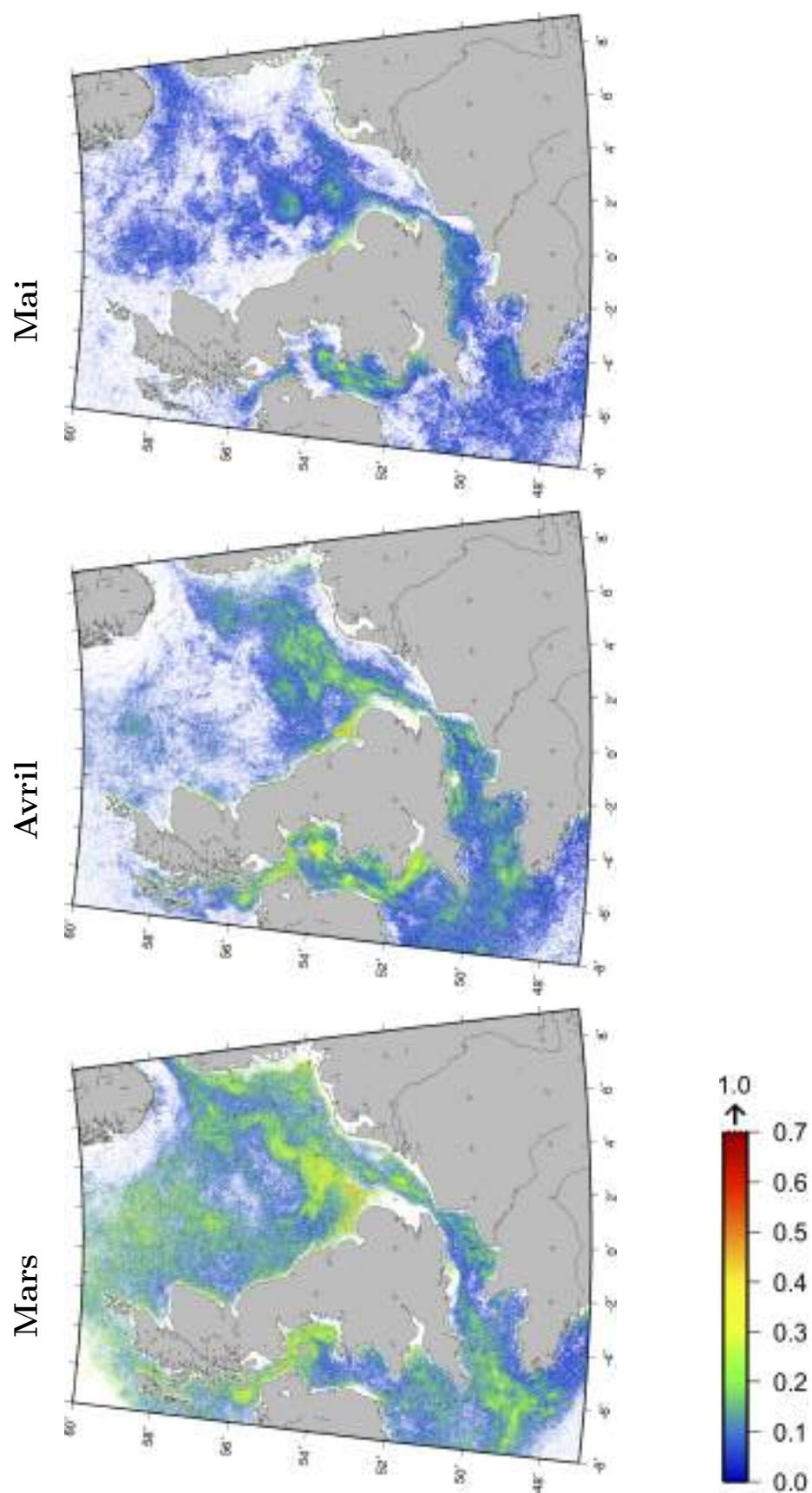


Figure 5.7 - (a) Fréquence de détection des classes de  $Ra(\lambda)$  associées à l'assemblage *Thalassionema nitzschoides* (légende page 200).



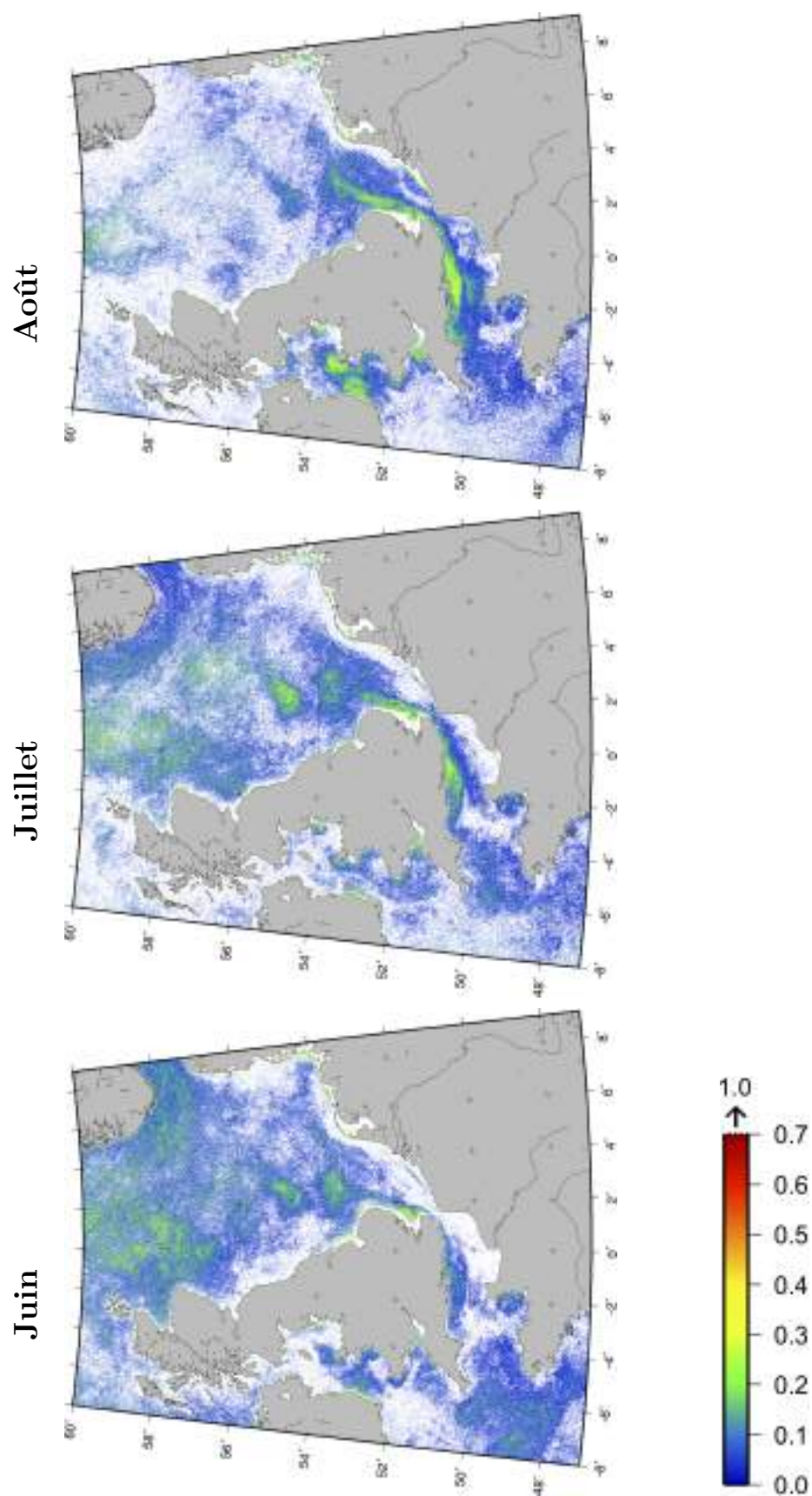


Figure 5.7 - (a) Fréquence de détection des classes de  $Ra(\lambda)$  associées à l'assemblage *Thalassionema nitzschoides* (légende page 200).

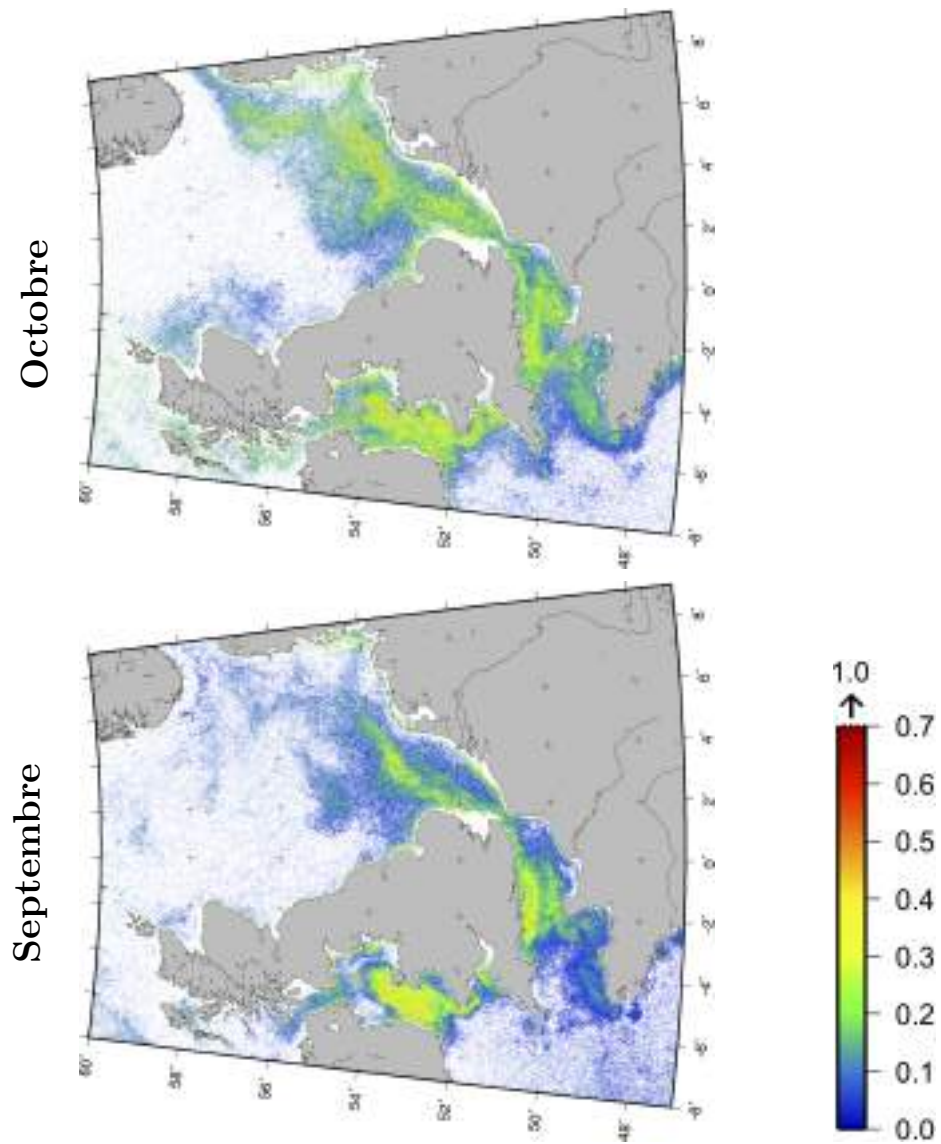


Figure 5.7 - (a) Fréquence de détection des classes de  $Ra(\lambda)$  associées à l'assemblage *Thalassionema nitzschooides* (légende page 200).

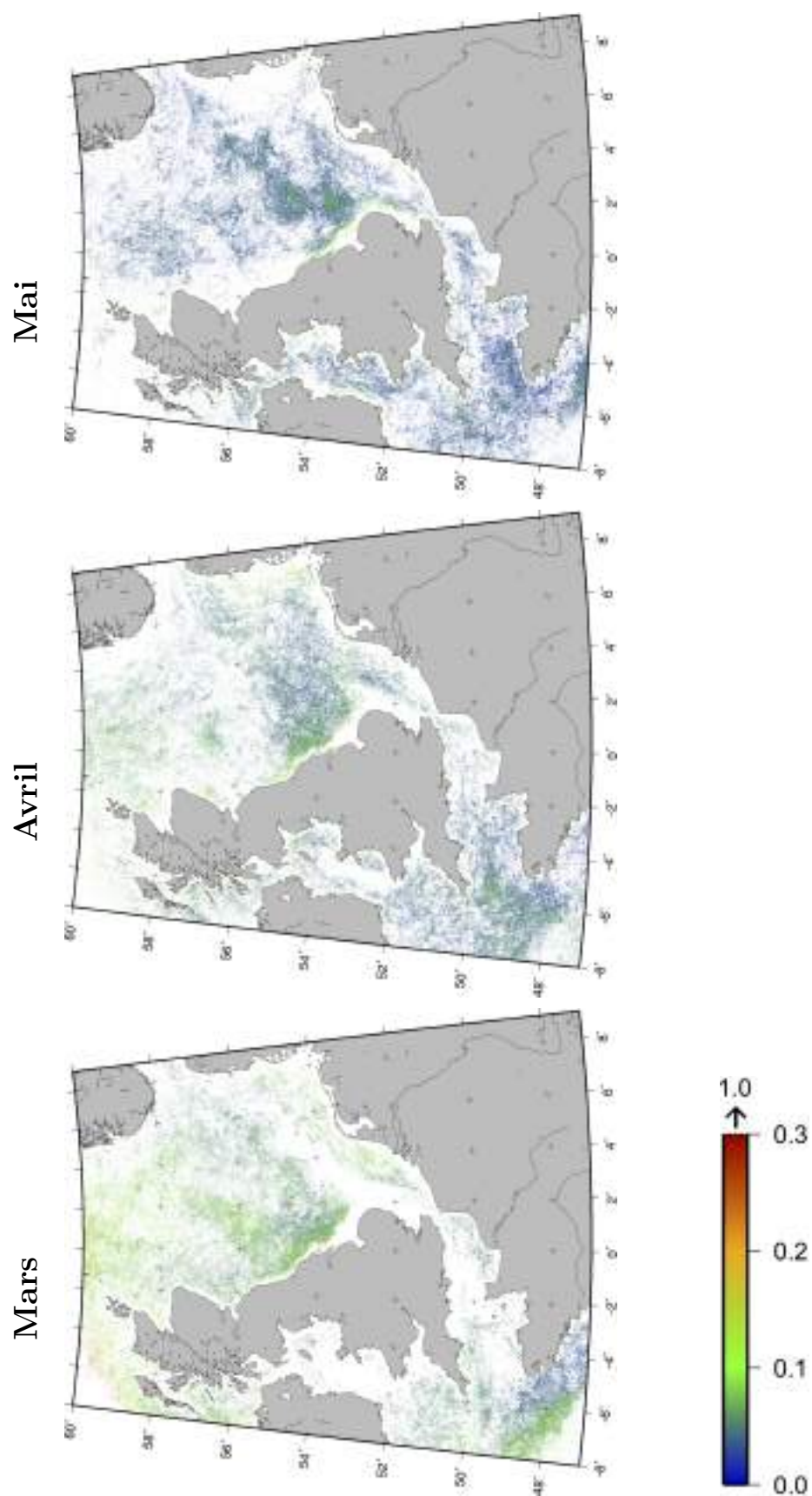


Figure 5.7 - (b) Fréquence de détection des classes de  $Ra(\lambda)$  associées à l'assemblage *Thalassiosira* spp. (légende page 200).

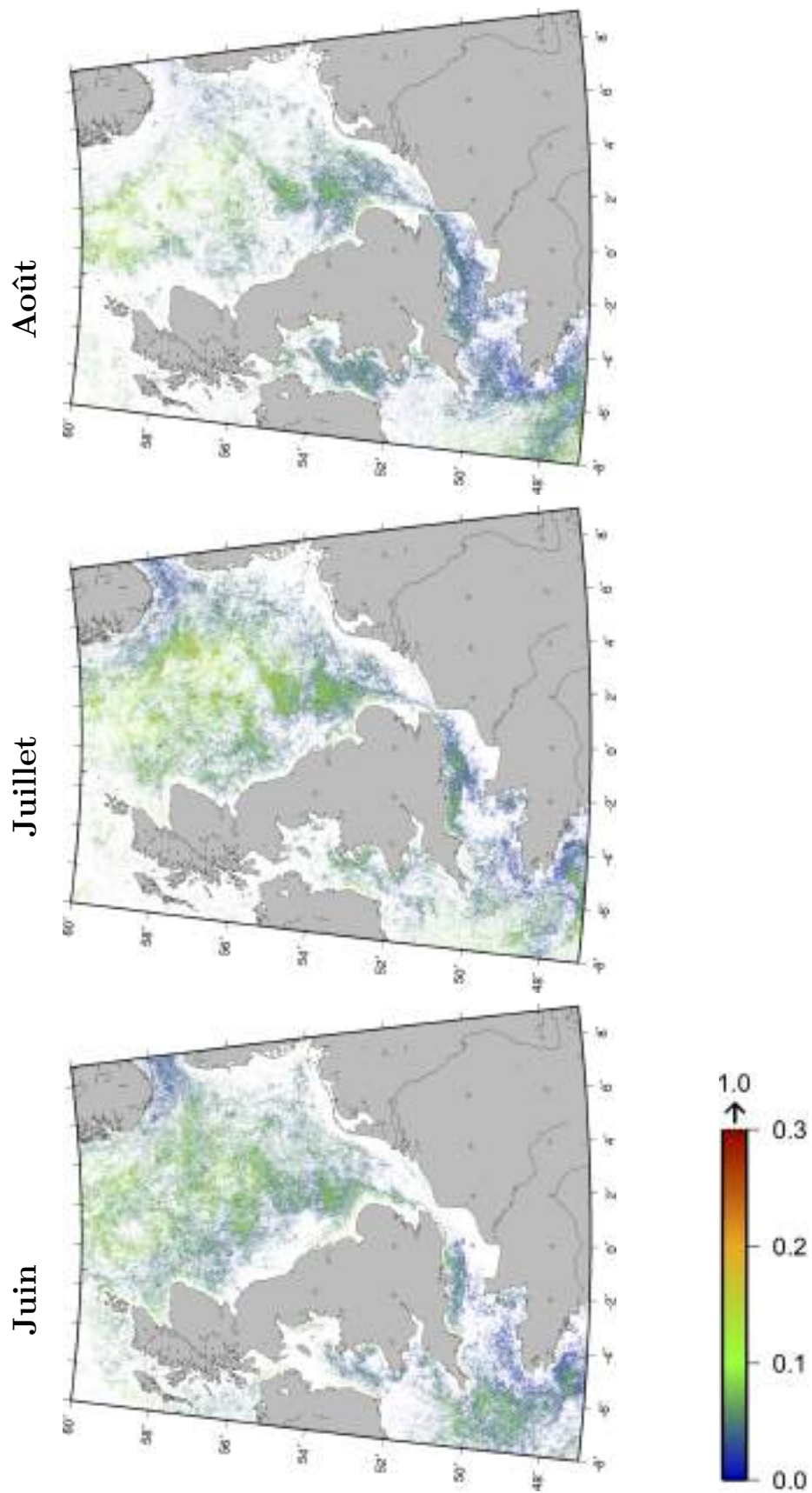


Figure 5.7 - (b) Fréquence de détection des classes de  $Ra(\lambda)$  associées à l'assemblage *Thalassiosira* spp. (légende page 200).

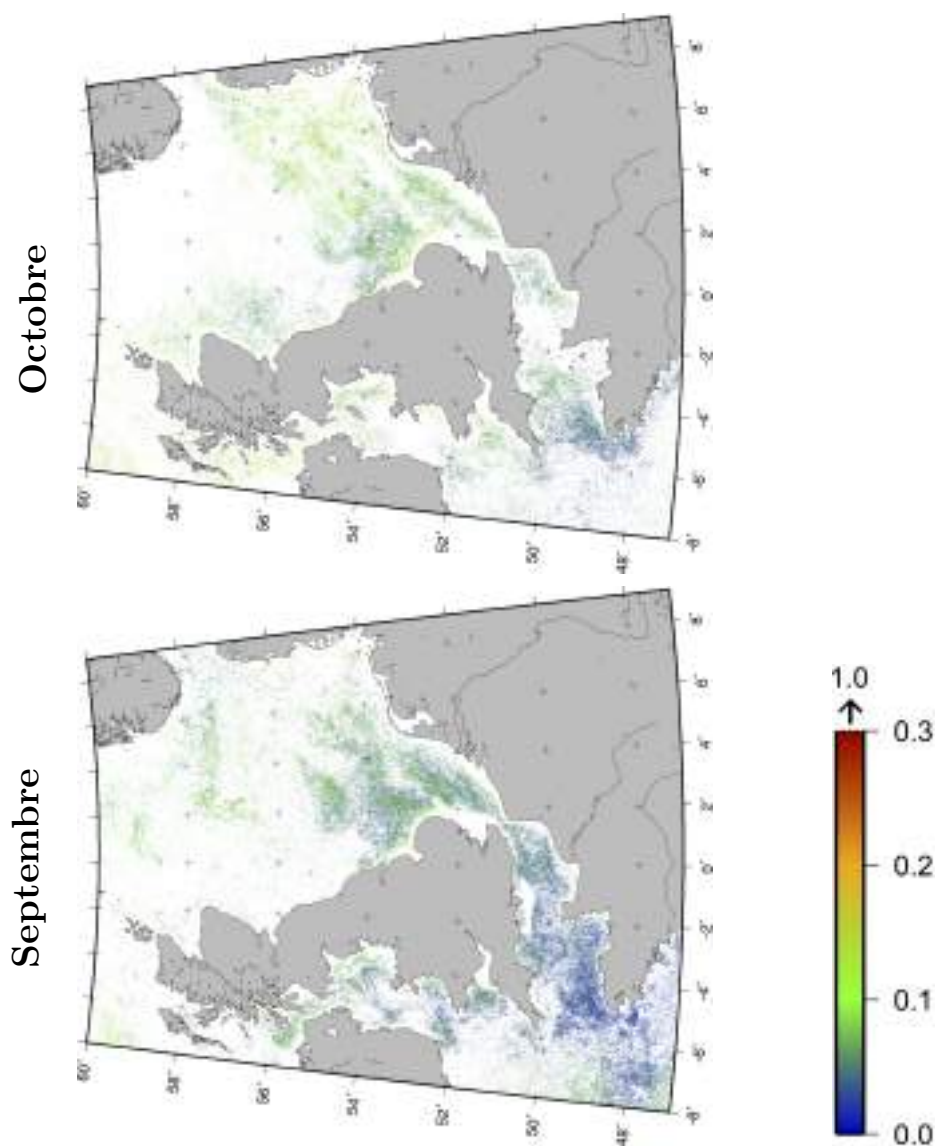


Figure 5.7 - (b) Fréquence de détection des classes de  $Ra(\lambda)$  associées à l'assemblage *Thalassiosira* spp. (légende page 200).

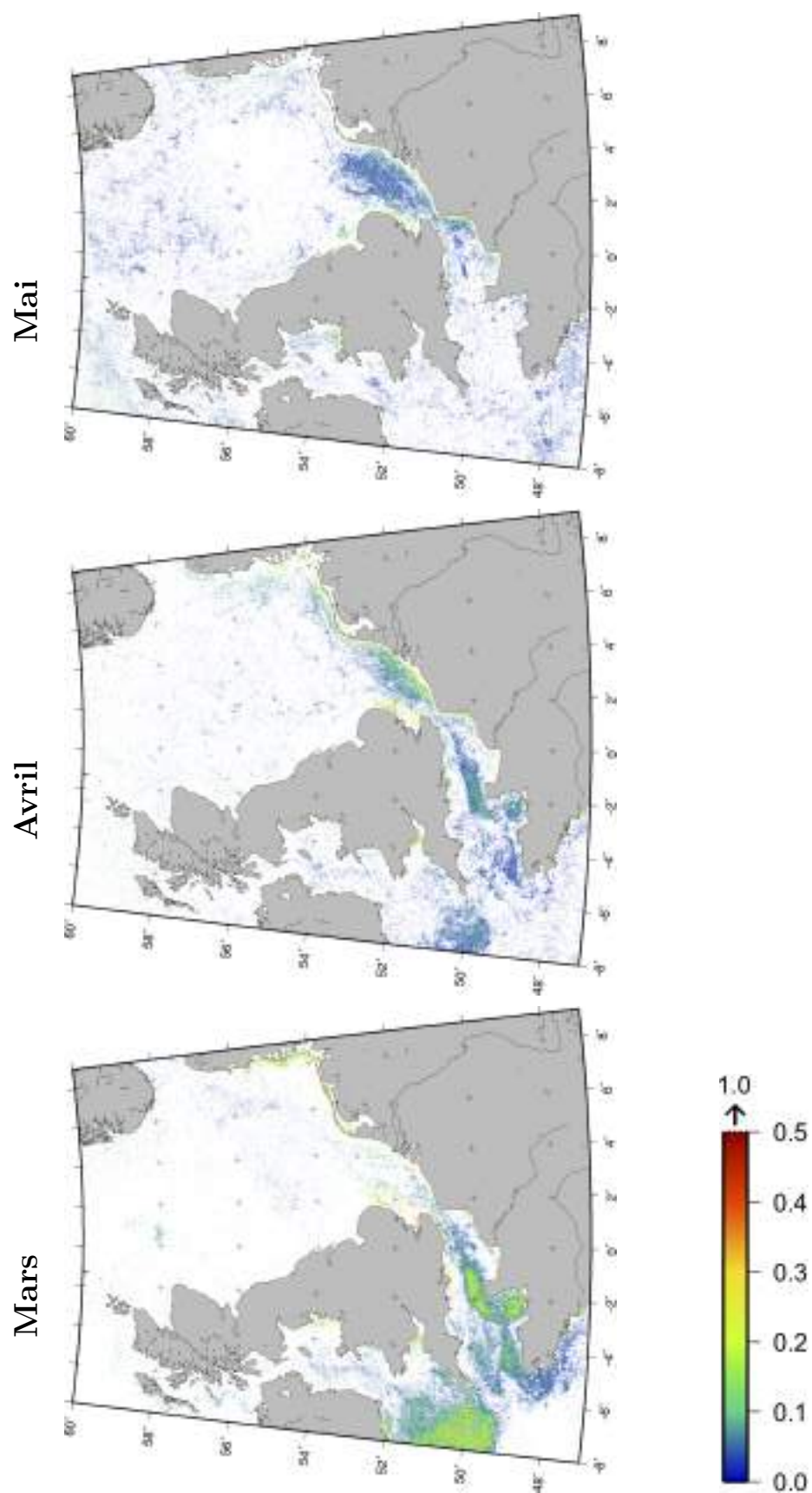


Figure 5.7 - (c) Fréquence de détection des classes de  $Ra(\lambda)$  associées à l'assemblage *Rhizosolenia imbricata* (légende page 200).

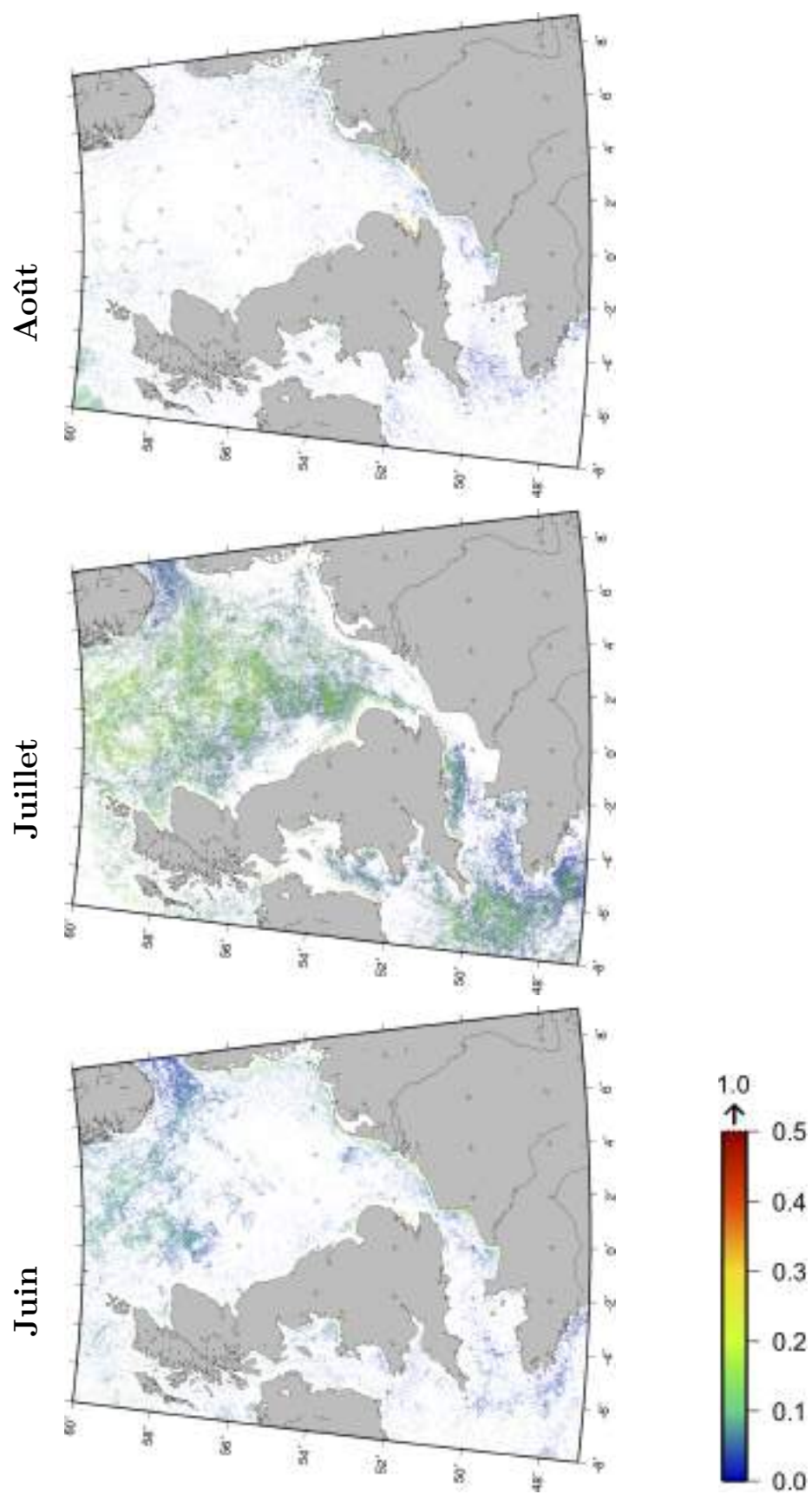


Figure 5.7 - (c) Fréquence de détection des classes de  $Ra(\lambda)$  associées à l'assemblage *Rhizosolenia imbricata* (légende page 200).

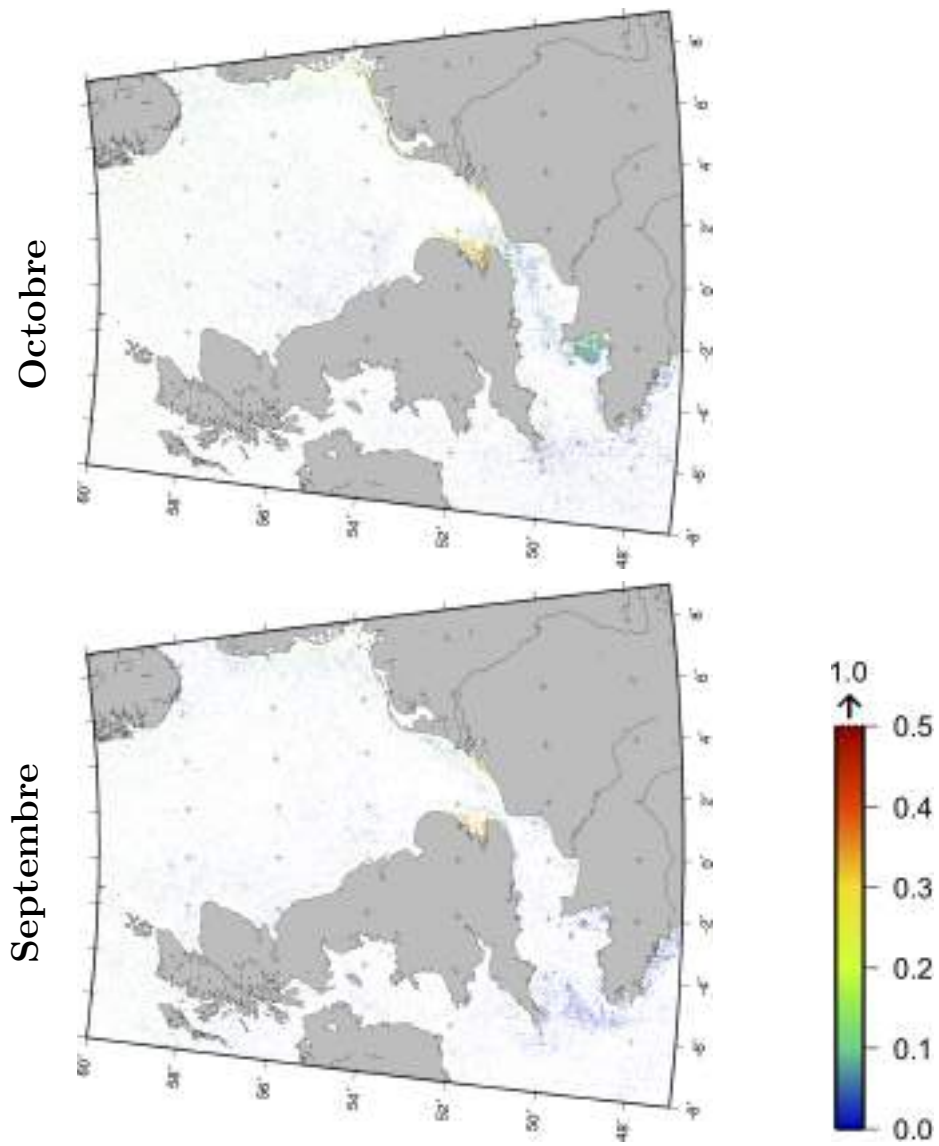


Figure 5.7 - (c) Fréquence de détection des classes de  $Ra(\lambda)$  associées à l'assemblage *Rhizosolenia imbricata* (légende page 200).



## 5.3 Discussion

### 5.3.1 Choix de la méthode *Indicator Value* pour la caractérisation des assemblages de diatomées *in-situ*

La méthode IndVal, combinée avec la classification des échantillons, a permis d'identifier six assemblages de diatomées, dont chacun a été caractérisé par une espèce de diatomée indicatrice particulière (qui sert de nom à l'assemblage). *In-situ*, ces assemblages ont des caractéristiques saisonnières et spatiales distinctes (Figure 5.8). Ils sont stables (qualitativement) sur la période de l'étude de 1998 à 2010.

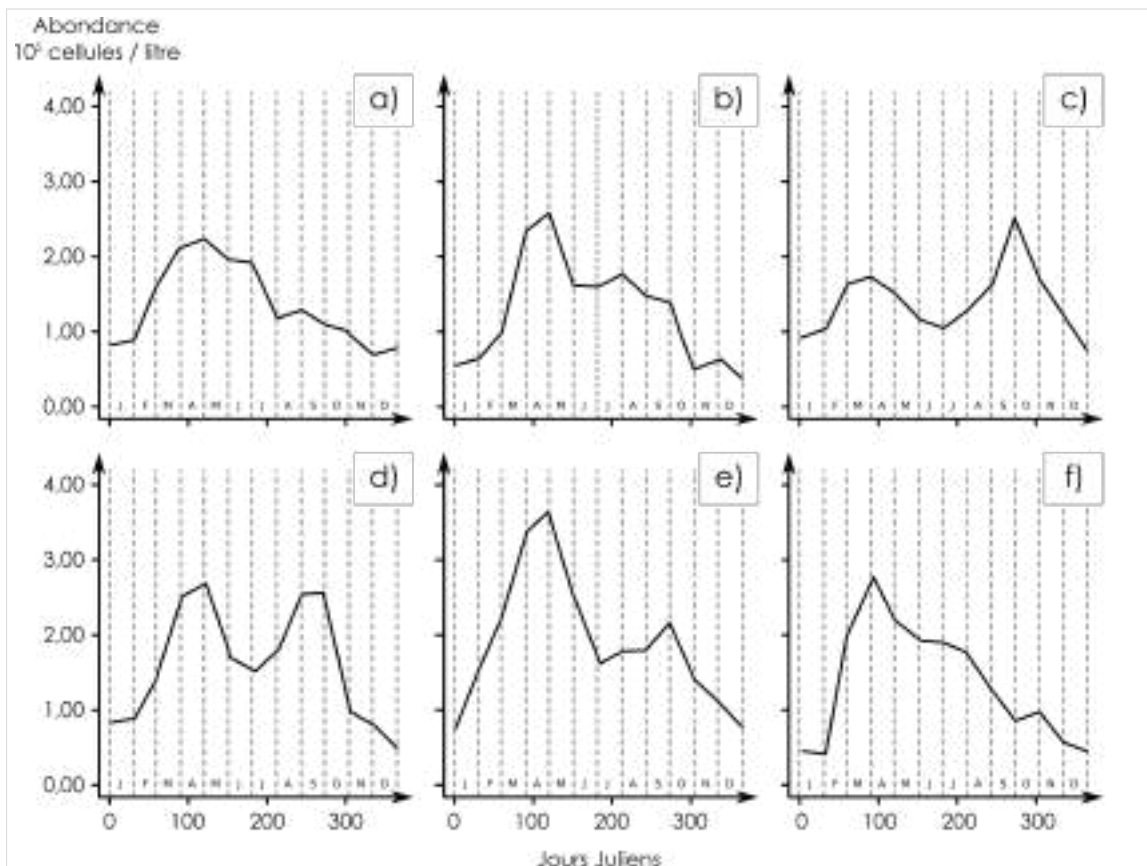


FIGURE 5.8 – Cycle saisonnier moyen des six assemblages de diatomées caractérisés par la méthode IndVal. L'axe des X représente les jours Julien et l'axe des Y représente l'abondance cumulée des espèces représentant les assemblages caractérisés par : a) *Chaetoceros sp2*. (n=855), b) *Chaetoceros sp1*. (n=961), c) *Thalassionema nitzschoides* (n=901), d) *Pseudonitzschia seriata* (n=1195), e) *Thalassiosira sp.* (n=1415) et f) *Rhizosolenia imbricata* (n=650). Pour un assemblage, les abondances des diatomées qui caractérisent l'assemblage sont additionnées pour un mois et moyennées mensuellement sur les années (1998-2010).

Pour cinq des six espèces ou taxa indicateurs identifiés *in-situ*, i.e., *Chaetoceros spp1*, *Chaetoceros spp2*, *T. nitzschoides*, *P. seriata*, and *R. imbricata*, la valeur de l'IndVal a

augmenté pour les cinq premiers niveaux de regroupement de la classification des échantillons CPR (ce qui correspond à six assemblages), puis l'IndVal diminue pour les niveaux de regroupement suivants. Cela indique que ces espèces sont dites *sténotopiques*, c'est-à-dire qu'elles ont une petite répartition de *niche écologique*. Ces espèces ont également une saisonnalité forte et une étendue spatiale relativement limitée, ce qui a été couramment rapporté en Mer du Nord et dans la Manche (e.g., Reid *et al.* (1990), Rousseau *et al.* (2002), Bresnan *et al.* (2009), Hinder *et al.* (2012), Hoppenrath (2004)).

Au contraire, pour le sixième taxon indicateur, *Thalassiosira spp.*, la valeur de l'IndVal décroît de manière significative au fur et à mesure des niveaux de regroupement. En effet, la mesure de fidélité de *Thalassiosira spp.* est assez haute à travers tous les assemblages, cependant, son abondance est plus importante dans l'assemblage qu'elle caractérise (mesure de spécificité plus élevée dans son assemblage que dans les autres). Ceci indique que ce taxon est eurytopique, c'est-à-dire qu'il a une étendue de niche relativement importante, même s'il a tendance à se développer de manière plus importante pour certains sites. Des résultats similaires ont déjà été rapportés, montrant notamment une augmentation de l'abondance et de la couverture spatiale de ce taxon en Mer-du-Nord (e.g., Bresnan *et al.* (2009), Hinder *et al.* (2012)). D'ailleurs, l'étude de (Trigueros et Orive, 2001) dans les eaux côtières du Nord de l'Espagne a montré que cette espèce était présente à toutes les saisons (avec un maximum au printemps), pour des salinités de 1,3 à 35,4 et des températures de 7,3 à 24,7°C. Dans notre étude, il ne nous a pas été possible de caractériser l'environnement lié à la présence de ce taxon. En revanche, son aspect ubiquiste peut être dû au fait que plusieurs espèces de *Thalassiosira spp.* composent probablement ce taxon.

Selon Dufrêne et Legendre (1997), les espèces eurytopiques pourraient être mieux définies qu'elles ne le sont actuellement dans la méthode IndVal. En effet, la plupart des espèces eurytopiques se montrent caractéristiques à des niveaux de groupement plus grand, comme montré pour *Thalassiosira spp.*, qui a été indicatrice dès le premier niveau de regroupement (L1), avec une forte valeur indicatrice (IndVal = 92,4). Cette espèce a donc une influence sur la classification des échantillons CPR effectuée au préalable du calcul de l'IndVal. Cependant, la classification des diatomées du CPR est une approche cohérente puisque les espèces sont subdivisées en assemblages précis, typiques de la combinaison de facteurs écologiques, mais aussi des échelles temporelles et spatiales. Par exemple, il est noté dans la littérature que les espèces *Chaetoceros spp.* et *Bacteriastrum spp.* forment souvent un assemblage printanier (Trigueros et Orive, 2001).

Dans cette étude, les assemblages de diatomées ont été utilisés comme un indicateur de la composition de la communauté phytoplanctonique totale. Cependant, les assemblages de diatomées ont des communautés associées qui varient, ce qui n'a pas pu être totalement évalué ici. Les données du CPR comportent également les abondances de trois autres groupes phytoplanctoniques, dont les dinoflagellés. Bien que les dinoflagellés n'aient pas

montré une spécificité particulière pour les assemblages définis selon la méthode IndVal, nous pouvons détailler ce que nous savons d’eux au sein des assemblages à partir des données CPR.

Les échantillons faisant partie de l’assemblage caractérisé par *Thalassionema nitzschoides* ont une abondance totale qui est due à 93%<sup>7</sup> aux diatomées, et à 7% aux dinoflagellés. Ce nombre descend à 83,2 % de diatomées pour l’assemblage caractérisé par *Thalassiosira sp.* et à 65,1% de diatomées pour l’assemblage caractérisé par *Rhizosolenia imbricata*. La communauté de dinoflagellés peut ainsi expliquer en partie la variabilité des caractéristiques des  $Ra(\lambda)$  labellisés (Figure 5.4), même si d’autres facteurs entrent sûrement en compte (voir la Section 5.3.3). Bien que les dinoflagellés soient présents aux côtés des diatomées dans tous les assemblages caractérisés ici, ils n’ont pas pu être évalués comme des indicateurs de ces assemblages au sens de Dufrene et Legendre (1997). Ceci est expliqué par le fait que les données CPR ne décrivent pas l’ensemble de la communauté de dinoflagellés au niveau spécifique<sup>8</sup>. Ainsi, il est possible qu’une espèce de dinoflagellés particulière puisse être indicatrice d’un assemblage spécifique. La méthode IndVal, appliquée aux groupes phytoplanctoniques du CPR<sup>9</sup>, a montré que le groupe des dinoflagellés est plutôt indicateur de l’été (mois de Juin à Septembre), dues des conditions environnementales propices à leur développement (notamment la stratification), alors que les diatomées sont plutôt indicatrices des périodes hivernales et printanières. Ainsi, la manière de considérer les espèces et les groupes phytoplanctoniques présents, i.e., le niveau taxonomique considéré, a un impact sur la caractérisation des espèces ou groupes indicateurs et permet d’établir des conclusions différentes et complémentaires.

Bien que la caractérisation des assemblages de diatomées ait été appropriée pour cette première évaluation du potentiel de PHYSAT dans la détection d’assemblage, une base de données *in-situ* plus détaillée serait absolument nécessaire dans le futur pour prendre en compte d’autres situations, compléter la labellisation et comprendre les observations mises en évidence dans cette étude et en définir les limites.

### 5.3.2 Distribution des classes de $Ra(\lambda)$ labellisées

À ce jour, dix classes de  $Ra(\lambda)$  sur les 85 classes potentielles ont pu être labellisées avec trois assemblages spécifiques de diatomées issus des données CPR disponibles dans les eaux de la Mer du Nord avec des scores de validation assez élevés. Les trois autres assemblages de diatomées n’ont pas pu être associés robustement à des classes de  $Ra(\lambda)$  à ce jour, selon nos critères (Sections 5.1.3.1 et 5.3).

---

7. Ne sont pris en compte dans ce pourcentage que les cellules microphytoplanctoniques disponibles dans les données CPR.

8. En effet, certains dinoflagellés, à coquilles dures, peuvent être compté au niveau spécifique. Pour les autres flagellés, à membranes molles, leur abondance est comptée en tant que "dinoflagellés" sans plus de détails.

9. Le principe est le même que celui décrit en matériel et méthode, sauf que la matrice "d’espèces" contient ici l’abondance des groupes CPR.

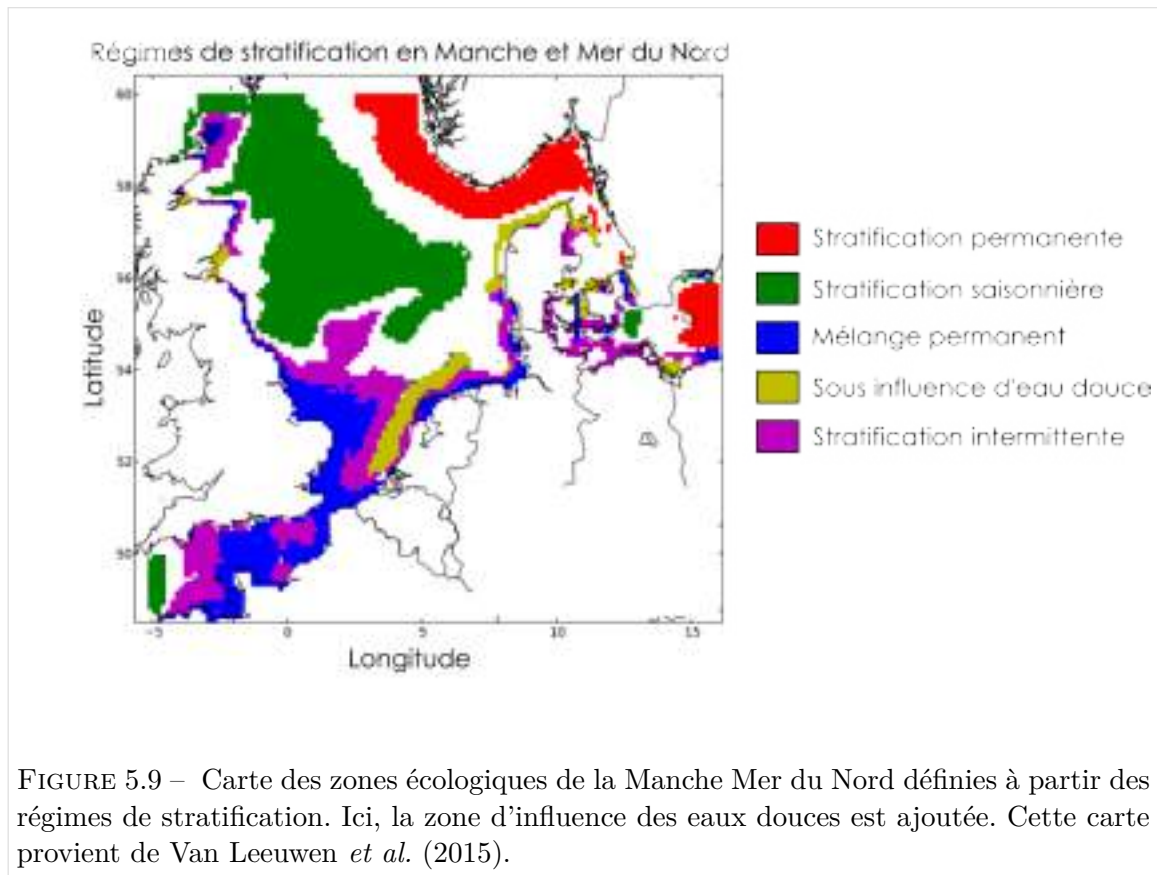


FIGURE 5.9 – Carte des zones écologiques de la Manche Mer du Nord définies à partir des régimes de stratification. Ici, la zone d’influence des eaux douces est ajoutée. Cette carte provient de Van Leeuwen *et al.* (2015).

Pour les dix classes de  $Ra(\lambda)$  qui ont été associées à des assemblages de diatomées, les cartes de fréquence mensuelles montrent, pour la première fois, une variabilité phytoplanctonique spatiale et temporelle en accord avec la littérature de cette région (Reid *et al.*, 1990; Beaugrand *et al.*, 2001). De plus, les distributions spatio-temporelles des  $Ra(\lambda)$  labellisés montrent des structures cohérentes avec les conditions physiques et hydrologiques en Mer du Nord qui induisent la distribution du phytoplancton (Reid *et al.*, 1990). En effet, quatre régions écologiques (Figure 5.9) sont communément décrites en Mer du Nord, sur la base d’études qui concernent le phytoplancton, le zooplancton et les poissons (Reid *et al.*, 1990; Fransz *et al.*, 1991; Daan *et al.*, 1990; Beaugrand *et al.*, 2001; Van Leeuwen *et al.*, 2015). Le front Flamborough (en blanc sur la Figure 5.9) est une région transitoire entre le Nord de la Mer du nord, saisonnièrement stratifié (en vert sur la Figure 5.9), et le Sud de la Mer du Nord, où les eaux sont plutôt mélangées (en bleu et violet sur la Figure 5.9) (Otto *et al.* (1990), Ducrotoy *et al.* (2000)). La quatrième région est le Skagerrak, qui est associé au courant Kattegat-Skagerrak (en rouge sur la Figure 5.9) (Otto *et al.*, 1990). Ces quatre régions ont des conditions différentes en termes d’apports de nutriments, de température, de lumière disponible, ce qui expliquerait que des assemblages de phytoplancton (incluant les diatomées) s’y développent différemment en accord avec leurs préférences physiologiques (Margalef, 1978). La distribution spatiale des dix  $Ra(\lambda)$  labellisés à ce jour distingue ces quatre régions écologiques.

À partir des données *in-situ*, nous avons vu que l’assemblage caractérisé par *Thalassionema nitzschooides* est dominant en hiver, a une efflorescence au début du printemps et en automne en Mer du Nord (Figure 5.8, c). Dans la littérature, l’espèce caractéristique de l’assemblage, *Thalassionema nitzschooides*, est plutôt localisée au Sud de la Mer du Nord, en relation avec des conditions environnementales spécifiques (e.g., Edwards *et al.* (2001), Rousseau *et al.* (2002), Wiltshire *et al.* (2010)). En effet, cet assemblage est commun dans les eaux *néritiques* (des côtes jusqu’au plateau continental) avec une grande quantité de nutriments (Van Iperen *et al.*, 1993), comme c’est le cas en automne et au printemps dans le Sud Mer du Nord (Commission, 2010). La distribution spatio-temporelle des Ra( $\lambda$ ) associés à ce jour montre beaucoup de similarités avec les connaissances des occurrences de cet assemblage.

Selon les données *in-situ*, l’assemblage caractérisé par *Rhizosolenia imbricata* a une efflorescence au début du printemps qui se maintient puis décroît progressivement jusqu’à l’automne (Figure 5.8, f), et qui d’après la littérature, est une espèce côtière (Hinder *et al.*, 2012). Ceci trouve une explication potentielle dans des expériences en laboratoire qui ont montré que *Rhizosolenia imbricata* utilise préférentiellement les nitrates ( $\text{NO}_3^-$ ) à l’ammonium ( $\text{NH}_4^+$ ), même si elle est également efficace dans l’incorporation de la forme ammonium (Villareal *et al.*, 1996). Ainsi, *Rhizosolenia imbricata* est assez peu limitée dans l’utilisation de ce nutriment, ce qui lui permet d’être présente sur une longue période de l’année dans les eaux mélangées ou peu stratifiées. Les Ra( $\lambda$ ) associés à cet assemblage ont une distribution spatio-temporelle similaire, localisée au Sud de la mer du Nord sur une grande période de l’année.

Pour *Thalassiosira spp.*, des études ont montré une dominance hivernale, une efflorescence printanière puis une seconde efflorescence automnale depuis les années 2000 (Hinder *et al.* (2012)), ce qui est également visible sur les données *in-situ* (Figure 5.8, e). En revanche, les résultats de notre étude montrent, à ce stade, que les Ra( $\lambda$ ) labellisés ont leur maximum de fréquence en été, au centre de la Mer du Nord. Cependant, ces différences peuvent être dues à un manque d’information pour correctement détecter cet assemblage actuellement. En effet, les mesures coïncidentes pour cet assemblage sont nombreuses en été : 71,5% des mesures coïncidentes de l’assemblage *Thalassiosira spp.* sont situées entre Mai et Août, alors qu’il reste peu de mesures coïncidentes réparties sur les autres mois (28.5%). De plus, uniquement une classe de Ra( $\lambda$ ) a été associée à cet assemblage, ce qui rend cette labellisation peu représentative pour le moment. Bien que ce Ra( $\lambda$ ) ait été correctement labellisé selon les scores de validation (environ 60% d’identification correcte), il y a probablement un ou plusieurs autres signaux qui correspondent à cet assemblage pour les autres périodes de l’année, ce qui est impossible à déterminer avec les mesures coïncidentes actuelles. Par ailleurs, comme détaillé dans les résultats de la méthode IndVal, le taxon *Thalassiosira spp.* est eurytopique, presque ubiquiste. Bien qu’il soit fortement

caractéristique d'un groupe de sites (ce qui a permis de l'associer à une classe de  $Ra(\lambda)$ ), il se peut que sa présence quasi-permanente forme une sorte de "bruit de fond" sur le signal d'anomalie de luminance. Cela expliquerait pourquoi, malgré le plus grand nombre de mesures coïncidentes de tous les assemblages caractérisés, celui-ci ne puisse pas être associé à un grand nombre de classes de  $Ra(\lambda)$ . Tout ceci restera bien sûr à confirmer ou infirmer à l'avenir.

### 5.3.3 Variabilité des caractéristiques spectrales des $Ra(\lambda)$ labellisés en amplitude, forme et phénologie

Les principales caractéristiques des spectres de  $Ra(\lambda)$  sont la forme et l'amplitude (Alvain *et al.*, 2005, 2008, 2012; Ben Mustapha *et al.*, 2014). Comme expliqué auparavant, les classes de  $Ra(\lambda)$  ont aussi été classées en accord avec leur phénologie (voir Chapitre 3, page 55), puis associées empiriquement avec des compositions spécifiques en diatomées.

#### 5.3.3.1 Variations des caractéristiques de $Ra(\lambda)$ intra- et inter-assemblages

##### Généralités

Un assemblage spécifique de diatomées peut être associé à plus d'une classe de  $Ra(\lambda)$  (par exemple, l'assemblage caractérisé par *Thalassionema nitzschoides* a été associé à six classes de  $Ra(\lambda)$ ). Ces spectres associés peuvent avoir des caractéristiques différentes (forme et amplitude). À ce stade de connaissances, cette observation n'est pas surprenante puisque la forme et l'amplitude des  $Ra(\lambda)$  varient avec les propriétés optiques inhérentes. En effet, selon l'étude théorique de Alvain *et al.* (2012), l'absorption phytoplanctonique, la rétrodiffusion particulaire et les matières dissoutes associées au phytoplancton sont une combinaison de facteurs qui influence les caractéristiques des  $Ra(\lambda)$  (voir Chapitre 2, page 27). Voici une liste d'hypothèses ou de considérations qui peuvent expliquer le fait que plusieurs classes de  $Ra(\lambda)$ , ayant des caractéristiques différentes, peuvent être associées à un même assemblage :

- Le stade de croissance des cellules peut induire différentes tailles cellulaires, différentes morphologies et différents arrangements inter-cellulaires (cellules libres, cellules formant des chaînes, voire des agrégats).
- La concentration et la composition en pigments peuvent varier en fonction de nombreux facteurs structuraux mais aussi en acclimatation à une modification de l'environnement lumineux.
- La concentration des matières dissoutes associées varie en fonction du stade de croissance de la *population* phytoplanctonique puisque, par exemple, la concentration en CDOM est plus importante en fin d'efflorescence.

- La communauté phytoplanctonique associée à l'assemblage de diatomées peut être très variée, ce qui n'est pas encore connu dans son intégralité actuellement sur la base des données *in-situ* disponible.

**Tous ces paramètres font varier dans des proportions différentes les propriétés optiques inhérentes de l'eau à la fois entre les assemblages ou au sein d'un même assemblage. Sans expérimentation en milieux contrôlés et avec un ensemble complet de mesures optiques et de caractérisation de la composition et des stades d'évolution du phytoplancton, il est impossible de connaître l'amplitude des modifications engendrées par tel ou tel facteur. Cette constatation illustre la nécessité d'utiliser des approches fines de classification des  $Ra(\lambda)$  à la place des grandes enveloppes statistiques grâce à l'utilisation de la phénologie des classes de  $Ra(\lambda)$  ; mais aussi la nécessité absolue de disposer dans le futur de grandes bases de données pour définir les limites de la méthode PHYSAT. Une approche globale, intégrant d'autres paramètres - environnementaux par exemple -, devra également être envisagée (détaillé au Chapitre Discussion générale).**

Cependant, dans l'état actuel des connaissances, nous pouvons d'ores et déjà conjecturer les sources de variation des caractéristiques des  $Ra(\lambda)$ , à la fois pour deux assemblages différents (paragraphe variations inter-assemblages), mais également au sein d'un même assemblage (paragraphe variations intra-assemblages).

### **Variations des $Ra(\lambda)$ inter-assemblages**

Les diatomées sont un groupe taxonomique diversifié. Les hypothèses suivantes peuvent être proposées pour expliquer les différences observées sur les caractéristiques des  $Ra(\lambda)$  associés entre les assemblages à ce jour :

- La morphologie des cellules des espèces caractérisant les assemblages sont différentes. En effet, les cellules sont de type "pennées" (symétrie axiale) ou "centriques" (symétrie radiaire). *Thalassionema nitzschoides* et *Rhizosolenia imbricata* font partie des diatomées pennées et ont donc plutôt une forme allongée alors que *Thalassiosira spp.* est une diatomée centrique et a donc une forme de cylindre.
- Les cellules peuvent s'organiser entre elles. En effet, certaines diatomées peuvent former des chaînes de cellules (c'est le cas de *Thalassionema nitzschoides*, *Thalassiosira spp.* et *Rhizosolenia imbricata*) qui ont également des formes variées : en étoile, en zig-zag ou encore linéaire.
- La taille des cellules est également variable entre les espèces qui caractérisent les assemblages (de 8 à 140  $\mu\text{m}$  pour les cellules libres), la plus petite ici étant les

*Chaetoceros spp.* (12 à 45  $\mu\text{m}$ ) et les *Thalassiosira sp* (8 à 60  $\mu\text{m}$ ) et la plus grande *Pseudonitzschia seriata* (75 à 140  $\mu\text{m}$ ) (Castellani et Edwards, 2017).

- Les assemblages peuvent être associés à d'autres communautés phytoplanctoniques.

### Variations des $Ra(\lambda)$ intra-assemblages

Les variations des  $Ra(\lambda)$  associés au sein d'un même assemblage pourraient également avoir des origines variées, dont les origines pourraient être les suivantes :

- Au niveau de l'organisation des cellules entre-elles, le basculement entre des cellules libres et des cellules enchainées pourrait avoir un impact sur l'anomalie de luminance. De plus, certaines espèces sont capables de former plusieurs types de chaines, comme *Thalassionema nitzschoides* qui forme des chaines en "zigzag" ou en "étoile" (Figure 5.10). Ces facteurs auraient potentiellement un impact sur la contribution du coefficient de rétrodiffusion aux caractéristiques des  $Ra(\lambda)$ .
- La taille cellulaire varie énormément pour une espèce au cours de sa vie en fonction du stade de croissance (en incluant les facteurs environnementaux) et du mode de reproduction (sexuée ou asexuée). Ces facteurs auraient une influence sur la contribution du coefficient de rétrodiffusion aux caractéristiques des  $Ra(\lambda)$ .
- Dans un même assemblage de diatomées, il est possible que la communauté phytoplanctonique associée se modifie, par exemple en passant par des compositions diverses en dinoflagellés, picophytoplancton et autres microphytoplancton. Cette succession phytoplanctonique, qui existe dans tous les systèmes biologiques, a un impact sur le signal optique mesuré. Ces facteurs auraient un impact sur la contribution du coefficient de rétrodiffusion et de l'absorption phytoplanctonique aux caractéristiques des  $Ra(\lambda)$ .
- En fonction des conditions de lumière, les cellules d'un même assemblage peuvent modifier (ou adapter) leurs pigments de manière qualitative et quantitative afin d'être plus efficace dans la réalisation de la photosynthèse, ce qui aurait un impact sur la contribution de l'absorption phytoplanctonique aux caractéristiques des  $Ra(\lambda)$ .
- Enfin, la présence plus ou moins importante de CDOM a un impact sur l'absorption du CDOM et donc sur les caractéristiques des  $Ra(\lambda)$ .



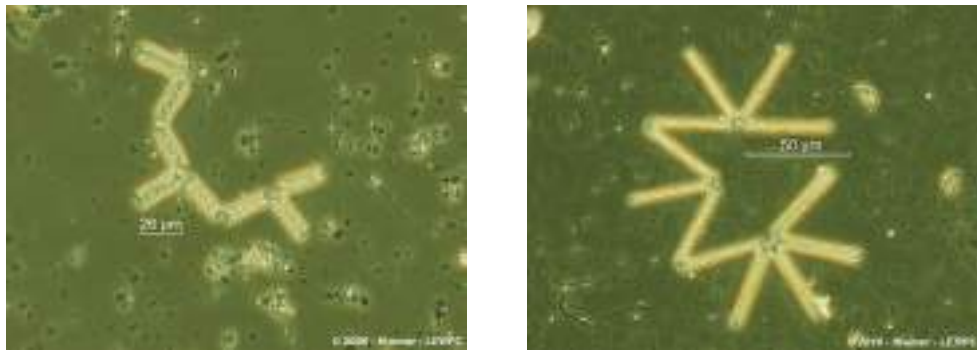


FIGURE 5.10 – Exemple d'arrangements de cellules qui concerne l'espèce de diatomée *Thalassionema nitzschooides* avec à gauche un arrangement en chaîne "zig-zag" et à droite un arrangement en chaîne "étoilée". Source : IFREMER Laboratoire LER/PC, <https://wwz.ifremer.fr/lerpc/> consulté en Août 2017.

### Perspectives sur la caractérisation des $Ra(\lambda)$

Toutes les hypothèses citées ci-dessus méritent des études plus approfondies, en considérant à la fois les paramètres optiques et une caractérisation aussi complète que possible des communautés phytoplanctoniques, et des matières dissoutes associées. Actuellement, il est impossible d'isoler la contribution des propriétés optiques inhérentes sur la variation des anomalies de luminances sans expériences en conditions contrôlées. Cependant, la robustesse des associations trouvées entre les  $Ra(\lambda)$  et les assemblages *in-situ*, ainsi que les scores de validation obtenus sont encourageants pour poursuivre les études visant à étudier le potentiel d'utilisation des  $Ra(\lambda)$  pour la détection d'assemblages de phytoplancton, notamment dans des régions bien documentées.

#### 5.3.3.2 Cas de "labellisations multiples"

La procédure de labellisation a été répliquée afin d'observer la robustesse de cette étape. Cette procédure a montré que certaines classes de  $Ra(\lambda)$  ont été associées à différents assemblages au cours des réplifications, en fonction du set de calibration (voir Figure 5.3).

Deux communautés phytoplanctoniques qui entraînent la même anomalie de luminance sont de rares cas selon l'étude théorique (Alvain *et al.*, 2012), mais restent possibles. Ainsi, il est possible de faire l'hypothèse, qu'en fonction de la combinaison de nombreux facteurs, que deux assemblages de diatomées que nous avons définis ici soient associés à la même anomalie de luminance. Ceci pourrait être expliqué par un signal optique porteur d'une anomalie de luminance non suffisamment spécifique. Par exemple, le cas de la classe de  $Ra(\lambda)$  n°25, qui a été labellisée pour moitié *Chaetoceros spp1.* et pour autre moitié *Chaetoceros spp2.* pourrait être associée à un autre groupe phytoplanctonique, ou une espèce particulière qui est souvent associée avec des diatomées de type *Chaetoceros spp.* Malheureusement, cet aspect n'est pas vérifiable avec les données actuelles puisque

le CPR n'enregistre que les abondances du microphytoplancton et qu'il est donc impossible de caractériser la communauté de pico- et nano-phytoplancton associée, qui peuvent cependant avoir une influence importante sur le signal optique (Alvain *et al.*, 2012).

Au stade actuel de l'étude, il est prématuré de tirer des conclusions sur le fait qu'un type d'anomalie de luminance pourrait être identique pour deux assemblages (ou plus). En effet, bien que nous considérons que ce cas est tout-à-fait possible (Alvain *et al.*, 2012), il est important de garder à l'esprit que notre étude s'est portée uniquement sur les diatomées de grande taille échantillonnées par le CPR. Ainsi, il reste, en Mer-du-Nord, une grande diversité d'autres groupes phytoplanctoniques qui pourraient être associés à ces cas. Le doute concernant cet aspect de PHYSAT ne pourra pas être levé tant que la communauté ne sera pas déterminée de manière exhaustive.

### 5.3.4 Validation des associations

Les scores de validation observés pour les trois assemblages sont assez élevés ( $\simeq 74$ , 95 et 64% respectivement pour les assemblages caractérisés par *Thalassionema nitzschoides*, *Thalassiosira spp.* et *Rhizosolenia imbricata*). Dans le calcul de ces scores de validation, l'absence d'une classe de  $Ra(\lambda)$  a été considérée comme une erreur. Par exemple, pour l'assemblage *Rhizosolenia imbricata*, chacune des trois classes de  $Ra(\lambda)$  associée (n°18, 29 et 36) a été validée indépendamment, c'est-à-dire que chaque classe labellisée a été cherchée dans les données de validation. Si on considère la théorie qui explique les caractéristiques des anomalies de luminance, ces trois  $Ra(\lambda)$  associés à l'assemblage *Rhizosolenia imbricata* pourraient apparaître à des périodes particulières en lien avec le stade de croissance du phytoplancton. On note que, si on réunit les  $Ra(\lambda)$  pour la validation, c'est-à-dire que la présence d'un seul d'entre-eux suffit pour comptabiliser une identification correcte, les taux de validation augmentent fortement, atteignant  $\simeq 100\%$  pour les trois assemblages robustement associés à des classes de  $Ra(\lambda)$ .

L'étude du cycle de vie d'un assemblage permettrait de comprendre pourquoi différentes amplitudes et formes spectrales des  $Ra(\lambda)$  se succèdent. Ce type d'étude permettra, à termes, de décider si les scores de validation doivent être conservés par classes de  $Ra(\lambda)$ , ou indépendamment de celles-ci. Des pistes de recherches sont données au Chapitre 6 (page 223).

### 5.3.5 La base de données de $Ra(\lambda)$ en Mer du Nord, limitations de l'approche et perspectives

Dans ce premier essai d'évaluation du potentiel des anomalies de luminance dans la détection de situations phytoplanctoniques très précises, au-delà la dominance, la méthode employée et les sélections ont été strictes, par exemple le seuil de 70% sur les taux de calibration. Les futures extensions de la méthode, avec des données *in-situ* supplémentaires

(non limitées aux diatomées), devraient permettre d’établir une balance entre ce degré de sélection, la disponibilité des données et les scores de validation.

Actuellement, la méthode proposée a permis d’identifier des associations entre assemblages *in-situ* et  $Ra(\lambda)$  pour 11,7% de la base de données  $Ra(\lambda)$  de la Manche Mer du Nord. À ce stade, il reste donc la majeure partie des classes de  $Ra(\lambda)$  qui sont disponibles pour étendre la méthode et la compréhension des résultats. En effet, il reste 88,3% de la base de données de classes de  $Ra(\lambda)$  qui n’a pas été associée avec une situation phytoplanctonique spécifique. Ces classes de  $Ra(\lambda)$  peuvent représenter :

- i Un des assemblages de diatomées décrit ici, à un stade de croissance différent, ou associé avec des espèces phytoplanctoniques différentes, dont les mesures coïncidentes ne sont pas disponibles actuellement ;
- ii Un autre assemblage de phytoplancton non-identifié. En effet, la Manche-Mer du nord est un environnement très variable pour le phytoplancton, et ayant une grande diversité d’espèces et de groupes ;
- iii Différents assemblages pour lesquels la différentiation est impossible à l’heure actuelle.

Même si la Mer du Nord est principalement dominée par les diatomées et les dinoflagellés (en ce qui concerne le microphytoplancton), des études ont également montré une présence non négligeable de *Prymnesiophyceae* (comme les espèces de type *Phaeocystis sp.*), de *Chlorophyceae*, de Cyanobacteries, et également de Coccolithophores (Gieskes et Kraay (1983), Reid *et al.* (1990), Hoppenrath (2004), Hoppenrath *et al.* (2007), Sapp *et al.* (2007), Masquelier *et al.* (2011), Bonato *et al.* (2015)).

Bien que cette étude ait montré le potentiel de détection de cas phytoplanctonique très précis grâce aux anomalies de luminance de PHYSAT, il reste de nombreux développements à réaliser pour caractériser la communauté phytoplanctonique de manière plus détaillée. Les observations PHYSAT issues du croisement empirique des données *in-situ* et des données satellites ne sont pas quantitatives, ce qui peut rester un problème pour les utilisateurs. En revanche, ces observations qualitatives peuvent être utiles pour estimer la contribution des assemblages à l’estimation de la concentration en chlorophylle-*a* (Alvain *et al.* (2006), Alvain *et al.* (2008)) et compléter des approches comme OC-PFT (Hirata *et al.*, 2011), et PhytoDAS (Bracher *et al.*, 2009).

Cette étude souligne, avant tout, le besoin de données *in-situ* exhaustives et répétitives qui permettront de mieux caractériser la communauté phytoplanctonique et de mieux comprendre les limitations de l’approche présentée. Quelques études préliminaires réalisées durant cette thèse montrent que les classes de  $Ra(\lambda)$  qui ont été associées de manière

robuste à un assemblage de diatomées ne sont pas associées à d'autres groupes de microphytoplancton, en particulier les dinoflagellés. Ceci pourrait confirmer que l'observation de cas de "labellisation multiple" dans la procédure répétitive de calibration serait effectivement due à la présence d'autres types phytoplanctoniques. D'après le premier test réalisé ici, les signaux obtenus et leurs répartitions spatiales et temporelles sont cohérents avec les connaissances passées et encourage à ne pas négliger le potentiel d'utilisation des classes de  $Ra(\lambda)$  pour la détection d'assemblages phytoplanctoniques. Les études futures devront en définir les limites en considérant les points suivants (détaillés en Chapitre 6) :

- Une caractérisation plus détaillée de la communauté phytoplanctonique sera nécessaire, ainsi que des mesures des propriétés optiques - notamment l'absorption du phytoplancton, le coefficient de rétrodiffusion et l'absorption des matières dissoutes - pendant un bloom entier (initiation / maximum et fin de bloom).
- Des expériences de mesures en milieu contrôlé, afin de développer notre compréhension théorique des signaux observés, et d'espérer à terme ajouter un indice de confiance sur les observations obtenues. *In fine*, des couplages des paramètres **IOPs** pourraient être utilisés comme paramètres de base, en amont des  $Ra(\lambda)$  qu'ils induisent.
- Une méthode de suivi spatial *lagrangienne*, avec de nombreuses mesures coïncidentes, pourrait nous permettre d'analyser les variations des caractéristiques des anomalies de luminance (forme et amplitude) en suivant l'évolution d'un bloom particulier dans son parcours spatio-temporel.
- Le besoin de réaliser des algorithmes régionaux devra également être estimé. D'autres études ont déjà montré que des algorithmes de couleur de l'océan spécifiques à une région sont parfois nécessaires (par exemple le cas de la mer Méditerranée, Volpe *et al.* (2007)). L'approche régionale développée dans ce Chapitre n'est en effet pas comparable à des approches plus globales à cause de l'utilisation de table de référence spécifique pour le calcul de l'anomalie lors de cette étude "test". Des algorithmes régionaux pourraient en effet être plus efficaces en limitant les possibilités des combinaisons d'IOPs conduisant à des  $Ra(\lambda)$  identiques.
- Comme énoncé dans le Chapitre 2, la méthode PHYSAT requière une table de référence spécifique à chaque capteur de couleur de l'océan. Actuellement, il est complexe d'évaluer les incertitudes liées à la projection d'une calibration de PHYSAT d'un capteur à l'autre. De plus, les impacts d'une intercalibration multi-capteurs doivent être étudiés dans le cadre de PHYSAT, bien que l'utilisation de bases de données standardisées a déjà été réalisée par Navarro *et al.* (2017) avec l'algorithme PHYSAT-MedOC-CCI. Il faut noter cependant que cette étude est régionale et utilise des algorithmes spécifiques à la mer Méditerranée (notamment pour l'esti-

mation de la concentration en Chl-*a*), ce qui réduit probablement la propagation des erreurs dans le calcul final de l’anomalie de luminance.

## 5.4 Conclusion

Bien qu’il reste des études essentielles à réaliser pour comprendre précisément les observations empiriques de cette étude, les scores de validation issus de cette méthodologie sont encourageants et mettent en évidence qu’il reste un grand nombre de classes de  $Ra(\lambda)$  non-attribuées à un assemblage spécifique. Cette observation est particulièrement rassurante. En effet, si l’on considère le peu d’informations *in-situ* disponibles à ce jour (uniquement le microphytoplancton de type diatomées), il aurait été anormal d’associer un grand nombre de classes de  $Ra(\lambda)$  aux données *in-situ*, sauf si les diatomées ont un grand impact sur le signal d’anomalie de luminance en Mer du Nord, ce qui reste à déterminer.

# Conclusion générale et perspectives

---

Dans le cadre des travaux présentés dans ce manuscrit, les éléments suivants ont été étudiés et ajoutés à la nouvelle chaîne de traitement PHYSAT :

*Chapitre 3* Une classification secondaire post-carte auto-organisatrice a été réalisée sur les classes d'anomalies de luminance afin de regrouper celles qui avaient des caractéristiques spectrales et spatio-temporelles similaires, *via* l'étude de leur *phénologie*. Cette étude a permis d'optimiser la classification des anomalies de luminance dans le cadre de la calibration de PHYSAT, au-delà de leur aspect spectral.

*Chapitre 4* Les classes d'anomalies, regroupées en fonction de leurs caractéristiques spectrales (*SOM*) et temporelles (*via* la phénologie), ont fait l'objet d'une nouvelle calibration qui a permis de les labelliser en termes d'*assemblages* déterminés *in-situ* à partir de concentrations en *pigments biomarqueurs*. Pour cela, une nouvelle approche d'analyse des données pigmentaires a été proposée.

*Chapitre 5* Une étude visant à évaluer le potentiel de PHYSAT pour la détection d'anomalies de luminance associées à des assemblages de phytoplancton a été menée en Mer du Nord, à partir des données issues du CPR (diatomées). Il a ainsi été montré, pour la première fois, qu'il était possible d'aller au-delà de la détection des cas de dominance.

Cependant, lors de ces études, de nombreuses questions ont été soulevées et constitueront la base de futures recherches. Parmi les points qu'il conviendra d'étudier, nous souhaitons cibler plus spécifiquement les limites de l'utilisation des anomalies de manière globale ou régionale. Dans ce dernier Chapitre, nous apportons ainsi des pistes ou éléments de réponses pour les travaux à venir.

## 6.1 Limites de la méthode PHYSAT : bilan actuel et propositions d'études *via* l'expérimentation (milieux contrôlés) et/ou l'approche théorique

Durant cette thèse, nous avons suggéré à plusieurs reprises que des mesures *in-situ* supplémentaires seraient nécessaires, voire des expérimentations ou de la modélisation, pour étendre et améliorer la calibration et validation de PHYSAT. Nous proposons donc ci-dessous quelques pistes qui permettraient, *in fine*, d'évaluer précisément les limitations de la méthode et d'apporter des améliorations qui pourraient en découler à plus long terme.

### 6.1.1 Limites liées à l'aspect empirique

En milieu naturel, l'aspect empirique de la méthode PHYSAT ne permet pas toujours de comprendre les caractéristiques des signaux d'anomalies de luminance observés. Dans cette optique, des mesures *in-situ* d'**IOPs** (absorption des matières dissoutes, absorptions phytoplanctoniques et rétrodiffusion), couplées à l'observation du phytoplancton, pourraient permettre de répondre aux questions liées à la spécificité du signal d'anomalie de luminance : **les gammes de variation des IOPs sont-elles spécifiques à la présence d'un groupe phytoplanctonique particulier ? La combinaison de ces IOPs induira-t-elle un signal spécifique en terme de forme et d'amplitude ?**

Pour cela, des campagnes en mer de type **lagrangiennes** seraient idéales<sup>1</sup>, puisqu'elles permettraient de suivre les IOPs d'une masse d'eau, tout en caractérisant sa composition et la succession phytoplanctonique de manière précise (par exemple, par comptage, **cytométrie en flux**, et **HPLC** en complément). Une zone d'étude idéale dans le cadre de PHYSAT devrait présenter : une bonne couverture satellite (pour augmenter les chances d'avoir une bonne couverture spatiale des mesures coïncidentes) et surtout une bonne représentation de la diversité phytoplanctonique détectée par PHYSAT avec une saisonnalité marquée. Des zones idéales (bien qu'elles soient nuageuses) seraient donc l'Océan Austral et l'Atlantique Nord, où les groupes de PHYSAT sont bien représentés (chaque groupe est fréquent sur une partie de cette zone pour une partie de l'année), et montrent une forte saisonnalité, avec une succession des groupes phytoplanctoniques (qui a été observée dans les versions précédentes de PHYSAT, dans la v2017, mais aussi d'après de nombreuses autres études *in-situ*).

L'approche lagrangienne permettrait aussi d'évaluer la modification de l'anomalie de luminance au cours de tous les stades d'une efflorescence phytoplanctonique, qu'elle ait une composition dominante ou mixte. En effet, des études ont montré que les IOPs pouvaient varier significativement en fonction du stade de croissance d'une efflorescence monospé-

---

1. Une approche non lagrangienne permettrait également d'obtenir des informations importantes pour PHYSAT, mais l'approche lagrangienne possède des avantages cités ensuite.

cifique (e.g., expérience en milieu contrôlé, Moutier *et al.* (2017)), et ainsi comprendre les variabilités observées des caractéristiques des spectres associés empiriquement à des groupes phytoplanctoniques : les anomalies de luminance sont-elles modifiées, et dans quelle mesure, au cours d'une efflorescence phytoplanctonique ?

### 6.1.2 Expériences en milieu contrôlé et simulations pour aller plus loin dans la compréhension théorique du signal d'anomalie de luminance

Actuellement, l'étude théorique de Alvain *et al.* (2012) a permis de mettre en évidence que ce sont les variations des combinaisons des IOPs inter-groupes qui induisent l'anomalie de luminance observée. Les auteurs en ont également déduit la moyenne de ces IOPs et un signal d'anomalie de luminance théorique pour chaque groupe.

Les résultats présentés dans ce manuscrit suggèrent également, comme l'étude de Alvain *et al.* (2012), que différents stades de croissance d'une efflorescence pourraient induire une variation dans l'anomalie de luminance (Chapitre 5, à partir de la page 167). Des expériences en milieu contrôlé, couplées à des simulations, pourraient permettre de comprendre l'évolution du signal détecté par PHYSAT au cours du cycle de vie d'une efflorescence.

En effet, des mesures de valeurs d'IOPs d'un milieu contrôlé, pour différents types de phytoplancton (cultures monospécifiques ou mélanges) au cours d'un cycle de vie, pourraient être utilisées pour estimer la luminance et l'anomalie de luminance telles qu'elles auraient été observées dans le milieu naturel (*via* des simulations *Hydrolight* paramétrisé comme dans l'étude de Alvain *et al.* (2012)). Ici, le but n'est pas de simuler une anomalie de luminance telle qu'elle aurait été observée dans la nature<sup>2</sup> (i.e., ses valeurs à chaque longueur d'onde), mais d'observer la succession de l'amplitude et de la forme spectrale de l'anomalie de luminance au cours d'un cycle de vie, et d'en estimer une variation plausible.

Ces observations pourraient nous permettre de mieux comprendre l'évolution du signal d'anomalie de luminance dans le temps au cours d'une efflorescence complète, pour observer et tenter de comprendre les variabilités des  $Ra(\lambda)$  entre les groupes phytoplanctoniques. Ce type d'étude pourrait aussi permettre d'évaluer la variabilité intrinsèque aux groupes phytoplanctoniques des anomalies de luminance pour un cycle de croissance. En effet, nous avons pu observer jusqu'à présent, que certains groupes étaient associés à des gammes de variations d'anomalies de luminance plus importante que d'autres groupes (par exemple, les diatomées par rapport aux *Synechococcus spp.-like*). Il s'agit de comprendre si cette diversité de types de spectres d'anomalies de luminance spécifiques est liée à la biodiversité phytoplanctonique du groupe, ou plutôt à son état physiologique ; comme cela a été fait

---

2. Comme dans l'étude de Alvain *et al.* (2012), nous souhaiterions étudier les différences entre tous les signaux théoriques pour des variations des IOPs des stades de croissance d'un bloom, bien que nous savons que ces valeurs ne seraient pas identiques aux signaux d'anomalie de luminance observés par satellite.



pour les variations des IOPs dans l'expérience en microcosme de Moutier *et al.* (2017).

## 6.2 Intégration de nouveaux paramètres dans l'algorithme PHYSAT : vers l'évaluation des probabilités de présence des groupes et assemblages phytoplanctoniques

Nous avons abordé le fait que la combinaison des IOPs qui forme l'anomalie de luminance pouvait induire des anomalies de luminance similaires pour des combinaisons d'IOPs différentes. Ce fait a été mis en évidence grâce à des simulations théoriques dans l'étude de Alvain *et al.* (2012). Actuellement, nous avons commencé à évaluer ce biais *via* l'utilisation d'une caractéristique supplémentaire de la classe d'anomalie de luminance, qu'est la phénologie d'apparition des classes (qui reflète assez bien leurs distributions spatiales). En effet, il est possible que des classes d'anomalies de luminance qui semblent similaires dans leurs caractéristiques spectrales ne le soient pas en ce qui concernent leur caractéristiques spatiales et temporelles. Actuellement, le regroupement secondaire des classes de  $Ra(\lambda)$  *via* la phénologie, est effectué après la classification spectrale *via* une carte auto-organisatrice. Ainsi, nous avons pu voir qu'il existe des classes dont les caractéristiques temporelles sont "bruitées", ce qui serait dû à une séparation trop stricte de deux types de  $Ra(\lambda)$  similaires spectralement.

À l'avenir, l'idée est d'intégrer d'autres paramètres, qui permettraient de mieux définir des classes à la fois spectrales, mais aussi écologiques.

### Vers des classes spectrales mieux définies

Par exemple, puisque des combinaisons d'IOPs peuvent induire des formes similaires de spectres, l'ajout des IOPs à la classification spectrale (carte auto-organisatrice) des anomalies pourraient permettre de prendre ces cas en considération. Ainsi, les classes de  $Ra(\lambda)$  seraient représentées dans toute leur diversité spectrale, induite par des combinaisons d'IOPs particulières : deux spectres "plats" pourraient par exemple être considérés comme différents (d'un point de vue "origine phytoplanctonique") si l'un provient d'une forte absorption phytoplanctonique compensée par des faibles valeurs d'absorption CDOM et de rétrodiffusion faibles (e.g., "petites cellules"); alors que l'autre provient de valeurs moyennes de ces trois paramètres (e.g., nanoeucaryotes) (Alvain *et al.*, 2012). Bien que l'estimation des IOPs et le calcul des anomalies de luminances proviennent de l'observation du même paramètre (la réflectance), ils portent des informations complémentaires non redondantes.

### Vers des classes écologiques

Une autre perspective de développement de PHYSAT est de définir des classes de  $Ra(\lambda)$  d'une manière plus "écologique", c'est-à-dire en y intégrant des paramètres environnementaux (idéalement estimés depuis l'espace, ou modélisés dans un premier temps pour évaluer l'ap-

proche) qui permettront d'affiner, *in fine*, le nombre de classes de  $Ra(\lambda)$  vers un nombre qui tend à respecter la diversité phytoplanctonique recherchée.

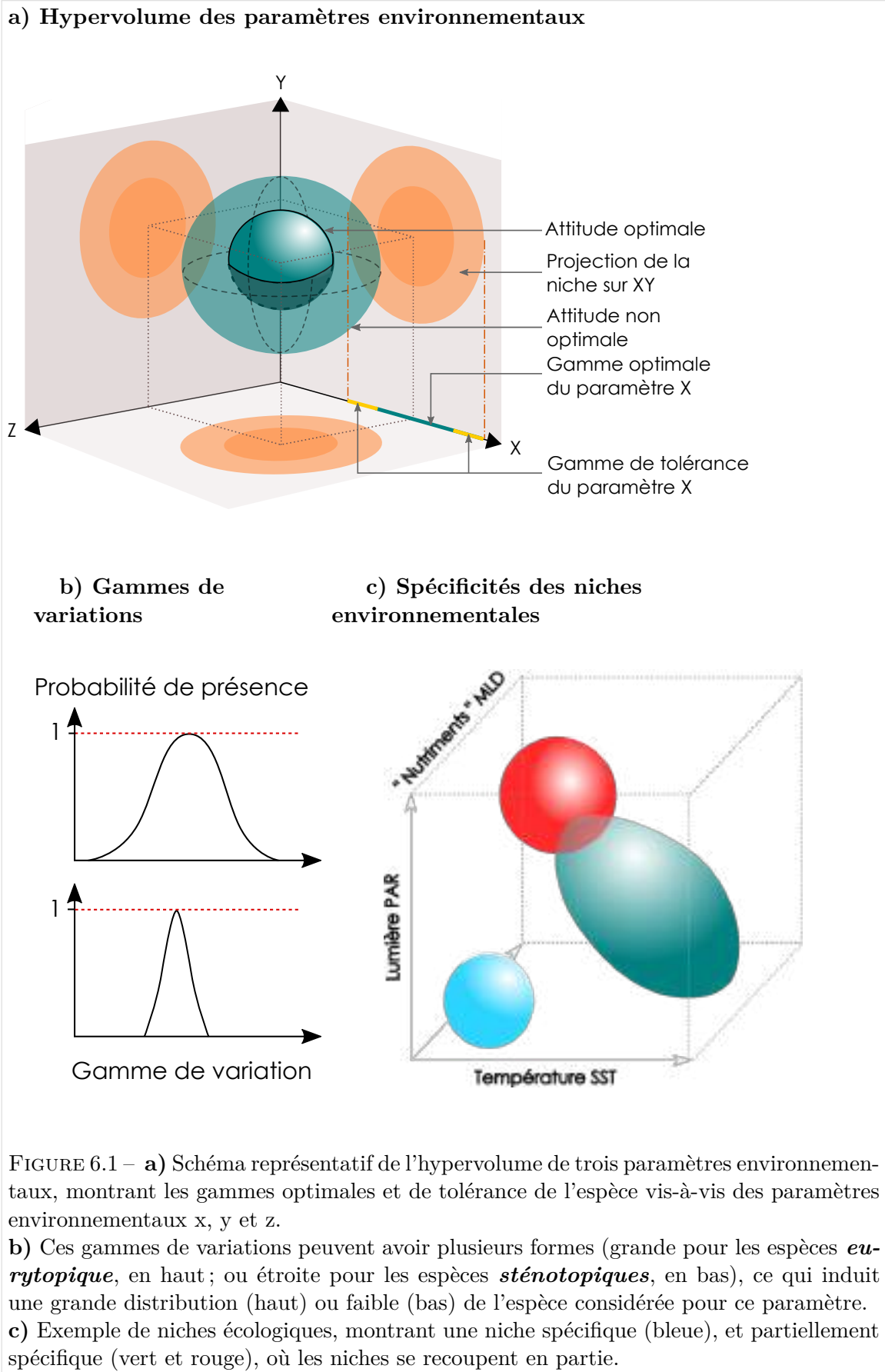
À partir de ces nouveaux outils, nous pourrions à l'avenir, et toujours à l'aide de davantage de mesures coïncidentes, définir des gammes de variations de ces paramètres (optiques et environnementaux) pour des groupes ou assemblages phytoplanctoniques spécifiques, et ainsi déduire des probabilités de présence de ces assemblages pour un pixel. Ce type d'étude pourrait être réalisé à partir du couplage de PHYSAT avec des modélisations du type "NP-PEN" (*Non-Parametric Probabilistic Ecological Niche model*) développé par Beaugrand *et al.* (2011). Ce type de modèle permet, en effet, de définir les *niches écologiques* à partir de données de présence uniquement (comme c'est le cas de PHYSAT), contrairement à d'autres modèles écologiques qui se basent sur des données de présence-absence<sup>3</sup> (e.g., BIOCLIM (Carpenter *et al.*, 1993), DOMAIN (Carpenter *et al.*, 1993), ENFA (Hirzel *et al.*, 2002), MAXENT (Phillips *et al.*, 2006)), ou des données quantitatives (e.g., McCullagh et Nelder (1989), Hastie et Tibshirani (1990), Guisan et Zimmermann (2000), Guisan et Thuiller (2005)). Un tel modèle pourrait être appliqué à plusieurs niveaux d'information sur les données PHYSAT (données brutes, i.e., classes de  $Ra(\lambda)$ ; données de groupes dominants; données d'assemblages); afin de répondre, par exemple, à la question suivante : **les niches écologiques des données PHYSAT sont-elles spécifiques ?** Après avoir défini les niches écologiques (Figure 6.1, a), c'est-à-dire les gammes de variations de paramètres environnementaux pour un groupe particulier de PHYSAT, nous pourrions répondre à la question : **les niches écologiques sont-elles particulières, ou ont-elles des zones de convergence ?** (Figure 6.1, b et c.).

Ces gammes de variation (Figure 6.1, b) pourraient ensuite servir à calculer des probabilités de présence des groupes sur chaque pixels en fonction des conditions environnementales, et ainsi donner un indice de confiance à la présence d'un groupe. Elles pourraient également servir à prédire la présence d'un groupe phytoplanctonique, dès lors que ce groupe aura des niches écologiques spécifiques, et que les données environnementales seraient disponibles à grande échelle.

Bien que le modèle NPPEN soit à "visée écologique" (Beaugrand *et al.*, 2011), ce type d'approche pourrait également être utilisé pour la définition des gammes de distributions des IOPs au sein des groupes de PHYSAT, afin de renforcer le calcul de la probabilité de présence d'un groupe particulier. C'est d'ailleurs une approche similaire qui a été utilisée dans Alvain *et al.* (2012), lorsqu'ils ont définis les valeurs moyennes des IOPs à partir de mesures *in-situ* pour les groupes de PHYSAT, afin de déduire une anomalie de luminance théorique pour chaque groupe dominant (expliqué au Chapitre 2, à partir de la page 27). L'idée serait de réaliser une étude similaire, mais qui tienne compte des gammes de

---

3. La notion de "présence-absence" requiert un effort d'échantillonnage important, afin que la notion d'absence ne soit pas due à un sous-échantillonnage du milieu. Ce type d'approche n'est pas applicable à PHYSAT, qui est basé sur la détection de présence fréquente des groupes phytoplanctoniques.



variations des IOPs, plutôt que des valeurs moyennes pour chaque groupe, et ainsi définir une collection de spectres d'anomalies de luminance théoriques pouvant convenir pour un groupe phytoplanctonique particulier.

Des études préliminaires ont été réalisées durant cette thèse pour l'évaluation des niches écologiques liées aux groupes détectés par PHYSAT (toutes les versions incluses, groupes dominants (Alvain *et al.*, 2008), v2017 et vMixte présentées dans cette thèse). Selon les résultats préliminaires, des environnements spécifiques se dessinent notamment pour les diatomées et les "petites cellules" (*Prochlorococcus spp.*, *Synechococcus spp.-like* et mélange des deux). Ceci a été obtenu en se basant sur la distribution des assemblages phytoplanctoniques PHYSAT, et des paramètres environnementaux suivants : profondeur de la couche de mélange (basée sur le modèle NOBM, Gregg et Casey (2007)), température de surface (AVHRR, Casey *et al.* (2010)), et la lumière disponible pour la photosynthèse (PAR SeaWiFS, Frouin *et al.* (2012)).

Cependant, ces études préliminaires ne sont pas optimales, et seront développées dès la fin de cette thèse. En effet, actuellement, seules les données de PHYSAT d'une année ont été comparées avec ces paramètres environnementaux, ce qui peut représenter une approximation qui conduit à des erreurs de définitions de ces niches environnementales (i.e., les niches n'ont pas un aspect "gaussien", idéalement, comme montré en Figure 6.1, b). Pour cette raison, il serait sans doute préférable de définir les niches écologiques à partir des données *in-situ*. En effet, l'utilisation de données *in-situ* pour la définition des niches écologiques présente deux avantages : i) une base de données *in-situ* conséquente a été analysée pour cette thèse (il suffira alors d'observer les paramètres environnementaux estimés ou modélisés de ces mesures *in-situ*) ; et ii) la définition des niches écologiques serait indépendante des données PHYSAT, ce qui évitera une redondance des données utilisées pour la définition des niches et pour le calcul des probabilités de présence associées aux groupes phytoplanctoniques. Une telle approche pour définir les niches écologiques à partir des données *in-situ* permettra probablement de réduire les approximations.

À ce stade, les résultats préliminaires obtenus sont disponibles en Table 6.1, et permettent de poser les bases des études à venir. À l'heure actuelle nous avons pu observer que :

- les nanoeucaryotes et l'assemblage des diatomées-nanoeucaryotes sont situés dans des environnements similaires (SST +, MLD - et PAR -) ;
- les *Prochlorococcus spp.* non dominants et l'assemblage de *Prochlorococcus spp.* et *Synechococcus spp.-like* sont également situés dans des environnements similaires (SST +, MLD - et PAR+), ainsi que pour les *Prochlorococcus spp.* et *Synechococcus spp.-like* dominants, mais où la température moyenne observée est plus faible ;
- les diatomées dominantes et "dinoflagellés" partagent des environnements similaires

(SST -, MLD + et PAR -).

L'utilisation de données *in-situ* pour la définition des niches écologiques permettra probablement d'affiner les observations ci-dessus, en définissant plus précisément les gammes de variations des trois paramètres (SST, MLD et PAR) ; ou en y intégrant plus de paramètres dans le cas où ces trois paramètres ne permettent pas une distinction suffisante des spécificités des niches écologiques. Par exemple, il existe des distributions (liées à des algorithmes semi-empiriques, empiriques, ou des modélisations) des estimations de concentrations en fer, nitrates, silice, etc., de la durée de la photopériode (en supplément du PAR), de la profondeur de la zone euphotique, etc. ; qui peuvent avoir un impact sur la distribution phytoplanctonique et pourraient permettre de définir des niches écologiques spécifiques.

Cependant, ces résultats préliminaires sont encourageants pour le développement de la méthode PHYSAT en ce sens. En effet, intégrer des probabilités de présence liées à l'environnement et aux IOPs, donnerait un indice de confiance de la répartition des groupes phytoplanctoniques.

TABLE 6.1 – Moyennes et écart-types des distributions des paramètres estimés : SST (en °C, température de surface AVHRR), MLD (en mètre, profondeur de la couche de mélange, modèle NOBM) et PAR (en Einstein/m<sup>2</sup>/jour, SeaWiFS Frouin *et al.* (2012)), pour les 10 groupes de la version Mixte, et les groupes de PHYSAT version 2008 (Alvain *et al.*, 2008).

<b>Version Mixte</b>						
	$\overline{SST}$	sd(SST)	$\overline{MLD}$	sd(MLD)	$\overline{PAR}$	sd(PAR)
Nanoeucaryotes dominants	20,41	8,23	34,15	30,54	43,54	10,14
<i>Prochlorococcus spp.</i> dominants	19,28	7,13	38,23	32,05	49,61	10,49
<i>Synechococcus spp.-like</i> dominants	17,58	7,51	42,10	35,92	52,50	9,74
Diatomées dominantes	15,80	8,08	40,30	39,87	52,99	9,04
<i>Phaeocystis spp.-like</i>	14,74	7,93	43,40	41,96	52,73	9,30
Coccolithes	17,37	8,26	43,12	33,87	52,72	9,22
<i>Prochlorococcus spp.</i> non dominants	22,66	6,44	36,80	27,13	49,51	8,85
Diatomées et nanoeucaryotes	20,74	7,17	37,87	32,98	43,73	10,85
<i>Synechococcus spp.-like</i> et <i>Prochlorococcus spp.</i>	20,97	7,49	36,29	30,85	50,31	8,57
"Dinoflagellés"	16,39	7,50	43,08	37,67	52,02	9,76
<b>Version 2008</b>						
	$\overline{SST}$	sd(SST)	$\overline{MLD}$	sd(MLD)	$\overline{PAR}$	sd(PAR)
Nanoeucaryotes dominants	20,96	7,42	36,42	31,74	43,32	10,40
<i>Prochlorococcus spp.</i> dominants	21,95	6,76	36,33	28,73	49,14	9,24
<i>Synechococcus spp.-like</i> dominants	20,54	7,29	39,43	30,01	51,60	8,97
Diatomées dominantes	16,12	7,95	40,78	38,97	52,70	9,16
<i>Phaeocystis spp.-like</i>	14,37	8,10	45,36	43,96	52,87	9,36
Coccolithes	17,37	8,26	43,12	33,87	52,72	9,22

## 6.3 Évaluation du potentiel de couplage multicapteurs dans le cadre de l'utilisation de PHYSAT

Les études qui concernent la méthode PHYSAT ont été réalisées principalement sur les données du capteur de couleur de l'océan SeaWiFS (e.g., Alvain *et al.* (2005, 2008); D'Ovidio *et al.* (2010); Navarro *et al.* (2014)), tout comme dans ce manuscrit.

Nous avons pu évaluer au cours des Chapitres la difficulté de l'utilisation d'autres capteurs, notamment en raison de l'étape de calibration empirique qui complique le passage d'une information d'un capteur à un autre. Ainsi, l'expertise *in-situ* issue de la base de données GEP&CO, l'une des plus complète pour notre recherche à ce jour, ne peut pas être facilement transposée sur les capteurs qui ont fonctionné / fonctionnent ultérieurement.

Les tests de calibration réalisés à ce jour pour d'autres capteurs ont montré que l'utilisation des périodes de recouvrements et le transfert empirique de la **labellisation** existante de SeaWiFS aux autres capteurs représente l'approche la plus fiable en considérant les informations disponibles. Sans base de données *in-situ* détaillée, la calibration transposée d'un capteur à l'autre reste la meilleure solution à envisager. Les bases de données *in-situ*, ultérieures au début de la période de calibration de PHYSAT, qui sont insuffisantes pour une calibration, pourraient alors être utilisées comme données de validation de la transposition de calibration d'un capteur à l'autre.

Dans le cadre d'une réflexion portant sur le passage d'un capteur à un autre, il est également essentiel de garder à l'esprit que les tables de référence  $nLw_{ref}(\lambda, Chl-a)$  sont dédiées à leur capteur, afin d'adapter le signal moyen (qui intègre également les spécificités des capteurs) pour le calcul des anomalies de luminance. Dans l'idéal, nous devrions déterminer, pour une concentration en Chlorophylle-*a*, un spectre de luminance moyen identique qui faciliterait la transposition d'une calibration de PHYSAT entre capteurs. Cependant, il existe de nombreuses différences dans les traitements des données brutes de couleur de l'océan, dans les approches techniques (e.g., angles de vue), dans les caractéristiques des capteurs (e.g., radiométrie, résolution) et dans les corrections atmosphériques ; qui empêchent actuellement l'utilisation d'une table de référence multi-capteurs, sans une étude plus poussée de son impact. La Figure 6.2 met en évidence les différences qui existent entre les tables de références des capteurs de SeaWiFS, MODIS, MERIS et VIIRS.

À l'heure actuelle, il n'est pas non plus évident d'utiliser les bases de données qui fusionnent les chaînes de traitement des capteurs existants (e.g., base de données OC-CCI, European Space Agency (2017)). En effet, au stade des connaissances actuelles, il est complexe d'estimer les impacts d'un traitement moyen, identique pour tous les capteurs, sur l'anomalie de luminance. Ceci devra faire l'objet d'une étude spécifique avant l'utilisation d'une base de données du type OC-CCI dans le cadre de PHYSAT. Ceci est prévu, dans



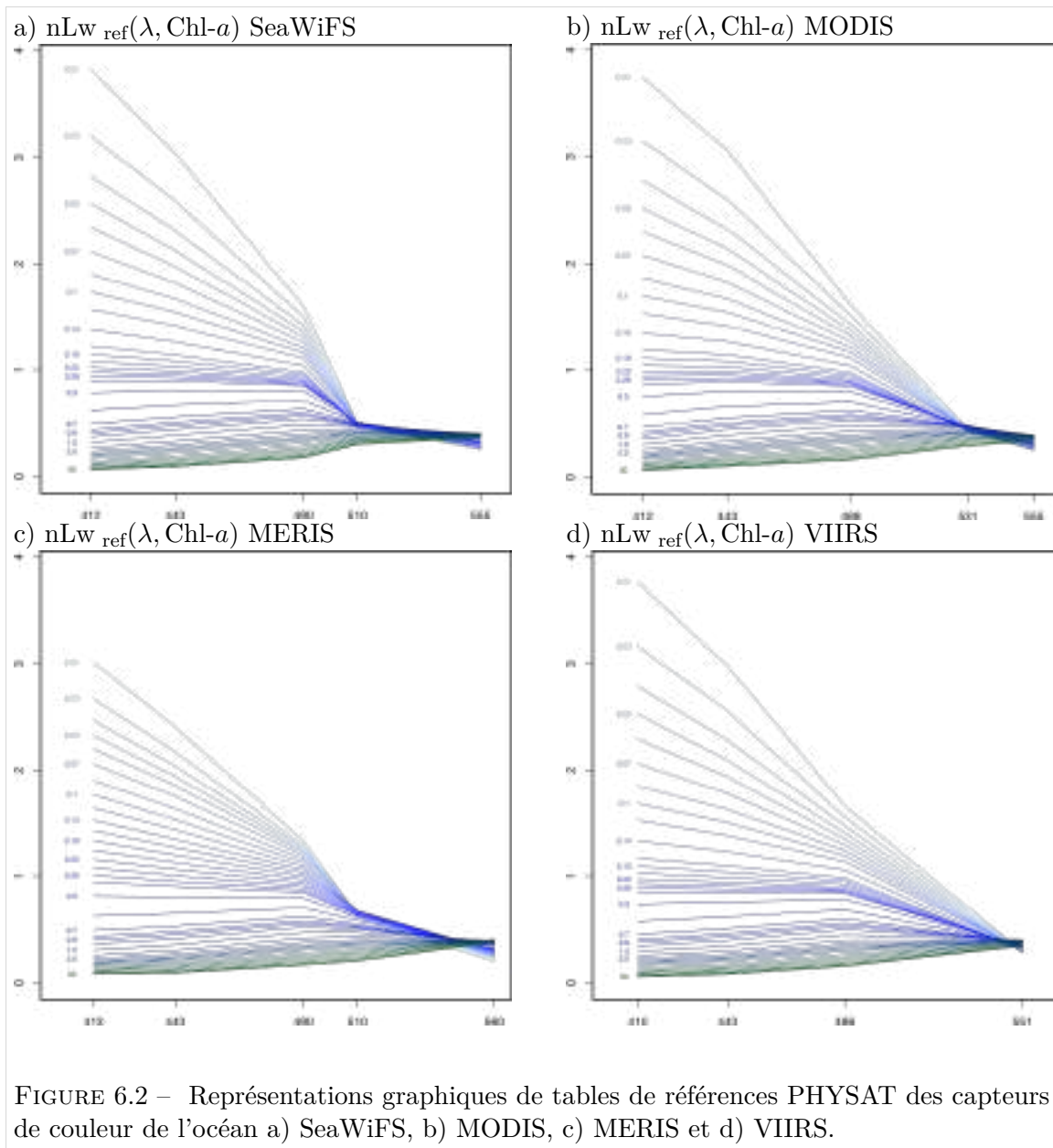


FIGURE 6.2 – Représentations graphiques de tables de références PHYSAT des capteurs de couleur de l'océan a) SeaWiFS, b) MODIS, c) MERIS et d) VIIRS.

un premier temps, sur des zones ciblées pour lesquelles un grand nombre de mesures *in-situ* est disponible, et continuent d'être complétées (par exemple, la Méditerranée, pour laquelle l'approche pourrait être comparée avec l'étude de Navarro *et al.* (2017)).

Des premiers essais d'utilisations de plusieurs capteurs, par transfert d'informations de labellisation *via* les périodes de recouvrements, montrent qu'au niveau des classes de  $Ra(\lambda)$ , la répartition des classes *som* SeaWiFS et MODIS présente des bonnes correspondances (Figure 6.3 a et b), ces dernières étant améliorées en tenant compte de la classification  $Ra(\lambda)$  *phénologie* (Figure 6.3 c et d). De la même manière que la calibration de Alvain *et al.* (2008) a été transférée à la classification *som* de Ben Mustapha *et al.* (2014), il est possible de transférer la calibration du capteur SeaWiFS vers le capteur MODIS en utilisant les périodes de recouvrement (périodes où les capteurs étaient tous les deux en activité). En revanche, ce type d'approche sur-estime probablement le nombre de classes de  $Ra(\lambda)$  labellisés; et ceci est d'autant plus probable que des approximations existaient déjà au niveau de la précédente version de PHYSAT (Alvain *et al.* (2008), Ben Mustapha *et al.* (2014)), avant les améliorations présentées dans ce manuscrit (v2017).

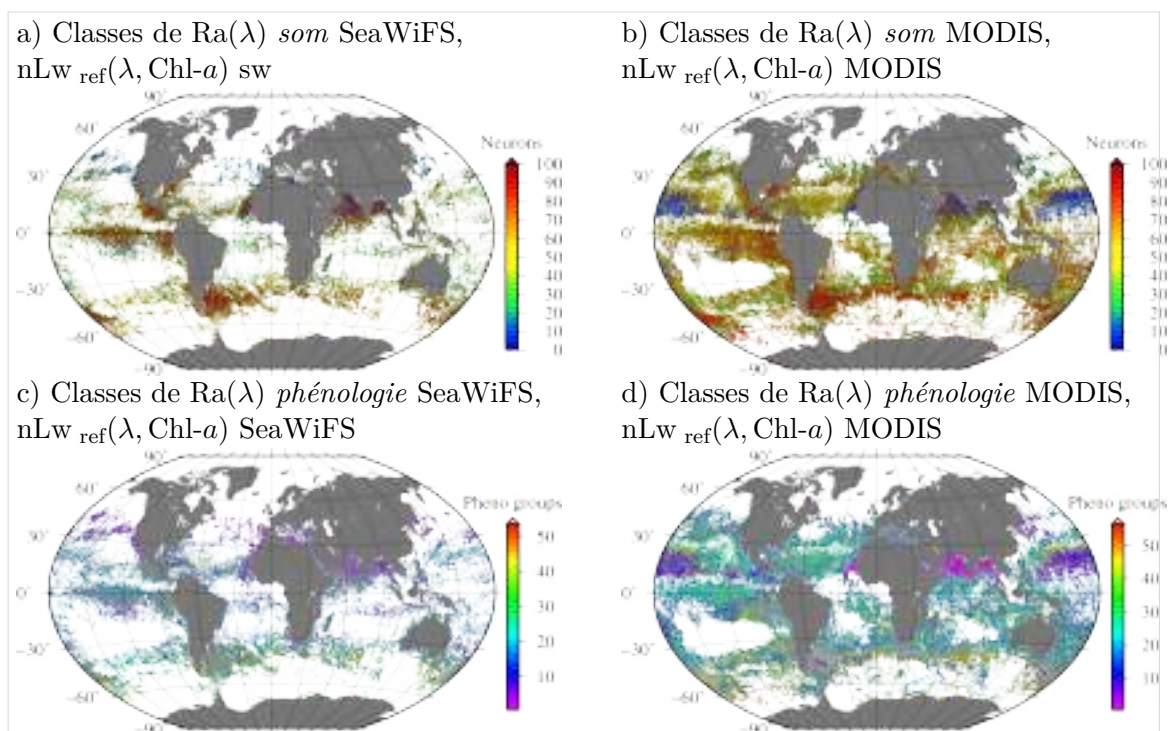


FIGURE 6.3 – Inter-comparaison des distributions spatiales de Janvier 2004 pour : a) les classes de  $Ra(\lambda)$  *som* SeaWiFS (basées sur une référence  $nLw_{ref}(\lambda, Chl-a)$  spécifique au capteur SeaWiFS), b) les classes de  $Ra(\lambda)$  *som* MODIS (basées sur une référence  $nLw_{ref}(\lambda, Chl-a)$  spécifique au capteur MODIS), c) les classes de  $Ra(\lambda)$  *phénologie* SeaWiFS (basées sur une référence  $nLw_{ref}(\lambda, Chl-a)$  spécifique au capteur SeaWiFS), d) les classes de  $Ra(\lambda)$  *phénologie* MODIS (basées sur une référence  $nLw_{ref}(\lambda, Chl-a)$  spécifique au capteur MODIS).

Il existe également une seconde manière de procéder à la labellisation des classes d'anomalies de luminance des autres capteurs. En effet, le calcul d'une référence  $nLw_{ref}(\lambda, Chl-a)$  commune à plusieurs capteurs pourrait permettre de réaliser une calibration *in-situ* commune. Dans la pratique, une table  $nLw_{ref}(\lambda, Chl-a)$  serait calculée à partir de l'ensemble des pixels pour deux capteurs (ou plus), puis une carte auto-organisatrice serait calculée. Ainsi, les classes de  $Ra(\lambda)$  *som* seraient communes pour les deux séries de données.

Une étude plus approfondie est nécessaire afin de définir dans quelle mesure des bases de données fusionnées doivent être utilisées, afin d'éviter la propagation des approximations, que ce soit dans l'établissement de la chaîne de traitement des anomalies de luminance ( $nLw_{ref}(\lambda, Chl-a)$ , calculs des  $Ra(\lambda)$ , carte auto-organisatrice, phénologie) et dans la labellisation.

## 6.4 Conclusion

Lors des travaux présentés dans ce manuscrit, nous nous sommes intéressés en premier lieu à l'amélioration de la classification des anomalies de luminance, *via* l'utilisation d'une classification basée sur les fréquences et les localisations des classes d'anomalies de luminance formées par une carte auto-organisatrice. Cette classification améliore les chances d'obtenir des mesures coïncidentes en évitant la division arbitraire du signal. Elle permet également de prendre en compte les spectres similaires en termes de caractéristiques spectrales, mais qui n'ont pas le même comportement de distribution spatiale et temporelle. Cette étude met en évidence qu'une classification uniquement spectrale des anomalies de luminance ne reflète pas toujours la diversité possible des classes de  $Ra(\lambda)$ , surtout si l'on considère que ces classes ne seront pas labellisées par les mêmes assemblages, et que certains assemblages présentent une variabilité d'anomalies de luminance spécifiques plus importante que d'autres (par exemple, les diatomées).

Nous avons vu qu'il reste des pistes de développements concernant la classification des anomalies de luminance, qui devraient permettre, *in fine*, d'observer si une anomalie de luminance spécifique (forme et amplitude) peut être due à plusieurs combinaisons d'IOPs (conformément à ce qui est prédit par Alvain *et al.* (2012)), et donc d'ajouter un indice de confiance aux données, mais également de donner un sens "écologique" à ces anomalies en y couplant des variables environnementales.

Une seconde étude s'est intéressée à la caractérisation des anomalies de luminance en termes d'assemblages phytoplanctoniques. Des études globales et régionales ont permis de montrer, pour la première fois, qu'il était possible d'associer empiriquement des anomalies de luminance spécifiques à la présence d'assemblages *in-situ*, à partir du moment où l'information de terrain est suffisamment détaillée, et que les signaux satellites sont étudiés

selon les approches présentées dans ce manuscrit. Ces assemblages ont été validés grâce à des mesures *in-situ*, avec de bons scores de validation en comparaison aux scores précédemment trouvés pour cette méthode. Ces résultats pourront être améliorés ou modifiés dès que de nouvelles données seront disponibles.

Là aussi, des pistes de recherches sont données afin de comprendre la variabilité des anomalies de luminance pour des assemblages, mais aussi pour différents stades de croissance d'une efflorescence, *via* des observations en milieux naturels ou en laboratoire.

Enfin, des pistes portant sur les études théoriques, nécessaires à l'avenir, sont également proposées, afin d'améliorer encore notre connaissance des anomalies de luminance, et d'en utiliser tout le potentiel (e.g., identification d'assemblages, définitions d'indices de confiance, modélisation des habitats phytoplanctoniques).



## Valorisation des travaux

### Publications des travaux de cette thèse :

*Acceptée, Annexe 12 (page 309)* A-H RÊVE-LAMARCHE, S. ALVAIN, M-F RACAULT, D. DESSAILLY, N. GUISELIN, C. JAMET, V. VANTREPOTTE, G. BEAUGRAND - **Estimation of the potential detection of diatom assemblages based on ocean color radiance anomalies in the North Sea** *Frontiers in Marine Science - Ocean Observation, Submitted on : 02 Aug 2017, accepted 29 Nov 2017.*

*Soumise* A-H RÊVE-LAMARCHE, S. ALVAIN, D. DESSAILLY, G. BEAUGRAND - **Monthly distribution of diatom from space based on PHYSAT approach** *Frontiers in Marine Science - Marine Biogeochemistry, Submitted on : 30 Nov 2017* in the Research Topic : *Biogeochemistry and Genomics of Silicification and Silicifiers*

### Participation aux travaux de :

Moutier *et al.* (2017) W. MOUTIER , L. DUFORÊT-GAURIER, M. THYSSEN, H. LOISEL, X. MÉRIAUX, L. COURCOT, D. DESSAILLY, A-H. RÊVE, G. GRÉGORI, S. ALVAIN, A. BARANI, L. BRUTIER, M. DUGENNE (2017) **Evolution of the scattering properties of phytoplankton cells from flow cytometry measurements** *PLoS ONE 12(7)*

### Communications :

*Posters* **Ocean Optics "Estimation of the potential detection of diatom assemblages based on ocean color radiance anomalies in the North Sea"** - Rêve A-H., Alvain S., Racault M-F., Dessailly D., Guiselin N., Jamet C., Vantrepotte V., Beaugrand G, en 2014, Portland Maine USA (résultats préliminaires) et en 2016, Victoria British Columbia Canada (résultats publiés).

*Oral* Journées des doctorants du LOG Juin 2015 (Prix du Jury) **Estimation of the potential detection of diatom assemblages based on ocean color radiance anomalies in the North Sea**

Juin 2016 (Prix du Jury) **Détermination des niches écologiques : un enjeu pour la télédétection du phytoplancton**

*Oral* Journées des éléphants de mer, CEBC, Chizé, Octobre 2014 et Avril 2016 **Détection des assemblages phytoplanctoniques depuis l'espace : la méthode PHYSAT ; Implications sur les performances de pêche des éléphants de mer du sud (feat C. Guinet (2014) et G. Labadie (2016))**



# Bibliographie

---

- ALVAIN, S., LE QUÉRÉ, C., BOPP, L., RACAULT, M.-F., BEAUGRAND, G., DESSAILLY, D. et BUITENHUIS, E. T. (2013). Rapid climatic driven shifts of diatoms at high latitudes. *Remote Sensing of Environment*, 132:195–201.
- 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(2):1070–1083.
- ALVAIN, S., MOULIN, C., DANDONNEAU, Y. et BRÉON, F. (2005). Remote sensing of phytoplankton groups in case 1 waters from global seawifs imagery. *Deep-Sea Research I*, 52:1989–2004.
- ALVAIN, S., MOULIN, C., DANDONNEAU, Y. et LOISEL, H. (2008). Seasonal distribution and successuïb of dominant phytoplankton groups in the global ocean : A satellite view. *Global Biogeochemical Cycles*, 22:1–15.
- ALVAIN, S., MOULIN, C., DANDONNEAU, Y., LOISEL, H. et BRÉON, F.-M. (2006). A species-dependant bio-optical model of case i waters for global ocean color processing. *Deep-Sea Research I*, 53:917–925.
- ANTOINE, D., SIEGEL, D. A., KOSTADINOV, T., MARRITORENA, S., NELSON, N. B., GENTILI, B., VELLUCCI, V. et GUILLOCHEAU, N. (2011). Variability in optical particle backscattering in contrasting bio-optical oceanic regimes. *Limnology and Oceanography*, 56(3):955–973.
- ARMBRUST, E. V. (2009). The life of diatoms in the world’s oceans. *Nature*, 459(7244):185–192.
- BABIN, M., STRAMSKI, D., FERRARI, G. M., CLAUSTRE, H., BRICAUD, A., OBOLENSKY, G. et HOEPFFNER, N. (2003). Variations in the light absorption coefficients of phytoplankton, nonalgal particles, and dissolved organic matter in coastal waters around europe. *Journal of Geophysical Research : Oceans*, 108(C7).
- 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(1):12–23.
- BALCH, W. M., GORDON, H. R., BOWLER, B. C., DRAPEAU, D. T. et BOOTH, E. S. (2005). Calcium carbonate measurements in the surface global ocean based on moderate-resolution imaging spectroradiometer data. *Journal of Geophysical Research : Oceans*, 110(C7).
- BATHMANN, U., SCHAREK, R., KLAAS, C., DUBISCHAR, C. et SMETACEK, V. (1997). Spring development of phytoplankton biomass and composition in major water masses of the atlantic sector of the southern ocean. *Deep Sea Research Part II : Topical Studies in Oceanography*, 44(1):51–67.
- BATTEN, S., CLARK, R., FLINKMAN, J., HAYS, G., JOHN, E., JOHN, A., JONAS, T., LINDLEY, J., STEVENS, D. et WALNE, A. (2003). Cpr sampling : the technical background, materials and methods, consistency and comparability. *Progress in Oceanography*, 58(2):193–215.
- BEAUFORT, L., PROBERT, I., de GARIDEL-THORON, T., BENDIF, E. M., RUIZ-PINO, D., METZL, N., GOYET, C., BUCHET, N., COUPEL, P., GRELAUD, M. *et al.* (2011). Sensitivity of coccolithophores to carbonate chemistry and ocean acidification. *Nature*, 476(7358):80–83.
- BEAUGRAND, G. (2004). Continuous plankton records : Plankton atlas of the north atlantic ocean (1958-1999). i. introduction and methodology. *Marine Ecology Progress Series*, 268:3–10.
- BEAUGRAND, G. (2015). *Marine biodiversity, climatic variability and global change*. Routledge.
- BEAUGRAND, G., IBAÑEZ, F. et LINDLEY J., A. (2001). Geographical distribution and seasonal and diel changes of the diversity of calanoid copepods in the north atlantic and north sea. *Marine Ecology Progress Series*,



219:189–203.

- BEAUGRAND, G., LENOIR, S., IBANEZ, F., MANTÉ, C. *et al.* (2011). A new model to assess the probability of occurrence of a species based on presence-only data. *Mar. Ecol. Prog. Ser.*, 424:175–190.
- BEHRENFELD, M. *et al.* FALKOWSKI, P. (1998). A consumer's guide to phytoplankton primary productivity models. *Oceanographic Literature Review*, 9(45):1576.
- BELVISO, S., MORROW, R. *et al.* MIHALOPOULOS, N. (2000). An atlantic meridional transect of surface water dimethyl sulfide concentrations with 10–15 km horizontal resolution and close examination of ocean circulation. *Journal of Geophysical Research : Atmospheres*, 105(D11):14423–14431.
- BEN MUSTAPHA, Z. (2013). *Téledétection des groupes phytoplanctoniques via l'utilisation conjointe de mesures satellites, in situ et d'une méthode de classification automatique*. Thèse de doctorat, Université du Littoral Côte d'Opale.
- BEN MUSTAPHA, Z., ALVAIN, S., JAMET, C., LOISEL, H. *et al.* DESSAILLY, D. (2014). 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*, 146:97–112.
- BONATO, S., CHRISTAKI, U., LEFEBVRE, A., LIZON, F., THYSSEN, M. *et al.* ARTIGAS, L. F. (2015). High spatial variability of phytoplankton assessed by flow cytometry, in a dynamic productive coastal area, in spring : The eastern english channel. *Estuarine, Coastal and Shelf Science*, 154:214–223.
- BOPP, L., AUMONT, O., CADULE, P., ALVAIN, S. *et al.* GEHLEN, M. (2005). Response of diatoms distribution to global warming and potential implications : A global model study. *Geophysical Research Letters*, 32(19).
- BOSS, E., SLADE, W. *et al.* HILL, P. (2009). Effect of particulate aggregation in aquatic environments on the beam attenuation and its utility as a proxy for particulate mass. *Optics Express*, 17(11):9408–9420.
- BOYCE, D. G., LEWIS, M. R. *et al.* WORM, B. (2010). Global phytoplankton decline over the past century. *Nature*, 466(7306):591.
- BOYD, P., LAROCHE, J., GALL, M., FREW, R. *et al.* MCKAY, R. M. L. (1999). Role of iron, light, and silicate in controlling algal biomass in subantarctic waters se of new zealand. *Journal of Geophysical Research : Oceans*, 104(C6):13395–13408.
- BRACHER, A., TAYLOR, M., TAYLOR, B., DINTER, T., ROETTIGERS, R. *et al.* STEINMETZ, F. (2015). Using empirical orthogonal functions derived from remote sensing reflectance for the prediction of phytoplankton pigments concentrations. *Ocean Science*, 11(1):139–158.
- BRACHER, A., VOUNTAS, M., DINTER, T., BURROWS, J., RÖTTIGERS, R. *et al.* PEEKEN, I. (2009). Quantitative observation of cyanobacteria and diatoms from space using PhytoDOAS on SCIAMACHY data. *Biogeosciences*, 6:751–764.
- BRAY, J. R. *et al.* CURTIS, J. T. (1957). An ordination of the upland forest communities of southern wisconsin. *Ecological monographs*, 27(4):325–349.
- BRESNAN, E., HAY, S., HUGHES, S., FRASER, S., RASMUSSEN, J., WEBSTER, L., SLESSER, G., DUNN, J. *et al.* HEATH, M. (2009). Seasonal and interannual variation in the phytoplankton community in the north east of scotland. *Journal of Sea Research*, 61(1):17–25.
- BREWIN, R. J., SATHYENDRANATH, S., HIRATA, T., LAVENDER, S. J., BARCIELA, R. M. *et al.* HARDMAN-MOUNTFORD, N. J. (2010). A three-component model of phytoplankton size class for the atlantic ocean. *Ecological Modelling*, 221(11):1472–1483.
- BREWIN, R. J., SATHYENDRANATH, S., JACKSON, T., BARLOW, R., BROTHAS, V., AIRS, R. *et al.* LAMONT, T. (2015). Influence of light in the mixed-layer on the parameters of a three-component model of phytoplankton size class. *Remote Sensing of Environment*,

- 168:437–450.
- BRICAUD, A., BABIN, M., MOREL, A. et CLAUSTRE, H. (1995). Variability in the chlorophyll-specific absorption coefficients of natural phytoplankton : Analysis and parameterization. *J. Geophys. Res.*, 100:13–321.
- 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(5):851–873.
- BRICAUD, A., CIOTTI, A. 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(1).
- BRICAUD, A. et MOREL, A. (1986). Light attenuation and scattering by phytoplanktonic cells : a theoretical modeling. *Applied Optics*, 25(4):571–580.
- 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 : Oceans*, 103(C13):31033–31044.
- BRICAUD, A., MOREL, A., PRIEUR, L. *et al.* (1981). Absorption by dissolved organic matter of the sea (yellow substance) in the uv and visible domains. *Limnol. Oceanogr.*, 26(1):43–53.
- BROWN, C. A., HUOT, Y., WERDELL, P. J., GENTILI, B. et CLAUSTRE, H. (2008). The origin and global distribution of second order variability in satellite ocean color and its potential applications to algorithm development. *Remote Sensing of Environment*, 112(12):4186–4203.
- BROWN, C. W. et YODER, J. A. (1994). Coccolithophorid blooms in the global ocean. *Journal of Geophysical Research : Oceans*, 99(C4):7467–7482.
- BROWNLEE, C. et TAYLOR, A. R. (2002). Algal calcification and silification. *eLS*.
- BUESSELER, K. O. (1998). The decoupling of production and particulate export in the surface ocean. *Global Biogeochemical Cycles*, 12(2):297–310.
- CAPONE, D. G., BURNS, J. A., MONTOYA, J. P., SUBRAMANIAM, A., MAHAFFEY, C., GUNDERSON, T., MICHAELS, A. F. et CARPENTER, E. J. (2005). Nitrogen fixation by *Trichodesmium spp.* : An important source of new nitrogen to the tropical and subtropical north atlantic ocean. *Global Biogeochemical Cycles*, 19(2).
- CARPENTER, G., GILLISON, A. et WINTER, J. (1993). DOMAIN : a flexible modelling procedure for mapping potential distributions of plants and animals. *Biodiversity & Conservation*, 2(6):667–680.
- CASEY, K. S., BRANDON, T. B., CORNILLON, P. et EVANS, R. (2010). The past, present, and future of the AVHRR Pathfinder SST program. *In Oceanography from space*, pages 273–287. Springer.
- CASTELLANI, C. et EDWARDS, M. (2017). *Marine Plankton : A Practical Guide to Ecology, Methodology, and Taxonomy*. Oxford University Press.
- CAVENDER-BARES, K. K., MANN, E. L., CHISHOLM, S. W., ONDRUSEK, M. E. et BIDIGARE, R. R. (1999). Differential response of equatorial pacific phytoplankton to iron fertilization. *Limnology and Oceanography*, 44(2):237–246.
- CHAZOTTES, A., BRICAUD, A., 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(31):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(18):3790–3799.
- CHEN, L., MI, X., COMITA, L. S., ZHANG, L., REN, H. et MA, K. (2010). Community-level consequences of density dependence and habitat association in a subtropical broad-leaved forest. *Ecology Letters*, 13(6):695–704.
- CIOTTI, A. M. et BRICAUD, A. (2006). Retrievals of a size parameter for phytoplankton and spectral light absorption by colored detrital matter from water-leaving radiances at SeaWiFS channels in a continental shelf region off Brazil. *Limnology and Oceanography : Methods*, 4:237–253.
- CIOTTI, A. M., LEWIS, M. R. et CULLEN, J. J. (2002). Assessment of the relationships between dominant cell size in natural phytoplankton communities and the spectral shape of the absorption coefficient. *Limnology and Oceanography*, 47(2):404–417.
- CLAUSTRE, H., HOOKER, S. B., VAN HEUKELEM, L., BERTHON, J.-F., BARLOW, R., RAS, J., SESSIONS, H., TARGA, C., THOMAS, C. S., van der LINDE, D. et al. (2004). An intercomparison of hplc phytoplankton pigment methods using in situ samples : application to remote sensing and database activities. *Marine Chemistry*, 85(1):41–61.
- CLAVANO, W., BOSS, E. et KARP-BOSS, L. (2007). Inherent optical properties of non-spherical marine-like particles from theory to observation. *Oceanography and marine biology : an annual review*, 45:1–38.
- COALE, K. H., JOHNSON, K. S., FITZWATER, S. E., GORDON, R. M. et al. (1996). A massive phytoplankton bloom induced by an ecosystem-scale iron fertilization experiment in the equatorial pacific ocean. *Nature*, 383(6600).
- COLEBROOK, J. (1982). Continuous plankton records-phytoplankton, zooplankton and environment, northeast atlantic and north-sea, 1958-1980. *Oceanologica Acta*, 5(4):473–480.
- COMMISSION, O. (2010). Bilan de santé 2010. *Commission OSPAR, London*, page 176.
- CONNOR, S. E., ROSS, S. A., SOBOTKOVA, A., HERIES, A. I., MOONEY, S. D., LONGFORD, C. et ILIEV, I. (2013). Environmental conditions in the se balkans since the last glacial maximum and their influence on the spread of agriculture into europe. *Quaternary Science Reviews*, 68:200–215.
- COTTRELL, M., LETRÉMY, P., ROUSSET, P. et IBOU, S. (2003). Cartes auto-organisées pour l’analyse exploratoire de données et la visualisation. *Journal de la société française de statistique*, 144:67–106.
- COUTO, B., BROTA, V., MÉLIN, F., GROOM, S. et SATHYENDRANATH, S. (2016). Inter-comparison of oc-cci chlorophyll-a estimates with precursor data sets. *International Journal of Remote Sensing*, 37(18):4337–4355.
- CRAIG, S. E., JONES, C. T., LI, W. K., LAZIN, G., HORNE, E., CAVERHILL, C. et CULLEN, J. J. (2012). Deriving optical metrics of coastal phytoplankton biomass from ocean colour. *Remote Sensing of Environment*, 119:72–83.
- CROOT, P. L., ANDERSSON, K., ÖZTÜRK, M. et TURNER, D. R. (2004). The distribution and speciation of iron along 6 e in the southern ocean. *Deep Sea Research Part II : Topical Studies in Oceanography*, 51(22):2857–2879.
- DAAN, N., BROMLEY, P., HISLOP, J. et NIELSEN, N. (1990). Ecology of north sea fish. *Netherlands Journal of Sea Research*, 26(2):343–386.
- 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 Part II : Topical Studies in Oceanography*, 51(1):303–318.
- DANDONNEAU, Y. et NIANG, A. (2007). Assemblages of phytoplankton pigments along a shipping line through the north atlantic and tropical pacific. *Progress in Oceanography*, 73(2):127–144.
- DARNIS, G., BARBER, D. G. et FORTIER, L. (2008). Sea ice and the onshore–offshore gradient in

- pre-winter zooplankton assemblages in southeastern beaufort sea. *Journal of Marine Systems*, 74(3):994–1011.
- DELBART, N., KERGOAT, L., LE TOAN, T., LHERMITTE, J. et PICARD, G. (2005). Determination of phenological dates in boreal regions using normalized difference water index. *Remote Sensing of Environment*, 97(1):26–38.
- DENMAN, K. L. (2008). Climate change, ocean processes and ocean iron fertilization. *Marine Ecology Progress Series*, 364:219–225.
- DESSAILLY, D., LOISEL, H., STEINMETZ, F., VANTREPOTTE, V., SATRA LE BRIS, C., RAMON, D. et MEILLON, J. (2016). Globcoast, a new dataset for coastal water ocr. In *Ocean Optics XXIII, October 23–28, 2016, Victoria, BC, Canada*.
- D’OVIDIO, F., DE MONTE, S., ALVAIN, S., DANDONNEAU, Y. et LÉVY, M. (2010). Fluid dynamical niches of phytoplankton types. *Proceedings of the National Academy of Sciences*, 107(43):18366–18370.
- DUBELAAR, G. B., VISSER, J. W. et DONZE, M. (1987). Anomalous behaviour of forward and perpendicular light scattering of a cyanobacterium owing to intracellular gas vacuoles. *Cytometry*, 8(4):405–412.
- DUCKLOW, H. W. et HARRIS, R. P. (1993). Introduction to the JGOFS North Atlantic bloom experiment. *Deep Sea Research Part II : Topical Studies in Oceanography*, 40(1-2):1–8.
- DUCROTOY, J.-P., ELLIOTT, M. et de JONGE, V. N. (2000). The north sea. *Marine Pollution Bulletin*, 41(1):5–23.
- DUFRENE, M. et LEGENDRE, P. (1997). Species assemblages and indicator species : the need for a flexible asymmetrical approach. *Ecological monographs*, 67(3):345–366.
- EDGINGTON, E. et ONGHENA, P. (2007). *Randomization tests*. CRC Press.
- EDWARDS, M., REID, P. et PLANQUE, B. (2001). Long-term and regional variability of phytoplankton biomass in the northeast atlantic (1960–1995). *ICES Journal of Marine Science : Journal du Conseil*, 58(1):39–49.
- EDWARDS, M. et RICHARDSON, A. J. (2004). Impact of climate change on marine pelagic phenology and trophic mismatch. *Nature*, 430(7002):881–884.
- EUROPEAN SPACE AGENCY (2017). Ocean colour climate change initiative dataset. available online at <http://www.esa-oceancolour-cci.org/>.
- FALKOWSKI, P. G. (1994). The role of phytoplankton photosynthesis in global biogeochemical cycles. *Photosynthesis research*, 39(3):235–258.
- FALKOWSKI, P. G. (1997). Evolution of the nitrogen cycle and its influence on the biological sequestration of co2 in the ocean. *Nature*, 387(6630):272–275.
- FIELD, C. B., BEHRENFELD, M. J., RANDERSON, J. T. et FALKOWSKI, P. (1998). Primary production of the biosphere : integrating terrestrial and oceanic components. *Science*, 281(5374):237–240.
- FRANZ, H., COLEBROOK, J., GAMBLE, J. et KRAUSE, M. (1991). The zooplankton of the north sea. *Netherlands Journal of Sea Research*, 28(1):1–52.
- FRANZ, B. (2009). Methods for assessing the quality and consistency of ocean color products. *NASA Goddard Space Flight Center, Ocean Biology Processing Group* [http://oceancolor.gsfc.nasa.gov/DOCS/methods/sensor\\_analysis\\_methods.html](http://oceancolor.gsfc.nasa.gov/DOCS/methods/sensor_analysis_methods.html).
- FRANZ, B. A. et WERDELL, P. J. (2010). A generalized framework for modeling of inherent optical properties in ocean remote sensing applications. *Proceedings of Ocean Optics, Anchorage, Alaska*, 27.
- FREDERIKSEN, M., EDWARDS, M., RICHARDSON, A. J., HALLIDAY, N. C. 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(6):1259–1268.

- FROUIN, R., MCPHERSON, J., UEYOSHI, K. et FRANZ, B. A. (2012). A time series of photosynthetically available radiation at the ocean surface from SeaWiFS and modis data. *Proc.SPIE*, 8525:8525 – 8525 – 12.
- FUJIWARA, A., HIRAWAKE, T., SUZUKI, K. et SAITOH, S.-I. (2011). Remote sensing of size structure of phytoplankton communities using optical properties of the chukchi and bering sea shelf region. *Biogeosciences*, 8(12):3567.
- GALL, M., BOYD, P., HALL, J., SAFI, K. et CHANG, H. (2001). Phytoplankton processes. part 1 : community structure during the southern ocean iron release experiment (soiree). *Deep Sea Research Part II : Topical Studies in Oceanography*, 48(11):2551–2570.
- GEHLEN, M., BOPP, L., EMPRIN, N., AUMONT, O., HEINZE, C. et RAGUENEAU, O. (2006). Reconciling surface ocean productivity, export fluxes and sediment composition in a global biogeochemical ocean model. *Biogeosciences*, 3(4):521–537.
- GIESKES, W. et KRAAY, G. (1983). Dominance of cryptophyceae during the phytoplankton spring bloom in the central north sea detected by hplc analysis of pigments. *Marine Biology*, 75(2-3):179–185.
- GOFFART, A., CATALANO, G. et HECQ, J.-H. (2000). Factors controlling the distribution of diatoms and phaeocystis in the ross sea. *Journal of Marine Systems*, 27(1):161–175.
- GORDON, H. R., BOYNTON, G. C., BALCH, W. M., GROOM, S. B., HARBOUR, D. S. et SMYTH, T. J. (2001). Retrieval of coccolithophore calcite concentration from seawifs imagery. *Geophysical Research Letters*, 28(8):1587–1590.
- GORDON, H. R. et MOREL, A. Y. (2012). *Remote assessment of ocean color for interpretation of satellite visible imagery : A review*, volume 4. Springer Science & Business Media.
- 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(3): 443–452.
- GOWER, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, pages 857–871.
- GREGG, W. W. et CASEY, N. W. (2007). Modeling coccolithophores in the global oceans. *Deep Sea Research Part II : Topical Studies in Oceanography*, 54(5):447–477.
- GROUP, N. O. B. P. (2015). Czcs level 3 binned chlorophyll data version 2014.
- GUISAN, A. et THUILLER, W. (2005). Predicting species distribution : offering more than simple habitat models. *Ecology letters*, 8(9):993–1009.
- GUISAN, A. et ZIMMERMANN, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological modelling*, 135(2):147–186.
- HAIRSTON, J. N. G. et HAIRSTON, S. N. G. (1993). Cause-effect relationships in energy flow, trophic structure, and interspecific interactions. *The American Naturalist*, 142(3):379–411.
- HASTIE, T. J. et TIBSHIRANI, R. J. (1990). *Generalized additive models*, volume 43. CRC Press.
- HELBLING, E. W., VILLAFANE, V. et HOLM-HANSEN, O. (1991). Effect of iron on productivity and size distribution of antarctic phytoplankton. *Limnology and Oceanography*, 36(8): 1879–1885.
- HINDER, S. L., HAYS, G. C., EDWARDS, M., ROBERTS, E. C., WALNE, A. W. et GRAVENOR, M. B. (2012). Changes in marine dinoflagellate and diatom abundance under climate change. *Nature Climate Change*, 2(4):271–275.
- HIRATA, T., HARDMAN-MOUNTFORD, N. J., BREWIN, R. J. W., AIKEN, J., BARLOW, R., SUZUKI, K., ISADA, T., HOWELL, E., HASHIOKA, T., NOGUCHI-AITA, M. et OTHERS (2011). Synoptic relationships between surface Chlorophyll-a and diagnostic pigments specific to phytoplankton functional types. *Biogeosciences*, 8(2):311–327.
- HIRZEL, A. H., HAUSSER, J., CHESSEL, D. et PER-

- RIN, N. (2002). Ecological-niche factor analysis : how to compute habitat-suitability maps without absence data ? *Ecology*, 83(7):2027–2036.
- HOEPPFNER, N. et SATHYENDRANATH, S. (1991). Effect of pigment composition on absorption properties of phytoplankton. *Marine Ecology Progress Series*, pages 11–23.
- HOOD, R. R., LAWS, E. A., ARMSTRONG, R. A., BATES, N. R., BROWN, C. W., CARLSON, C. A., CHAI, F., DONEY, S. C., FALKOWSKI, P. G., FEELY, R. A. *et al.* (2006). Pelagic functional group modeling : Progress, challenges and prospects. *Deep Sea Research Part II : Topical Studies in Oceanography*, 53(5):459–512.
- HOPPE, C., LANGER, G. et ROST, B. (2011). *Emiliania huxleyi* shows identical responses to elevated pco<sub>2</sub> in {TA} and {DIC} manipulations. *Journal of Experimental Marine Biology and Ecology*, 406:54 – 62.
- HOPPENRATH, M. (2004). A revised checklist of planktonic diatoms and dinoflagellates from helgoland (north sea, german bight). *Helgoland Marine Research*, 58(4):243–251.
- HOPPENRATH, M., BESZTERI, B., DREBES, G., HALLIGER, H., VAN BEUSEKOM, J. E., JANISCH, S. et WILTSHIRE, K. H. (2007). Thalassiosira species (bacillariophyceae, thalassiosirales) in the north sea at helgoland (german bight) and sylt (north frisian wadden sea)—a first approach to assessing diversity. *European Journal of Phycology*, 42(3):271–288.
- HU, C., LEE, Z. et FRANZ, B. (2012). Chlorophyll a algorithms for oligotrophic oceans : A novel approach based on three-band reflectance difference. *Journal of Geophysical Research : Oceans*, 117(C1).
- IOCCG report (2000). Remote sensing of ocean colour in coastal, and other optically-complex, waters. *IOCCG report*, 3.
- IOCCG report (2014). Phytoplankton functional types from space. (*Reports of the International Ocean-Colour Coordinating Group (IOCCG); 15*), pages 1–156. Authored by Sathyendranath, S and Aiken, J and Alvain, S and Barlow, R and Bouman, H and Bracher, A and Brewin, R and Bricaud, A and Brown, CW and Ciotti, AM and others.
- JAIN, A. K. et DUBES, R. C. (1988). *Algorithms for Clustering Data*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.
- JEFFREY, S., MANTOURA, R. et WRIGHT, S. (1997). Phytoplankton pigments in oceanography : guidelines to modern methods. *Monographs on oceanographic methodology*.
- JERLOV, N. G. (1976). *Marine optics*, volume 14. Elsevier.
- JI, R., EDWARDS, M., MACKAS, D. L., RUNGE, J. A. et THOMAS, A. C. (2010). Marine plankton phenology and life history in a changing climate : current research and future directions. *Journal of plankton research*, 32(10):1355–1368.
- JOHNSEN, G., SAMSET, O., GRANSKOG, L. et SAKSHAUG, E. (1994). In vivo absorption characteristics in 10 classes of bloom-forming phytoplankton : taxonomic characteristics and responses to photoadaptation by means of discriminant and hplc analysis. *Marine Ecology Progress Series*, pages 149–157.
- KARL, D. M., CHURCH, M. J., DORE, J. E., LETELIER, R. M. et MAHAFFEY, C. (2012). Predictable and efficient carbon sequestration in the north pacific ocean supported by symbiotic nitrogen fixation. *Proceedings of the National Academy of Sciences*, 109(6):1842–1849.
- KIRK, J. T. (1994). *Light and photosynthesis in aquatic ecosystems*. Cambridge university press.
- KOHONEN, T. (2013). Essentials of the self-organizing map. *Neural Networks*, 37:52–65.
- KOSTADINOV, T., SIEGEL, D. et MARITORENA, S. (2009). Retrieval of the particle size distribution from satellite ocean color observations. *Journal of Geophysical Research : Oceans (1978–2012)*, 114(9):1–22.
- KOSTADINOV, T., SIEGEL, D. et MARITORENA, S.

- (2010). Global variability of phytoplankton functional types from space : assessment via the particle size distribution. *Biogeosciences*, 7(10):3239–3257.
- KOSTADINOV, T. S., CABRÉ, A., VEDANTHAM, H., MARINOV, I., BRACHER, A., BREWIN, R. J., BRICAUD, A., HIRATA, T., HIRAWAKE, T., HARDMAN-MOUNTFORD, N. J. *et al.* (2017). Inter-comparison of phytoplankton functional type phenology metrics derived from ocean color algorithms and earth system models. *Remote Sensing of Environment*, 190:162–177.
- LANCELOT, C., WASSMANN, P. et BARTH, H. (1994). Ecology of phaeocystis-dominated ecosystems. *Journal of marine systems*, 5(1):1–4.
- LAUBSCHER, R., PERISSINOTTO, R. et MCQUAID, C. (1993). Phytoplankton production and biomass at frontal zones in the atlantic sector of the southern ocean. *Polar Biology*, 13(7):471–481.
- LE CLAINCHE, Y., VÉZINA, A., LEVASSEUR, M., CROPP, R. A., GUNSON, J. R., VALLINA, S. M., VOGT, M., LANCELOT, C., ALLEN, J., ARCHER, S. D. *et al.* (2010). A first appraisal of prognostic ocean dms models and prospects for their use in climate models. *Global biogeochemical cycles*, 24(3).
- LE QUÉRÉ, C., HARRISON, S. P., COLIN PRENTICE, I., BUITENHUIS, E. T., AUMONT, O., BOPP, L., CLAUSTRE, H., COTRIM DA CUNHA, L., GEIDER, R., GIRAUD, X. *et al.* (2005). Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. *Global Change Biology*, 11(11):2016–2040.
- LEGENDRE, P. et LEGENDRE, L. F. (2012). *Numerical ecology*, volume 24. Elsevier.
- LETERME, S. C., SEURONT, L. et EDWARDS, M. (2006). Differential contribution of diatoms and dinoflagellates to phytoplankton biomass in the ne atlantic ocean and the north sea. *Marine Ecology Progress Series*, 312:57–65.
- LI, Z., LI, L., SONG, K. et CASSAR, N. (2013). Estimation of phytoplankton size fractions based on spectral features of remote sensing ocean color data. *Journal of Geophysical Research : Oceans*, 118(3):1445–1458.
- LINDLEY, J. et WILLIAMS, R. (1980). Plankton of the fladen ground during flex 76 ii. population dynamics and production of thysanoessa inermis (crustacea : Euphausiacea). *Marine Biology*, 57(2):79–86.
- LOHBECK, K. T., RIEBESELL, U. et REUSCH, T. B. (2012). Adaptive evolution of a key phytoplankton species to ocean acidification. *Nature Geoscience*, 5(5):346–351.
- LOISEL, H., LUBAC, B., DESSAILLY, D., DUFORET-GAURIER, L. et VANTREPOTTE, V. (2010). Effect of inherent optical properties variability on the chlorophyll retrieval from ocean color remote sensing : an in situ approach. *Optics Express*, 18(20):20949–20959.
- LOISEL, H. et MOREL, A. (1998). Light scattering and chlorophyll concentration in case 1 waters : A reexamination. *Limnology and Oceanography*, 43(5):847–858.
- LOISEL, H., NICOLAS, J.-M., SCIANDRA, A., STRAMSKI, D. et POTEAU, A. (2006). Spectral dependency of optical backscattering by marine particles from satellite remote sensing of the global ocean. *Journal of Geophysical Research : Oceans*, 111(C9).
- LOISEL, H., VANTREPOTTE, V., DESSAILLY, D., STEINMETZ, F. et HAN, B. (2016). Remote sensing of suspended particulate matter variability of the global coastal waters over the last decade. In *Environmental Science and Sustainable Development : International Conference on Environmental Science and Sustainable Development (ICESSD 2015)*, pages 118–123. World Scientific.
- LONGHURST, A. R. (2010). *Ecological geography of the sea*. Academic Press.
- MALIN, G., TURNER, S. M. et LISS, P. S. (1992). Sulfur : the plankton/climate connection. *Journal of Phycology*, 28(5):590–597.
- MARAÑÓN, E. (2015). Cell size as a key determinant of phytoplankton metabolism and com-

- community structure. *Annual review in Marine Science*.
- MARGALEF, R. (1978). Life-forms of phytoplankton as survival alternatives in an unstable environment. *Oceanologica acta*, 1(4):493–509.
- MARGALEF, R. (1997). Turbulence and marine life. *Scientia marina*, 61:109–123.
- MARTIN, J. H. et FITZWATER, S. E. (1988). Iron deficiency limits phytoplankton growth in the north-east pacific subarctic. *Nature*, 331(6154):341–343.
- MASQUELIER, S., FOULON, E., JOUENNE, F., FERRÉOL, M., BRUSSAARD, C. P. et VAULOT, D. (2011). Distribution of eukaryotic plankton in the english channel and the north sea in summer. *Journal of Sea Research*, 66:112–122.
- MCCULLAGH, P. et NELDER, J. A. (1989). *Generalized linear models*, volume 37. CRC press.
- MEADOW, J. F., ALTRICHTER, A. E., KEMBEL, S. W., MORIYAMA, M., O’CONNOR, T. K., WOMACK, A. M., BROWN, G., GREEN, J. L. et BOHANNAN, B. J. (2014). Bacterial communities on classroom surfaces vary with human contact. *Microbiome*, 2(7):1–7.
- MOBLEY, C. D. (1994). *Light and water : radiative transfer in natural waters*. Academic press.
- MOORE, J. K., DONEY, S. C., GLOVER, D. M. et FUNG, I. Y. (2001). Iron cycling and nutrient-limitation patterns in surface waters of the world ocean. *Deep Sea Research Part II : Topical Studies in Oceanography*, 49(1):463–507.
- MOREL, A. (1988). Optical modeling of the upper ocean in relation to its biogenous matter content(case i waters). *Journal of Geophysical Research*, 93(10):749–10.
- MOREL, A. (2009). Are the empirical relationships describing the bio-optical properties of case 1 waters consistent and internally compatible? *Journal of Geophysical Research : Oceans*, 114(C1).
- MOREL, A., AHN, Y.-H., PARTENSKY, F., VAULOT, D. et CLAUSTRE, H. (1993). Prochlorococcus and Synechococcus : A comparative study of their optical properties in relation to their size and pigmentation. *Journal of Marine Research*, 51(3):617–649.
- MOREL, A., ANTOINE, D. et GENTILI, B. (2002). Bidirectional reflectance of oceanic waters : accounting for raman emission and varying particle scattering phase function. *Applied Optics*, 41(30):6289–6306.
- MOREL, A. et GENTILI, B. (1991). Diffuse reflectance of oceanic waters : its dependence on sun angle as influenced by the molecular scattering contribution. *Applied optics*, 30(30):4427–4438.
- MOREL, A. et MARITORENA, S. (2001). Bio-optical properties of oceanic waters : A reappraisal. *Journal of Geophysical Research : Oceans*, 106(C4):7163–7180.
- MOREL, A. et PRIEUR, L. (1977). Analysis of variations in ocean color. *Limnology and oceanography*, 22(4):709–722.
- MOUTIER, W., DUFORÊT-GAURIER, L., THYSSEN, M., LOISEL, H., MÉRIAUX, X., COURCOT, L., DESSAILLY, D., RÊVE, A.-H., GRÉGORI, G., ALVAIN, S. *et al.* (2017). Evolution of the scattering properties of phytoplankton cells from flow cytometry measurements. *PloS one*, 12(7):e0181180.
- MOUW, C. B., HARDMAN-MOUNTFORD, N. J., ALVAIN, S., BRACHER, A., BREWIN, R. J., BRICAUD, A., CIOTTI, A. M., DEVRED, E., FUJIWARA, A., HIRATA, T. *et al.* (2017). A Consumer’s Guide to Satellite Remote Sensing of Multiple Phytoplankton Groups in the Global Ocean. *Frontiers in Marine Science*, 4(41).
- MOUW, C. B. et YODER, J. A. (2010). Optical determination of phytoplankton size composition from global seawifs imagery. *Journal of Geophysical Research : Oceans*, 115(C12).
- NAHMANI, J., LAVELLE, P. et ROSSI, J.-P. (2006). Does changing the taxonomical resolution alter the value of soil macroinvertebrates as bioindicators of metal pollution? *Soil Bio-*



- logy and Biochemistry*, 38(2):385–396.
- NAIR, A., SATHYENDRANATH, S., PLATT, T., MORALES, J., STUART, V., FORGET, M.-H., DEVRED, E. et BOUMAN, H. (2008). Remote sensing of phytoplankton functional types. *Remote Sensing of Environment*, 112(8):3366–3375.
- NASA GODDARD SPACE FLIGHT CENTER, O. E. L. (2014). Seawifs ocean color data.
- NAVARRO, G., ALMARAZ, P., CABALLERO, I., VAZQUEZ, A. et HUERTAS, I. E. (2017). Reproduction of spatio-temporal patterns of major mediterranean phytoplankton groups from remote sensing oc-cci data. *Frontiers in Marine Science*, 4:246.
- NAVARRO, G., ALVAIN, S., VANTREPOTTE, V. et HUERTAS, I. E. (2014). Identification of dominant phytoplankton functional types in the mediterranean sea based on a regionalized remote sensing approach. *Remote Sensing of Environment*, 152:557–575.
- NELSON, D. M. et SMITH, W. (1991). Sverdrup revisited : Critical depths, maximum chlorophyll levels, and the control of southern ocean productivity by the irradiance-mixing regime. *Limnology and Oceanography*, 36(8):1650–1661.
- 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(2):257 – 271.
- O'REILLY, J. E., MARITORENA, S., MITCHELL, B. G., SIEGEL, D. A., CARDER, K. L., GARVER, S. A., KAHRU, M. et MCCLAIN, C. (1998). Ocean color chlorophyll algorithms for seawifs. *Journal of Geophysical Research : Oceans (1978–2012)*, 103(C11):24937–24953.
- O'REILLY, J. E., MARITORENA, S., O'BRIEN, M., SIEGEL, D., TOOLE, D., MENZIES, D., SMITH, R., MUELLER, J., MITCHELL, B. G., KAHRU, M. et al. (2000). SeaWiFS postlaunch calibration and validation analyses, part 3. *NASA tech. memo*, 206892(11):3–8.
- ORGANELLI, E., NUCCIO, C., LAZZARA, L., UITZ, J., BRICAUD, A. et MASSI, L. (2017). On the discrimination of multiple phytoplankton groups from light absorption spectra of assemblages with mixed taxonomic composition and variable light conditions. *Appl. Opt.*, 56(14):3952–3968.
- OTTO, L., ZIMMERMAN, J., FURNES, G., MORK, M., SAETRE, R. et BECKER, G. (1990). Review of the physical oceanography of the north sea. *Netherlands Journal of Sea Research*, 26(2):161–238.
- PARTENSKY, F., BLANCHOT, J. et VAULOT, D. (1999). Differential distribution and ecology of prochlorococcus and synechococcus in oceanic waters : a review. *Bulletin institut océanographique de Monaco*, pages 457–476.
- PASSOW, U. et CARLSON, C. A. (2012). The biological pump in a high co2 world. *Marine Ecology Progress Series*, 470:249–271.
- PELOQUIN, J., SWAN, C., GRUBER, N., VOGT, M., CLAUSTRE, H., RAS, J., UITZ, J., BARLOW, R., BEHRENFELD, M., BIDIGARE, R. et al. (2013). The maredat global database of high performance liquid chromatography marine pigment measurements. *Earth System Science Data*, 5(1):109.
- PHILLIPS, S. J., ANDERSON, R. P. et SCHAPIRE, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological modelling*, 190(3):231–259.
- PLATT, T., FUENTES-YACO, C. et FRANK, K. T. (2003). Marine ecology : spring algal bloom and larval fish survival. *Nature*, 423(6938):398–399.
- PLATT, T. et JASSBY, A. D. (1976). The relationship between photosynthesis and light for natural assemblages of coastal marine phytoplankton. *Journal of Phycology*, 12(4):421–430.
- PLATT, T. et SATHYENDRANATH, S. (2008). Ecological indicators for the pelagic zone of the ocean from remote sensing. *Remote Sensing of Environment*, 112(8):3426–3436.

- PLATT, T., WHITE, G. N., ZHAI, L., SATHYENDRANATH, S. et ROY, S. (2009). The phenology of phytoplankton blooms : Ecosystem indicators from remote sensing. *Ecological Modelling*, 220(21):3057–3069.
- POLOVINA, J. J., HOWELL, E. A., KOBAYASHI, D. R. et SEKI, M. P. (2017). The transition zone chlorophyll front updated : Advances from a decade of research. *Progress in Oceanography*, 150:79 – 85. Advances in understanding of the North Pacific subtropical front and transition zone ecosystems.
- PRIEUR, L. et SATHYENDRANATH, S. (1981). An optical classification of coastal and oceanic waters based on the specific spectral absorption curves of phytoplankton pigments, dissolved organic matter, and other particulate materials. *Limnology and Oceanography*, 26(4):671–689.
- R CORE TEAM (2015). *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- RACAULT, M.-F., LE QUÉRÉ, C., BUITENHUIS, E., SATHYENDRANATH, S. et PLATT, T. (2012). Phytoplankton phenology in the global ocean. *Ecological Indicators*, 14(1):152–163.
- RAITSOS, D. E., LAVENDER, S. J., MARAVELIAS, C. D., HARALABOUS, J., RICHARDSON, A. J. et REID, P. C. (2008). Identifying four phytoplankton functional types from space : An ecological approach. *Limnology and oceanography*, 53(2):605–613.
- REID, P., COLEBROOK, J., MATTHEWS, J., AIKEN, J. et TEAM, C. P. R. (2003). The continuous plankton recorder : concepts and history, from plankton indicator to undulating recorders. *Progress in Oceanography*, 58(2): 117–173.
- REID, P., LANCELOT, C., GIESKES, W., HAGMEIER, E. et WEICHART, G. (1990). Phytoplankton of the north sea and its dynamics : a review. *Netherlands Journal of Sea Research*, 26(2): 295–331.
- REYDONDEAU, G., LONGHURST, A., MARTINEZ, E., BEAUGRAND, G., ANTOINE, D. et MAURY, O. (2013). Dynamic biogeochemical provinces in the global ocean. *Global Biogeochemical Cycles*, 27(4):1046–1058.
- RICHARDSON, A., WALNE, A., JOHN, A., JONAS, T., LINDLEY, J., SIMS, D., STEVENS, D. et WITT, M. (2006). Using continuous plankton recorder data. *Progress in Oceanography*, 68(1):27–74.
- RICHARDSON, A. J. et SCHOEMAN, D. S. (2004). Climate impact on plankton ecosystems in the northeast atlantic. *Science*, 305(5690):1609–1612.
- ROBINSON, G. (1970). Continuous plankton records : variation in the seasonal cycle of phytoplankton in the north atlantic. *Bulletins of Marine Ecology*.
- ROUSSEAU, V., LEYNAERT, A., DAUD, N., LANCELOT, C. et al. (2002). Diatom succession, silicification and silicic acid availability in belgian coastal waters (southern north sea). *Marine Ecology Progress Series*, 236:61–73.
- SADEGHI, A., DINTER, T., VOUNTAS, M., TAYLOR, B., ALTENBURG-SOPPA, M., PEEKEN, I. et BRACHER, A. (2012). Improvement to the PhytoDOAS method for identification of coccolithophores using hyper-spectral satellite data. *Ocean Science*, 8(6):1055.
- SAHFOS (2015). Dinoflagellate, diatom, chlorophytes, coccolithophores and trichodesmium data for the north atlantic.
- SAILLEY, S., VOGT, M., DONEY, S., AITA, M., BOPP, L., BUITENHUIS, E., HASHIOKA, T., LIMA, I., QUÉRÉ, C. L. et YAMANAKA, Y. (2013). Comparing food web structures and dynamics across a suite of global marine ecosystem models. *Ecological Modelling*, 261-262.
- SAPP, M., WICHELS, A., WILTSHIRE, K. H. et GERDTS, G. (2007). Bacterial community dynamics during the winter–spring transition in the north sea. *FEMS Microbiology Ecology*, 59(3):622–637.
- SARTHOU, G., TIMMERMANS, K. R., BLAIN, S. et

- TRÉGUER, P. (2005). Growth physiology and fate of diatoms in the ocean : a review. *Journal of Sea Research*, 53(1):25–42.
- SATHYENDRANATH, S., COTA, G., STUART, V., MAASS, H. et PLATT, T. (2001). Remote sensing of phytoplankton pigments : a comparison of empirical and theoretical approaches. *International Journal of Remote Sensing*, 22(2-3):249–273.
- SATHYENDRANATH, S., LAZZARA, L. et PRIEUR, L. (1987). Variations in the spectral values of specific absorption of phytoplankton. *Limnology and Oceanography*, 32(2):403–415.
- SATHYENDRANATH, S. et MOREL, A. (1983). Light emerging from the sea—interpretation and uses in remote sensing. *In Remote sensing applications in marine science and technology*, pages 323–357. Springer.
- SATHYENDRANATH, S., WATTS, L., DEVRED, E., PLATT, T., CAVERHILL, C. et MAASS, H. (2004). Discrimination of diatoms from other phytoplankton using ocean-colour data. *Marine ecology progress series*, 272:59–68.
- SCHLOSS, I. R., ABELE, D., MOREAU, S., DEMERS, S., BERS, A. V., GONZÁLEZ, O. et FERREYRA, G. A. (2012). Response of phytoplankton dynamics to 19-year (1991–2009) climate trends in potter cove (antarctica). *Journal of marine Systems*, 92(1):53–66.
- SCHOEMANN, V., BECQUEVORT, S., STEFELS, J., ROUSSEAU, V. et LANCELOT, C. (2005). Phaeocystis blooms in the global ocean and their controlling mechanisms : a review. *Journal of Sea Research*, 53(1):43–66.
- SCHRÖDER, M., SONDERMANN, M., SURES, B. et HERING, D. (2015). Effects of salinity gradients on benthic invertebrate and diatom communities in a german lowland river. *Ecological Indicators*, 57:236–248.
- SCHWARTZ, M. D. (2003). *Phenology : an integrative environmental science*. Springer.
- SHIMODAIRA, H. (2002). An approximately unbiased test of phylogenetic tree selection. *Systematic biology*, 51(3):492–508.
- SHIMODAIRA, H. *et al.* (2004). Approximately unbiased tests of regions using multistep-multiscale bootstrap resampling. *The Annals of Statistics*, 32(6):2616–2641.
- SIEBURTH, J. M., SMETACEK, V. et LENZ, J. (1978). Pelagic ecosystem structure : heterotrophic compartments of the plankton and their relationship to plankton size fractions. *Limnology and oceanography*, 23(6):1256–1263.
- SIEGEL, D., DONEY, S. et YODER, J. (2002). The north atlantic spring phytoplankton bloom and sverdrup’s critical depth hypothesis. *Science*, 296(5568):730–733.
- SIEGEL, D., MARITORENA, S., NELSON, N., BEHRENFELD, M. et MCCLAIN, C. (2005). Colored dissolved organic matter and its influence on the satellite-based characterization of the ocean biosphere. *Geophysical Research Letters*, 32(20).
- 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.
- SMETACEK, V., KLAAS, C., STRASS, V. H., ASSMY, P., MONTRESOR, M., CISEWSKI, B., SAVOYE, N., WEBB, A., D’OVIDIO, F., ARRIETA, J. M. *et al.* (2012). Deep carbon export from a southern ocean iron-fertilized diatom bloom. *Nature*, 487(7407):313–319.
- SMETACEK, V., SCHAREK, R. et NOTHIG, E. M. (1990). Seasonal and regional variation in the pelagical and its relationships to the life cycle of krill in antarctic ecosystems. *Ecological change and conservation*.
- SMITH, W. O., DENNETT, M. R., MATHOT, S. et CARON, D. A. (2003). The temporal dynamics of the flagellated and colonial stages of phaeocystis antarctica in the ross sea. *Deep Sea Research Part II : Topical Studies in Oceanography*, 50(3):605–617.
- SMYTH, T. J., MOORE, G. F., GROOM, S. B., LAND, P. E. et TYRRELL, T. (2002). Optical modeling and measurements of a coccolithophore

- bloom. *Applied Optics*, 41(36):7679–7688.
- 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. *Appl. Opt.*, 40(18):2929–2945.
- STRAMSKI, D. et KIEFER, D. A. (1991). Light scattering by microorganisms in the open ocean. *Progress in Oceanography*, 28(4):343–383.
- STRAMSKI, D., REYNOLDS, R. A., BABIN, M., KACZMAREK, S., LEWIS, M. R., RÖTTGERS, R., SCIANDRA, A., STRAMSKA, M., TWARDOWSKI, M. S., FRANZ, B. A. et CLAUSTRE, H. (2008). Relationships between the surface concentration of particulate organic carbon and optical properties in the eastern south pacific and eastern atlantic oceans. *Biogeosciences*, 5(1):171–201.
- SUBRAMANIAM, A., CARPENTER, E. J. et FALKOWSKI, P. G. (1999). Bio-optical properties of the marine diazotrophic cyanobacteria *Trichodesmium* spp. II. A reflectance model for remote sensing. *Limnol. Oceanogr.*, 44(3):618–627.
- SUNDA, W., KIEBER, D., KIENE, R. et HUNTSMAN, S. (2002). An antioxidant function for dmsp and dms in marine algae. *Nature*, 418(6895):317.
- SUNDA, W. G., HARDISON, R., KIENE, R. P., BUCCIARELLI, E. et HARADA, H. (2007). The effect of nitrogen limitation on cellular dmsp and dms release in marine phytoplankton : climate feedback implications. *Aquatic Sciences-Research Across Boundaries*, 69(3):341–351.
- SUNDA, W. G. et HUNTSMAN, S. A. (2004). Relationships among photoperiod, carbon fixation, growth, chlorophyll a, and cellular iron and zinc in a coastal diatom. *Limnology and oceanography*, 49(5):1742–1753.
- SUZUKI, R. et SHIMODAIRA, H. (2006). Pvcust : an r package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*, 22(12):1540–1542.
- SVERDRUP, H. U. (1953). On conditions for the vertical blooming of phytoplankton. *Journal du Conseil*, 18(3):287–295.
- THYSSEN, M., ALVAIN, S., LEFÈVRE, A., DESSAILLY, D., RIJKEBOER, M., GUISELIN, N., CREACH, V. et ARTIGAS, L.-F. (2015). High-resolution analysis of a north sea phytoplankton community structure based on in situ flow cytometry observations and potential implication for remote sensing. *Biogeosciences*, 12(13):4051–4066.
- 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.
- TRIGUEROS, J. M. et ORIVE, E. (2001). Seasonal variations of diatoms and dinoflagellates in a shallow, temperate estuary, with emphasis on neritic assemblages. *Hydrobiologia*, 444(1-3):119–133.
- 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 : Oceans*, 111(C8).
- VAN IPEREN, J., VAN BENNEKOM, A. et VAN WEEERING, T. (1993). Diatoms in surface sediments of the Indonesian archipelago and their relation to hydrography. In *Twelfth International Diatom Symposium*, pages 113–128. Springer.
- VAN LEEUWEN, S., TETT, P., MILLS, D. et van der MOLEN, J. (2015). Stratified and nonstratified areas in the north sea : Long-term variability and biological and policy implications. *Journal of Geophysical Research : Oceans*, 120(7):4670–4686.
- VANTREPOTTE, V., LOISEL, H., DESSAILLY, D. et MÉRIAUX, X. (2012). Optical classification of contrasted coastal waters. *Remote Sensing of Environment*, 123:306–323.
- VILLAREAL, T. A., WOODS, S., MOORE, J. K. et

- CULVERRYMSZA, K. (1996). Vertical migration of rhizosolenia mats and their significance to no<sub>3</sub>- fluxes in the central north pacific gyre. *Journal of Plankton Research*, 18(7):1103–1121.
- VOGT, M., HASHIOKA, T., PAYNE, M., BUITENHUIS, E., QUÉRÉ, C. L., ALVAIN, S., AITA, M., BOPP, L., DONEY, S., HIRATA, T. *et al.* (2013). The distribution, dominance patterns and ecological niches of plankton functional types in dynamic green ocean models and satellite estimates. *Biogeosciences Discussions*, 10(11):17193–17247.
- VOLPE, G., SANTOLERI, R., VELLUCCI, V., D'ALCALÀ, M. R., MARULLO, S. *et al.* (2007). The colour of the Mediterranean Sea : Global versus regional bio-optical algorithms evaluation and implication for satellite chlorophyll estimates. *Remote Sensing of Environment*, 107(4):625 – 638.
- WARNER, A. *et al.* HAYS, G. (1994). Sampling by the continuous plankton recorder survey. *Progress in Oceanography*, 34(2):237–256.
- WATERBURY, J. B., WATSON, S. W., VALOIS, F. W. *et al.* FRANKS, D. G. (1986). Biological and ecological characterization of the marine unicellular cyanobacterium *synechococcus*. *Can Bull Fish Aquat Sci*, 214(71):120.
- WERDELL, P. J. *et al.* BAILEY, S. W. (2005). An improved in-situ bio-optical data set for ocean color algorithm development and satellite data product validation. *Remote sensing of environment*, 98(1):122–140.
- WERDELL, P. J., FRANZ, B. A., BAILEY, S. W., FELDMAN, G. C., BOSS, E., BRANDO, V. E., DOWELL, M., HIRATA, T., LAVENDER, S. J., LEE, Z., LOISEL, H., MARITORENA, S., MÉLIN, F., MOORE, T. S., SMYTH, T. J., ANTOINE, D., DEVRED, E., D'ANDON, O. H. F. *et al.* MANGIN, A. (2013). Generalized ocean color inversion model for retrieving marine inherent optical properties. *Appl. Opt.*, 52(10):2019–2037.
- WESTBERRY, T. K., SIEGEL, D. A. *et al.* SUBRAMANIAM, A. (2005). An improved bio-optical model for the remote sensing of *Trichodesmium spp.* blooms. *Journal of Geophysical Research*, 110(C6).
- WHITMIRE, A. L., PEGAU, W. S., KARP-BOSS, L., BOSS, E. *et al.* COWLES, T. J. (2010). Spectral backscattering properties of marine phytoplankton cultures. *Optics Express*, 18(14):15073–15093.
- WILLIAMS, R. *et al.* LINDLEY, J. (1980). Plankton of the fladen ground during flex 76 i. spring development of the plankton community. *Marine Biology*, 57(2):73–78.
- WILTSHIRE, K. H., KRABERG, A., BARTSCH, I., BOERSMA, M., FRANKE, H.-D., FREUND, J., GEBÜHR, C., GERDTS, G., STOCKMANN, K. *et al.* WICHELS, A. (2010). Helgoland roads, north sea : 45 years of change. *Estuaries and Coasts*, 33(2):295–310.
- WILTSHIRE, K. H., MALZAHN, A. M., WIRTZ, K., GREVE, W., JANISCH, S., MANGELSDORF, P., MANLY, B. F. *et al.* BOERSMA, M. (2008). Resilience of north sea phytoplankton spring bloom dynamics : An analysis of long-term data at helgoland roads. *Limnology and Oceanography*, 53(4):1294–1302.
- ZUBKOV, M. V., SLEIGH, M. A., BURKILL, P. H. *et al.* LEAKEY, R. J. (2000). Picoplankton community structure on the Atlantic Meridional Transect : a comparison between seasons. *Progress in oceanography*, 45(3):369–386.
- ZWIRGLMAIER, K., HEYWOOD, J. L., CHAMBERLAIN, K., WOODWARD, E. M. S., ZUBKOV, M. V. *et al.* SCANLAN, D. J. (2007). Basin-scale distribution patterns of picocyanobacterial lineages in the atlantic ocean. *Environmental Microbiology*, 9(5):1278–1290.

---

---

# ANNEXES

---

---



# Annexe 1 : Liste des radiomètres principaux à portée globale

---

## ***CZCS - Coastal Zone Colour Scanner :***

Satellite : Nimbus-7 (NASA)

Période d'observation de l'océan : octobre 1978 à juin 1986

Résolution spatiale d'une image : 825 m

Nombre de bandes spectrales : 6

## ***POLDER (1, 2 et 3) - Polarization and directionality of the earth's reflectances :***

Satellite : Adéos (POLDER 1), Adéos 2 (POLDER 2), Parasol (POLDER 3); capteurs CNES/CNRS embarqués sur satellites Japonais (Adéos 1 et 2), puis français (Parasol)

Période d'observation de l'océan : août 1996 à juillet 1997 (POLDER 1), décembre 2002 à octobre 2003 (POLDER 2), décembre 2004 à 2012 (POLDER 3)

Résolution spatiale d'une image : 6000 m

Nombre de bandes spectrales : 9

## ***SeaWiFS - Sea-viewing Wide-Field-of-View :***

Satellite : SeaStar (NASA)

Période d'observation de l'océan : septembre 1997 à décembre 2010

Résolution spatiale d'une image : 1100 m

Nombre de bandes spectrales : 8

## ***MERIS - Medium Resolution Imaging Spectrometer :***

Satellite : Envisat (ESA)

Période d'observation de l'océan : mars 2002 à avril 2011

Résolution spatiale d'une image : 300 et 1200 m

Nombre de bandes spectrales : 15



**MODIS - *Moderate Resolution Imaging Spectroradiometer* :**

Satellite : Aqua ou EOS/PM-1 (NASA)

Période d'observation de l'océan : mai 2002 à aujourd'hui (2017)

Résolution spatiale d'une image : 1000 m

Nombre de bandes spectrales : 36

**VIIRS - *Visible Infrared Imager Radiometer Suite* :**

Satellite : Suomi-NPP (NASA)

Période d'observation de l'océan : octobre 2011 à aujourd'hui (2017)

Résolution spatiale d'une image : 270 et 740 m

Nombre de bandes spectrales : 22

# Annexe 2 : Cartes de Kohonen pour les capteurs MODIS, VIIRS, et MERIS

---

Spectres de références (informations condensées et classées de toutes les entrées) des unités de la carte auto-organisatrice des anomalies de radiance des capteurs de couleur de l'eau MODIS (Figure 1, ANNEXE 2), VIIRS (Figure 2, ANNEXE 2), et MERIS (Figure 3, ANNEXE 2), avec écarts types (lignes grises), à l'échelle globale. Pour chaque graphique, l'axe des abscisses représentent la longueur d'onde\*, et l'axe des ordonnées représente l'anomalie de radiance (sans unité), réalisé avec l'algorithme de Kohonen (Kohonen, 2013), sur la base de Ben Mustapha *et al.* (2014).

\* **MODIS** : 412, 443, 488, 531, 555 nm ; **VIIRS** : 410, 443, 486, 551 nm ; **MERIS** : 413, 443, 490, 510, 560 nm

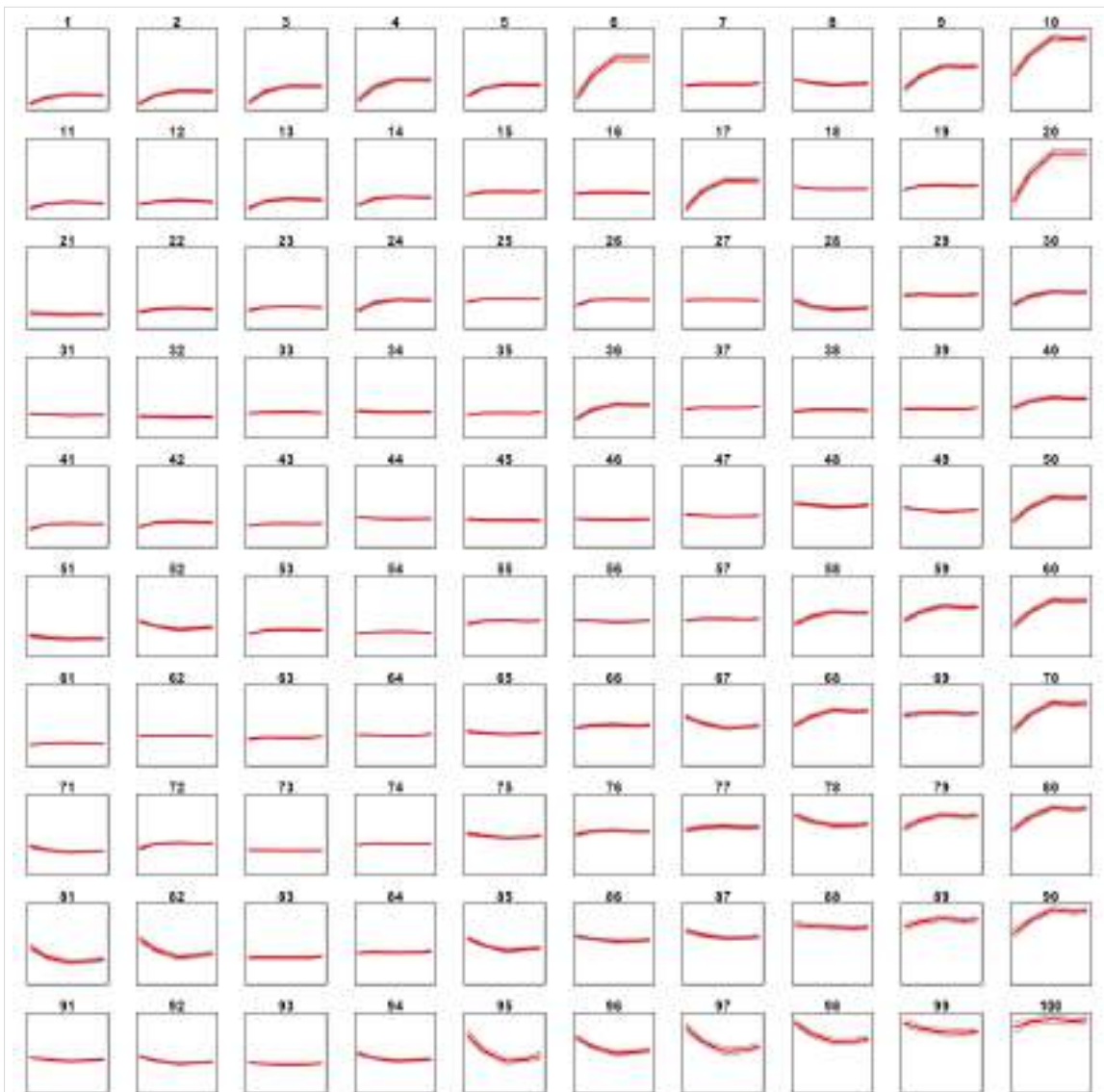


FIGURE 1, ANNEXE 2 – Carte auto-organisatrice du capteur MODIS.

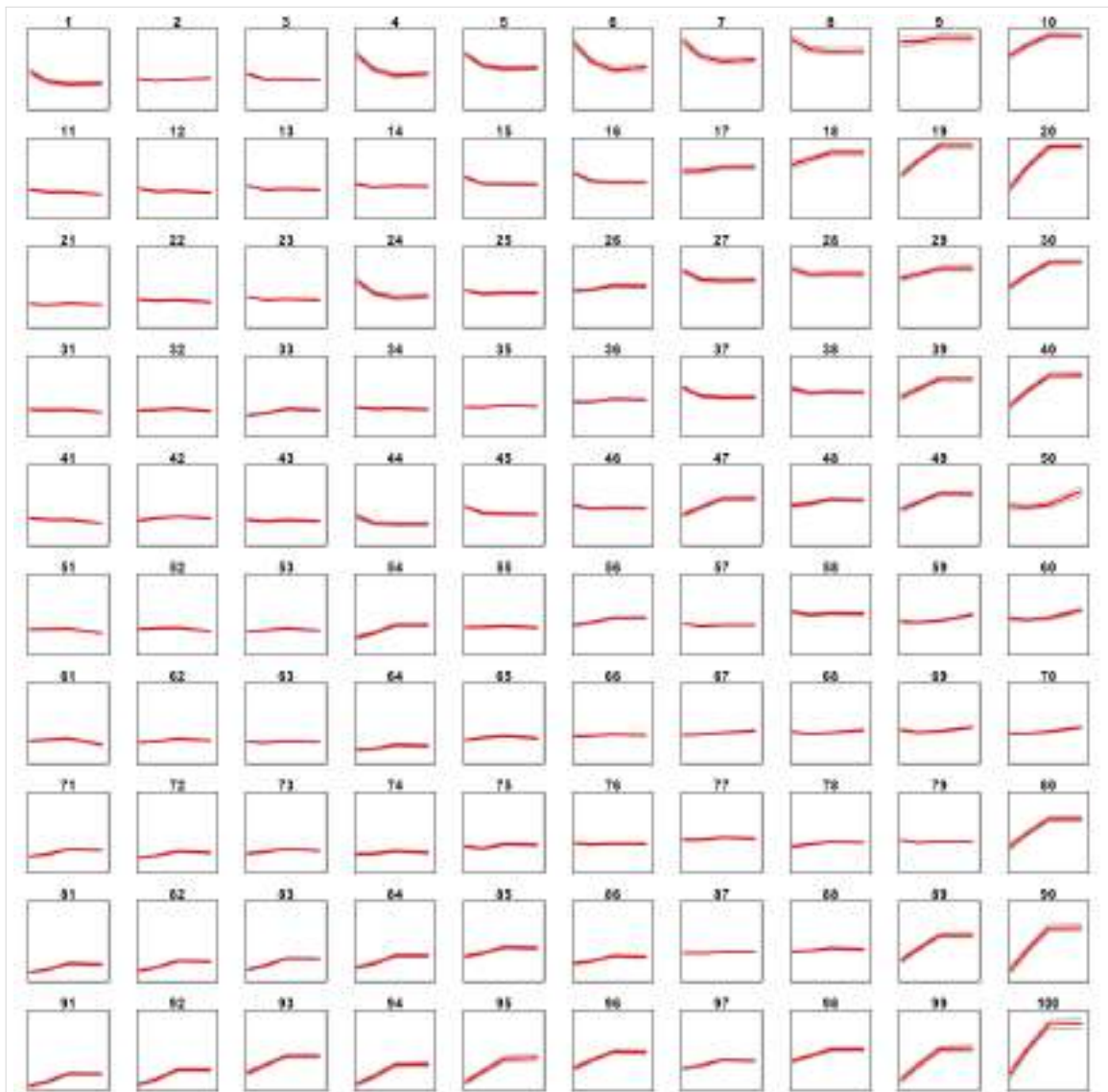


FIGURE 2, ANNEXE 2 – Carte auto-organisatrice du capteur VIIRS.

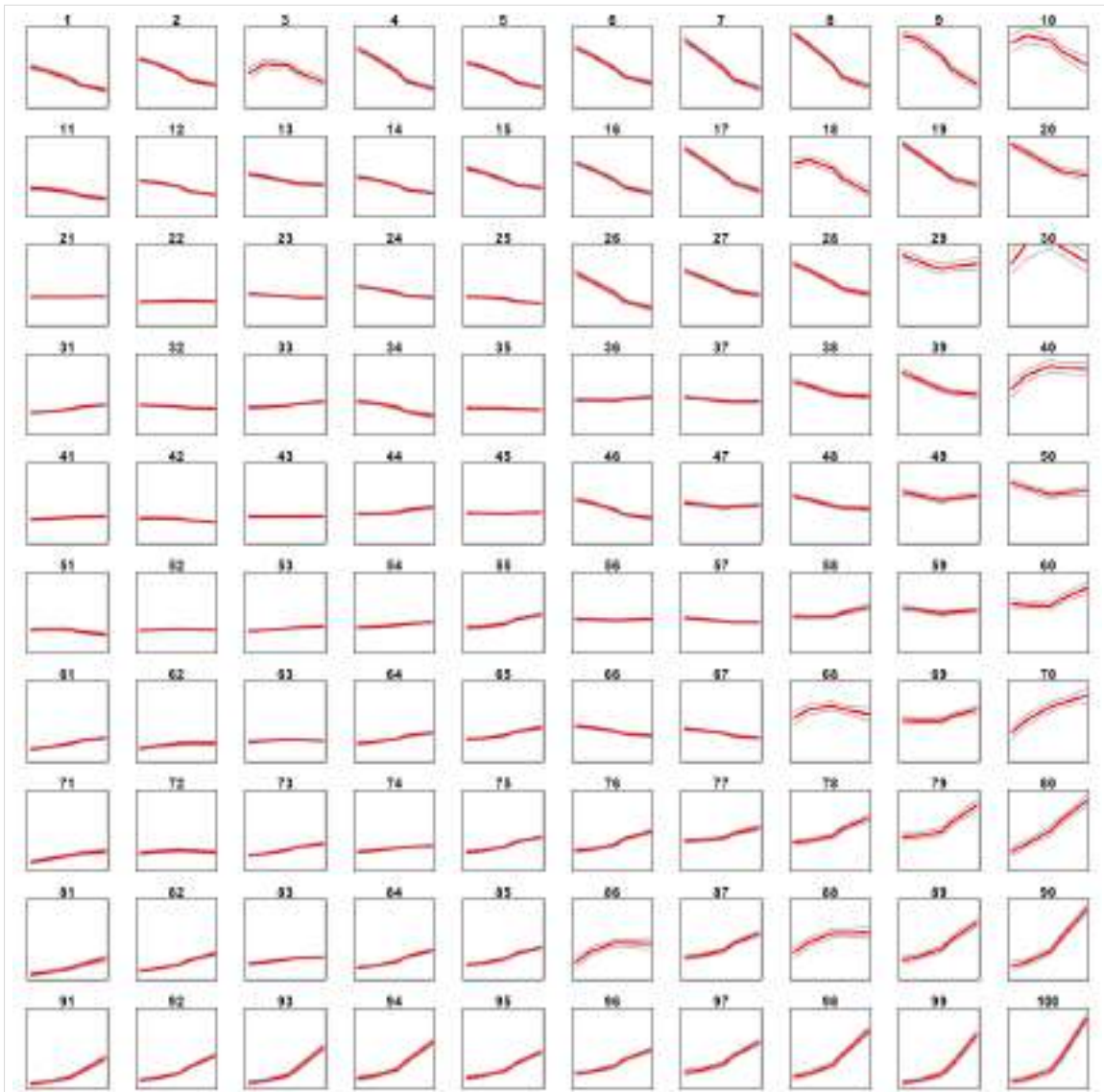


FIGURE 3, ANNEXE 2 – Carte auto-organisatrice du capteur MERIS.

Annexe 3 : Seuils des  
 $Ra(\lambda)$  définissant les signatures  
spectrales pour les groupes  
dominants dans la version Alvain  
*et al.* (2005)

---

TABLE 1, ANNEXE 3 – Seuils des  $Ra(\lambda)$  définissant les signatures spectrales pour le capteur de couleur de l'eau SeaWiFS et pour les groupes dominants de la version Alvain *et al.* (2005). Les spectres entrants dans ces critères ont donc été considérés comme spécifiques pour les groupes dominants suivants : les nanoeucaryotes dominants, les *Prochlorococcus spp.* dominants, les *Synechococcus spp.-like* dominants et les diatomées dominantes ; et pour le capteur de couleur de l'eau SeaWiFS.

	Ra(412nm)	Ra(443nm)	Ra(490nm)	Ra(510nm)	Ra(555nm)	Critères additionnels
Nanoeucaryotes (min)	0,4	0,55	0,6	0,6	0,6	Ra(412nm) < Ra(443nm)
Nanoeucaryotes (max)	0,8	0,9	0,95	1,0	1,0	Ra(443nm) > Ra(490nm)
<i>Prochlorococcus spp.</i> (min)	0,8	0,85	0,85	0,85	0,8	
<i>Prochlorococcus spp.</i> (max)	1,0	1,0	1,0	1,0	1,0	
<i>Synechococcus spp.-like</i> (min)	1,0	0,95	0,9	0,9	0,9	Ra(412nm) > Ra(443nm)
<i>Synechococcus spp.-like</i> (max)	1,3	1,2	1,2	1,2	1,2	Ra(412nm) < Ra(490nm)
Diatomées (min)	1,3	1,2	1,1	1,1	1,1	Ra(412nm) > Ra(490nm)
Diatomées (max)	2,4	2,0	1,7	1,6	1,6	Ra(490nm) > Ra(555nm)

# Annexe 4 : Séries temporelles de fréquences de détection des 100 classes de $Ra(\lambda)$ (capteur SeaWiFS) issues de la carte auto-organisatrice

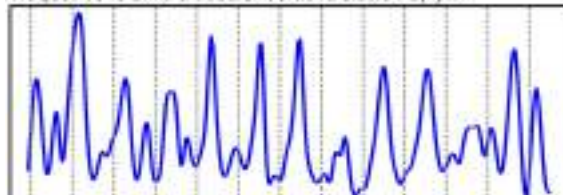
---

Les graphiques des pages suivantes représentent les fréquences de détection à 14 jours des signaux d'anomalies de radiance, groupées en 100 classes par une carte auto-organisatrice (Ben Mustapha *et al.*, 2014), pour le capteur de couleur de l'eau SeaWiFS à l'échelle globale. Les axes des ordonnées représentent la fréquence relative d'une classe par rapport à toutes les autres classes (normalisées par leurs maximums\*). Les axes des abscisses représentent le temps, avec un trait vertical en pointillés pour le 1er janvier de chaque année, de 1998 (première année de données complète pour le capteur SeaWiFS) à 2010 (fin d'activité du capteur SeaWiFS).

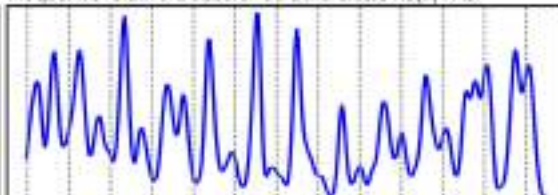
\* comme il l'a été précisé dans le Chapitre 3, les valeurs de la fréquence restent subjectives car elles dépendent de la représentativité de la classe au sein de la carte auto-organisatrice (i.e., le nombre de pixels attribués à cette classe). C'est pourquoi ces graphiques ont été normalisés par leurs maximums de fréquence respectifs.



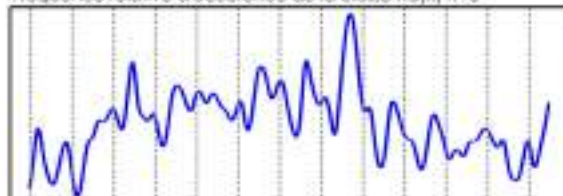
Fréquence relative d'occurrence de la classe Ra(A) n° 1



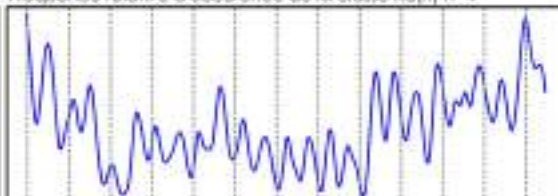
Fréquence relative d'occurrence de la classe Ra(A) n° 2



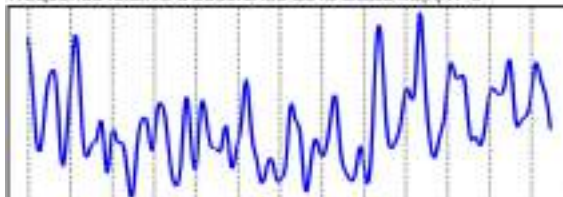
Fréquence relative d'occurrence de la classe Ra(A) n° 3



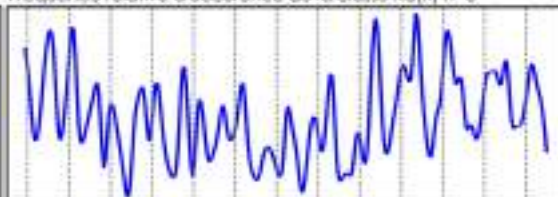
Fréquence relative d'occurrence de la classe Ra(A) n° 4



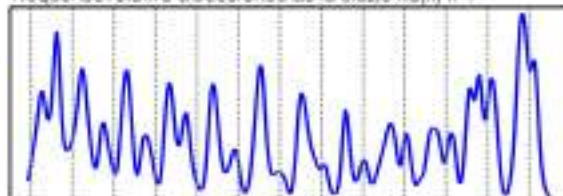
Fréquence relative d'occurrence de la classe Ra(A) n° 5



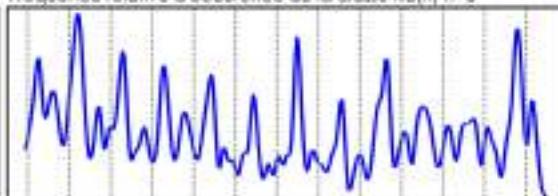
Fréquence relative d'occurrence de la classe Ra(A) n° 6



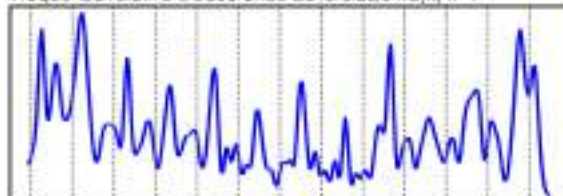
Fréquence relative d'occurrence de la classe Ra(A) n° 7



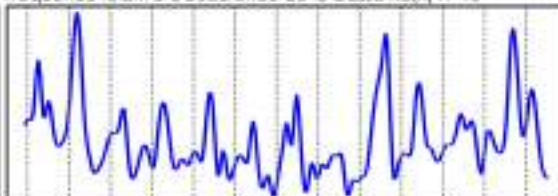
Fréquence relative d'occurrence de la classe Ra(A) n° 8



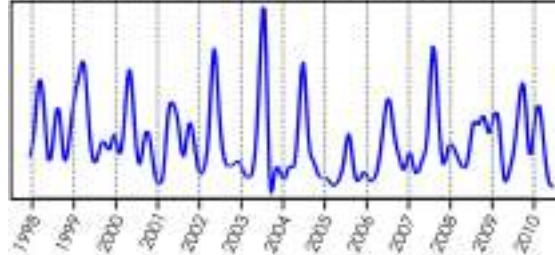
Fréquence relative d'occurrence de la classe Ra(A) n° 9



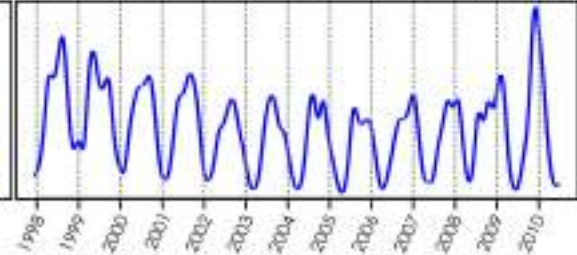
Fréquence relative d'occurrence de la classe Ra(A) n° 10



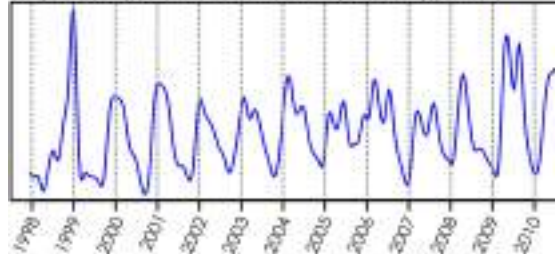
Fréquence relative d'occurrence de la classe Ra( $\lambda$ ) n° 11



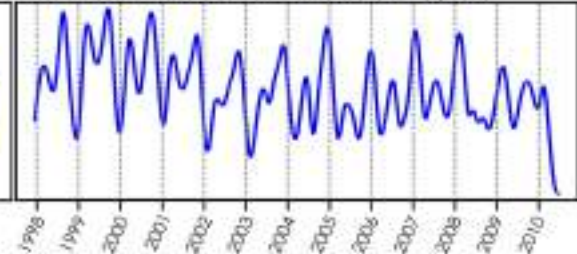
Fréquence relative d'occurrence de la classe Ra( $\lambda$ ) n° 12



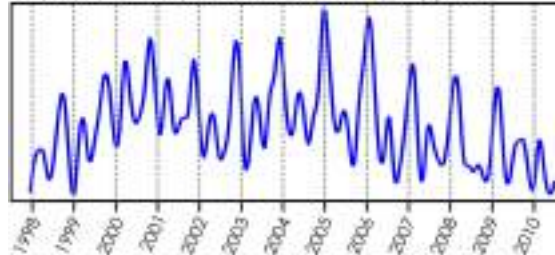
Fréquence relative d'occurrence de la classe Ra( $\lambda$ ) n° 13



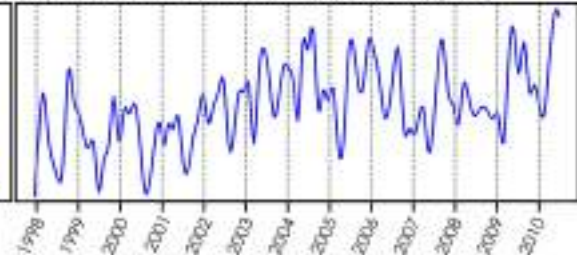
Fréquence relative d'occurrence de la classe Ra( $\lambda$ ) n° 14



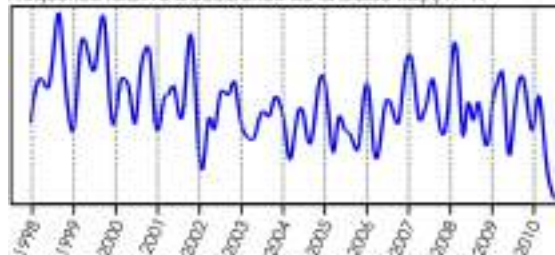
Fréquence relative d'occurrence de la classe Ra( $\lambda$ ) n° 15



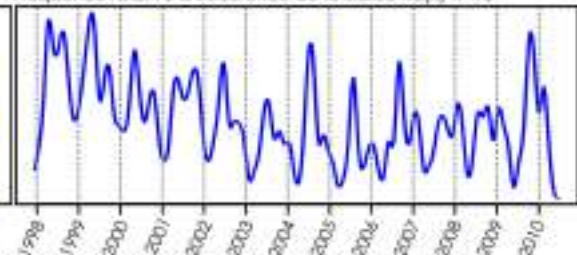
Fréquence relative d'occurrence de la classe Ra( $\lambda$ ) n° 16



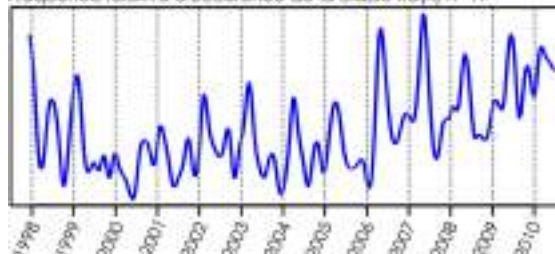
Fréquence relative d'occurrence de la classe Ra( $\lambda$ ) n° 17



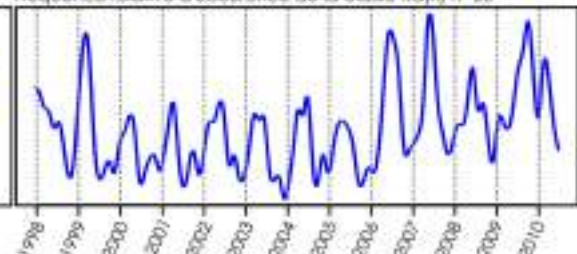
Fréquence relative d'occurrence de la classe Ra( $\lambda$ ) n° 18



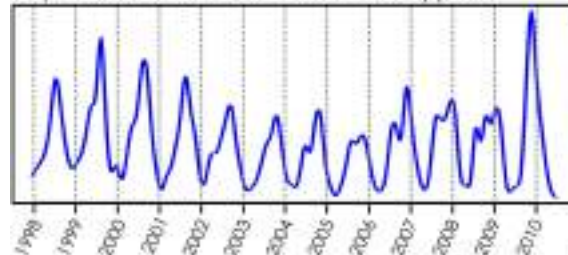
Fréquence relative d'occurrence de la classe Ra( $\lambda$ ) n° 19



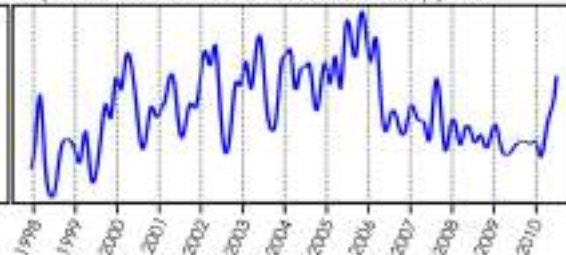
Fréquence relative d'occurrence de la classe Ra( $\lambda$ ) n° 20



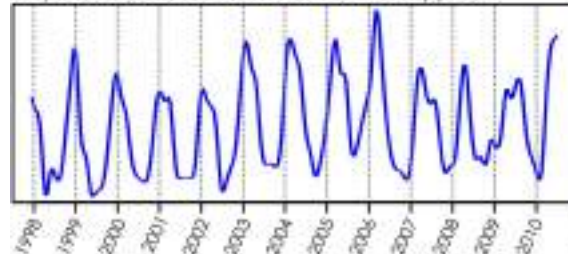
Fréquence relative d'occurrence de la classe Ra(k) n° 21



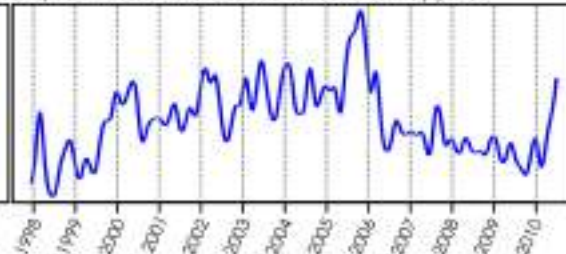
Fréquence relative d'occurrence de la classe Ra(k) n° 22



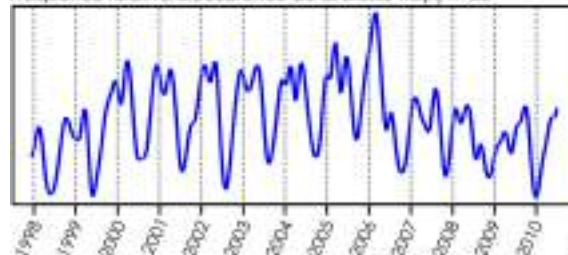
Fréquence relative d'occurrence de la classe Ra(k) n° 23



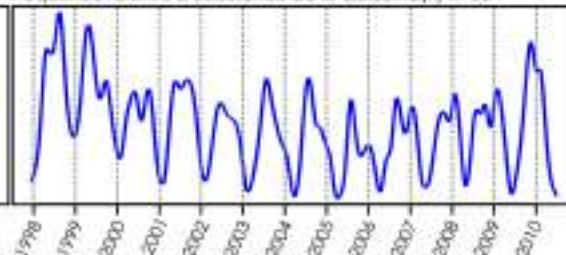
Fréquence relative d'occurrence de la classe Ra(k) n° 24



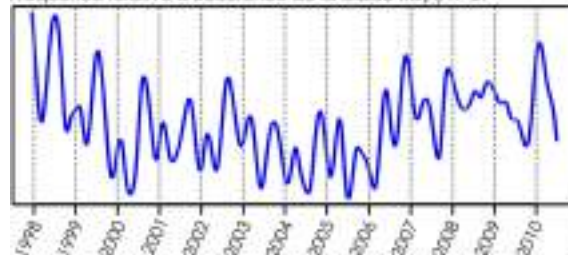
Fréquence relative d'occurrence de la classe Ra(k) n° 25



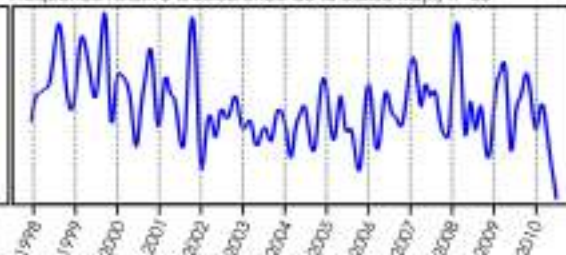
Fréquence relative d'occurrence de la classe Ra(k) n° 26



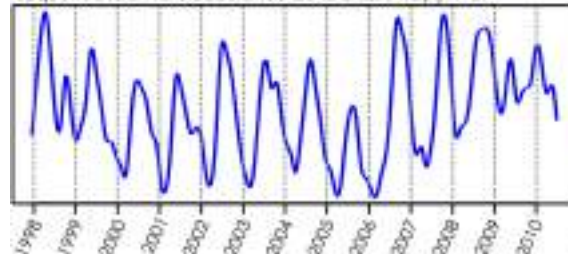
Fréquence relative d'occurrence de la classe Ra(k) n° 27



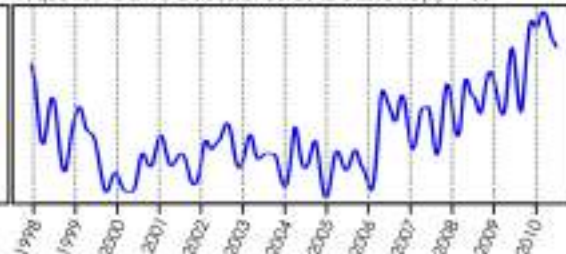
Fréquence relative d'occurrence de la classe Ra(k) n° 28

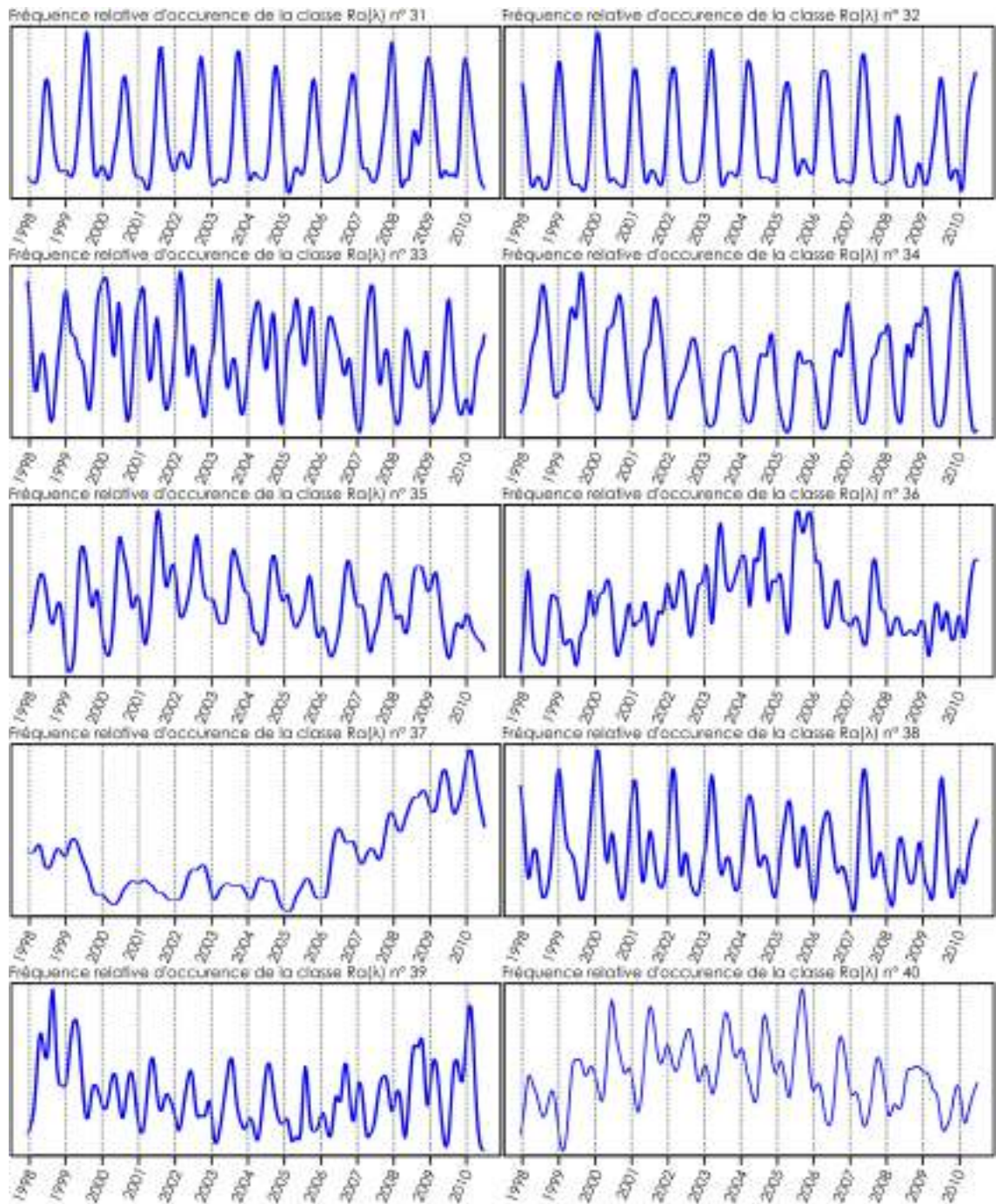


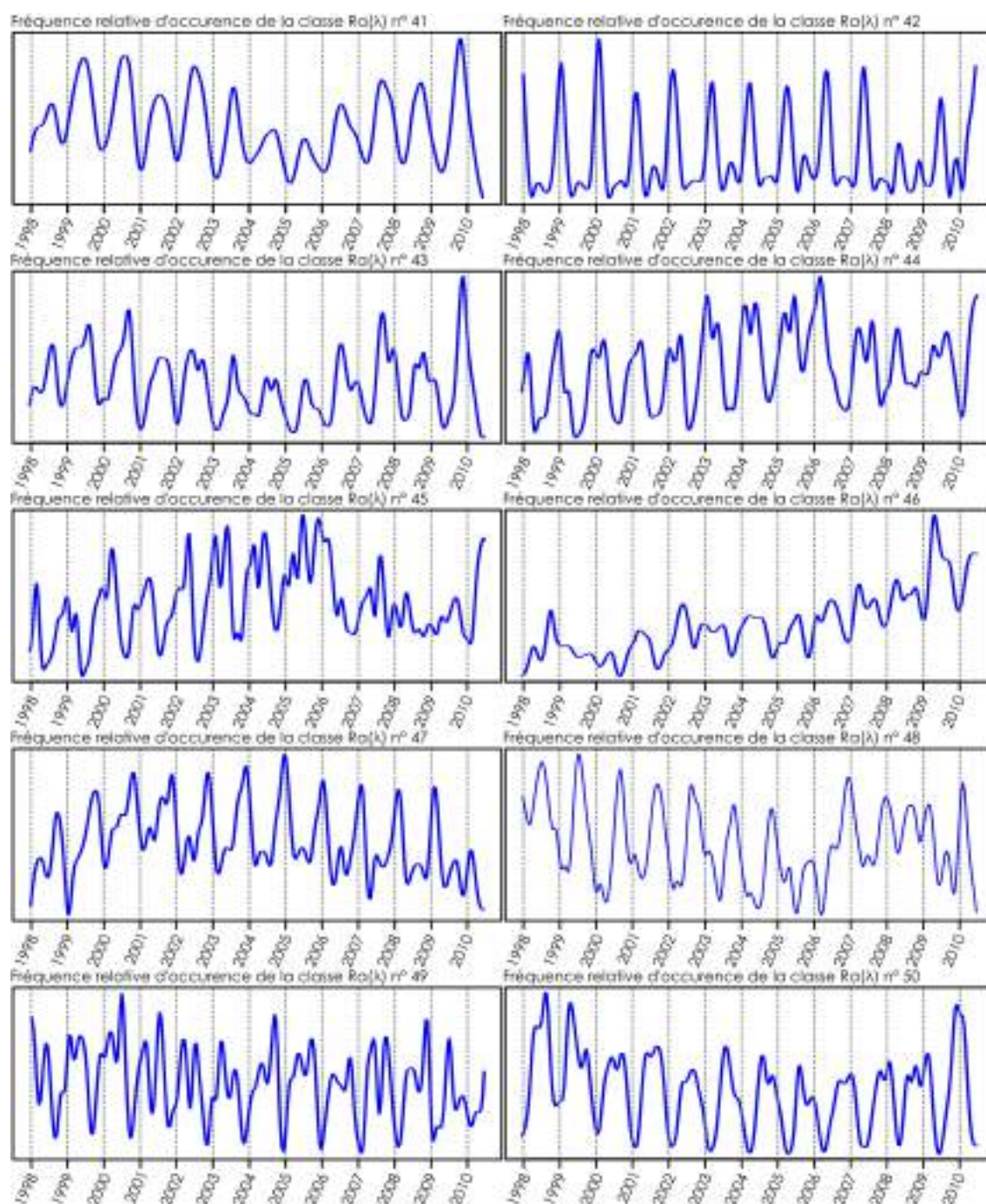
Fréquence relative d'occurrence de la classe Ra(k) n° 29

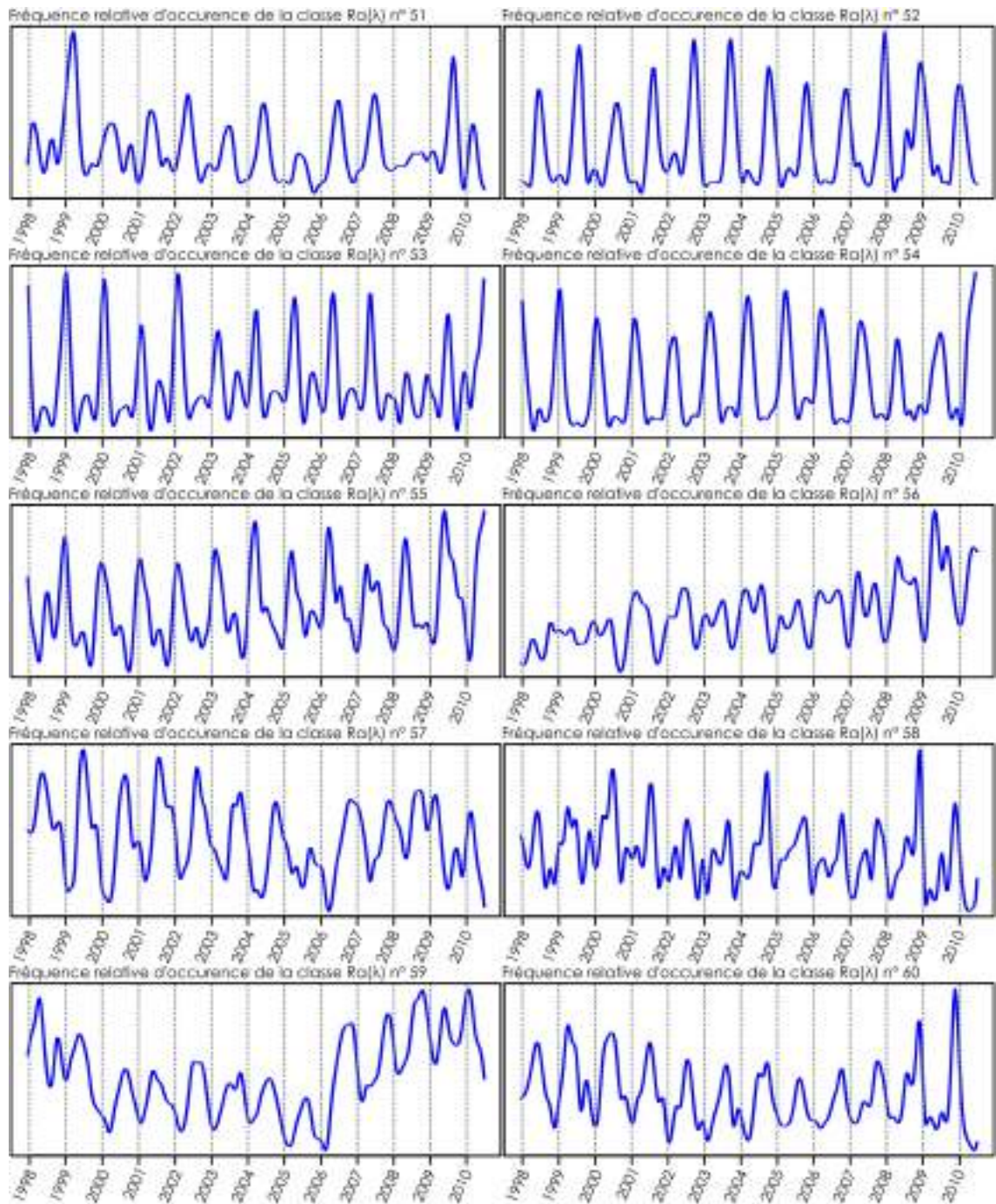


Fréquence relative d'occurrence de la classe Ra(k) n° 30

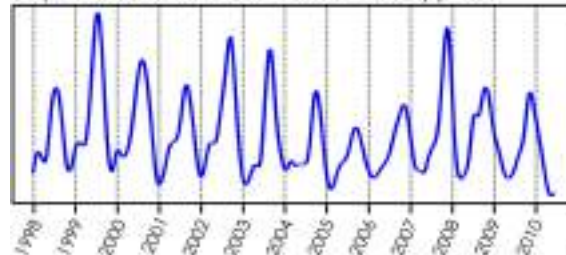




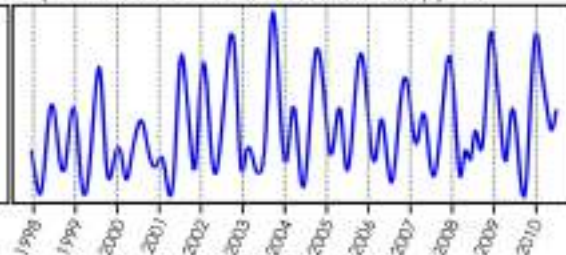




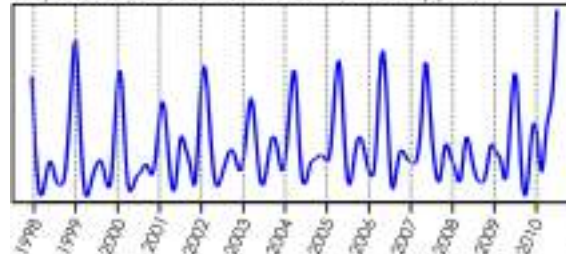
Fréquence relative d'occurrence de la classe  $Ra(\lambda)$  n° 61



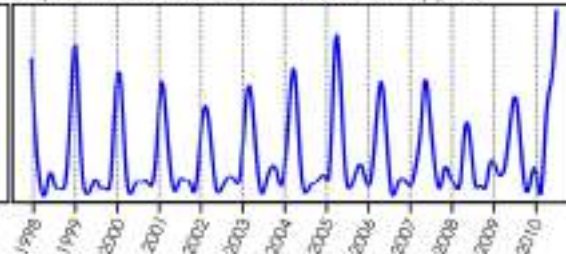
Fréquence relative d'occurrence de la classe  $Ra(\lambda)$  n° 62



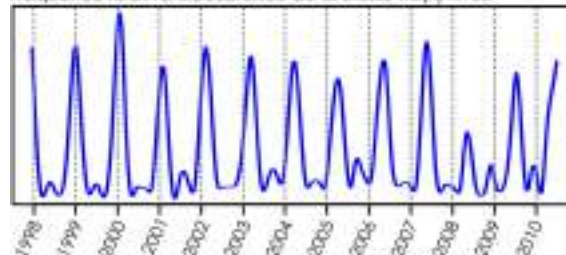
Fréquence relative d'occurrence de la classe  $Ra(\lambda)$  n° 63



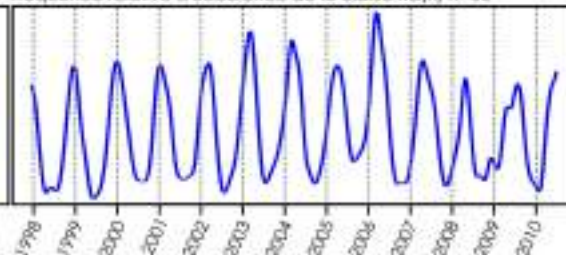
Fréquence relative d'occurrence de la classe  $Ra(\lambda)$  n° 64



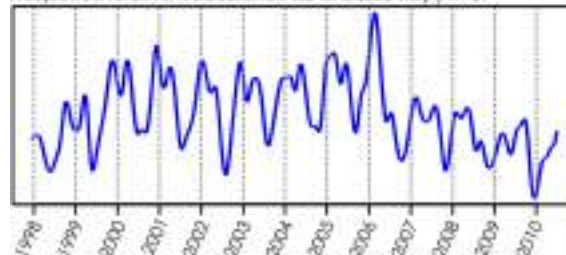
Fréquence relative d'occurrence de la classe  $Ra(\lambda)$  n° 65



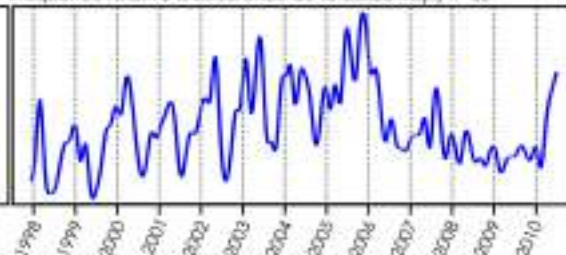
Fréquence relative d'occurrence de la classe  $Ra(\lambda)$  n° 66



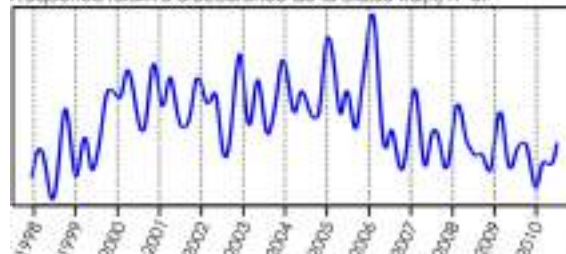
Fréquence relative d'occurrence de la classe  $Ra(\lambda)$  n° 67



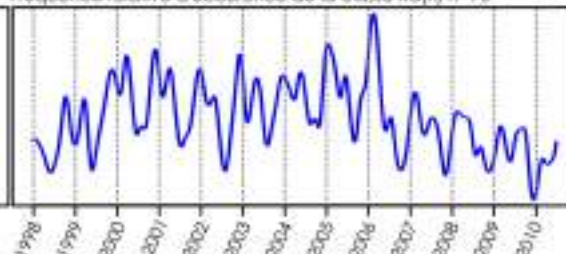
Fréquence relative d'occurrence de la classe  $Ra(\lambda)$  n° 68



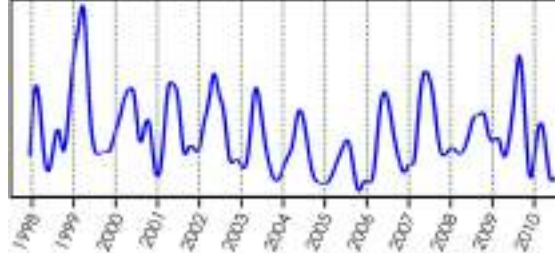
Fréquence relative d'occurrence de la classe  $Ra(\lambda)$  n° 69



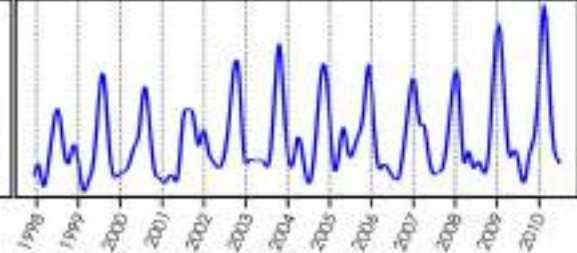
Fréquence relative d'occurrence de la classe  $Ra(\lambda)$  n° 70



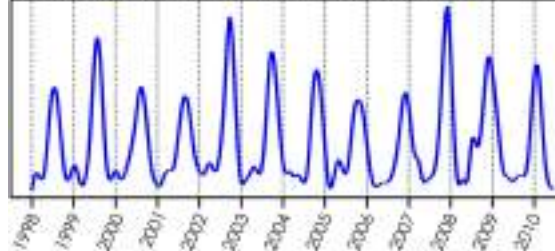
Fréquence relative d'occurrence de la classe Ra( $\lambda$ ) n° 71



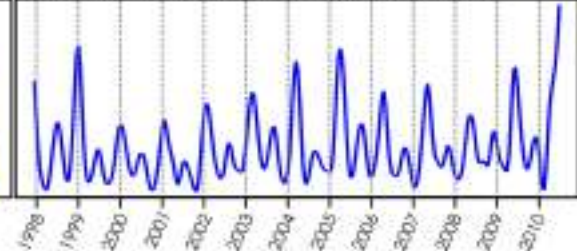
Fréquence relative d'occurrence de la classe Ra( $\lambda$ ) n° 72



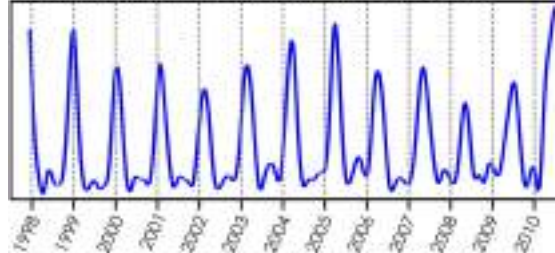
Fréquence relative d'occurrence de la classe Ra( $\lambda$ ) n° 73



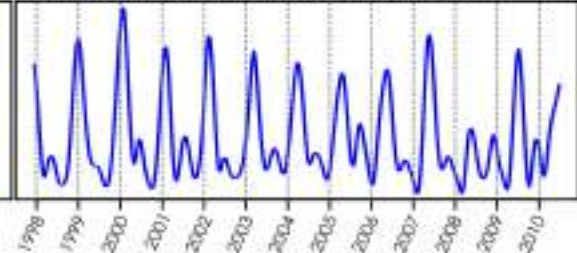
Fréquence relative d'occurrence de la classe Ra( $\lambda$ ) n° 74



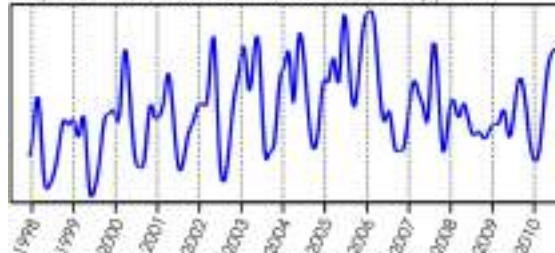
Fréquence relative d'occurrence de la classe Ra( $\lambda$ ) n° 75



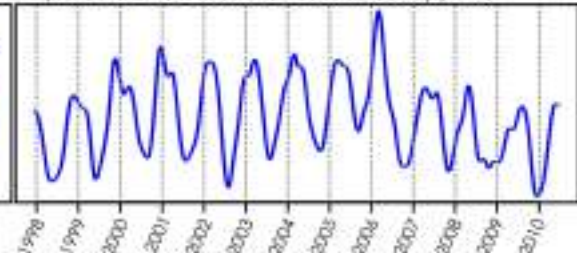
Fréquence relative d'occurrence de la classe Ra( $\lambda$ ) n° 76



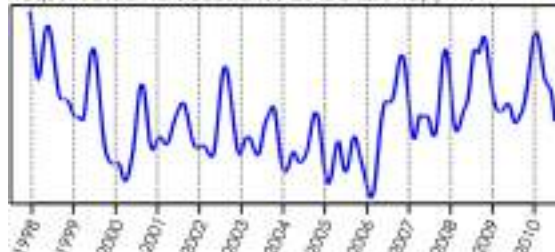
Fréquence relative d'occurrence de la classe Ra( $\lambda$ ) n° 77



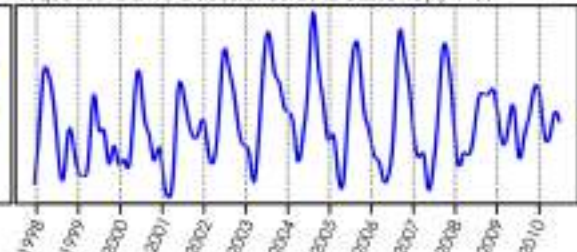
Fréquence relative d'occurrence de la classe Ra( $\lambda$ ) n° 78



Fréquence relative d'occurrence de la classe Ra( $\lambda$ ) n° 79

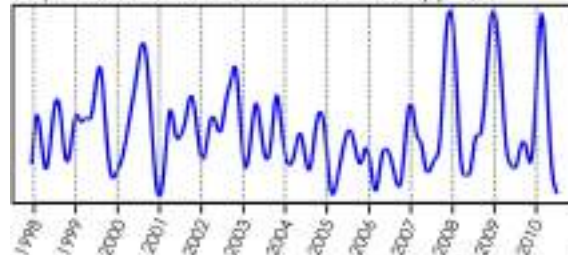


Fréquence relative d'occurrence de la classe Ra( $\lambda$ ) n° 80

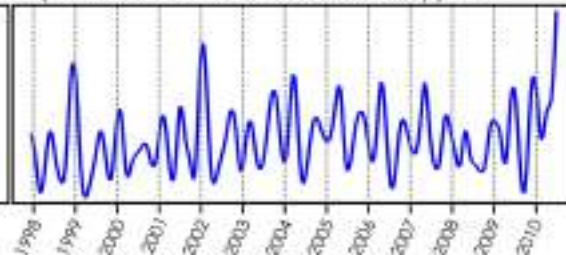




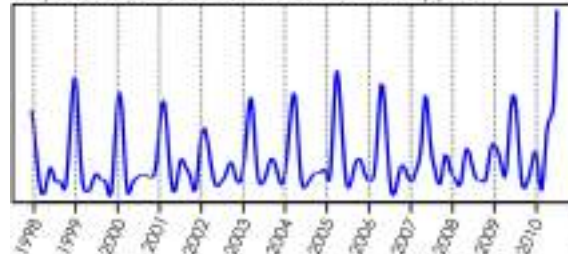
Fréquence relative d'occurrence de la classe  $Ra(\lambda)$  n° 81



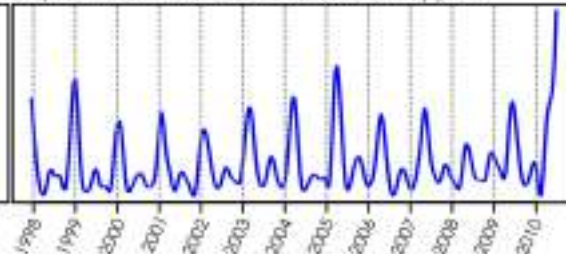
Fréquence relative d'occurrence de la classe  $Ra(\lambda)$  n° 82



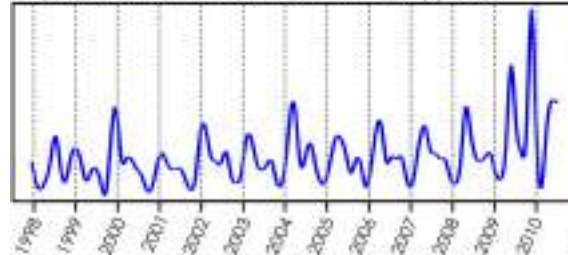
Fréquence relative d'occurrence de la classe  $Ra(\lambda)$  n° 83



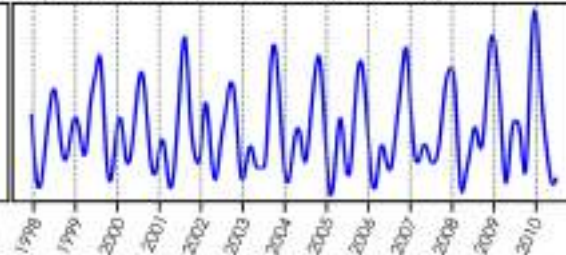
Fréquence relative d'occurrence de la classe  $Ra(\lambda)$  n° 84



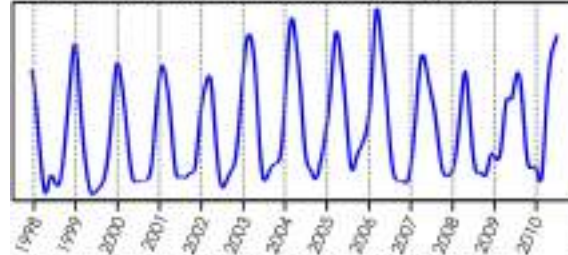
Fréquence relative d'occurrence de la classe  $Ra(\lambda)$  n° 85



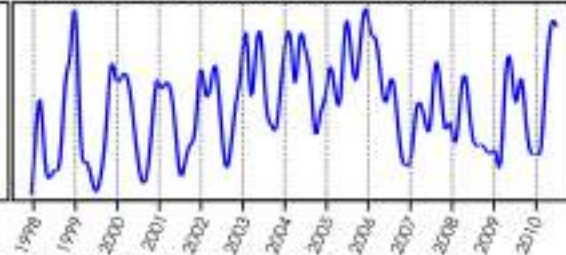
Fréquence relative d'occurrence de la classe  $Ra(\lambda)$  n° 86



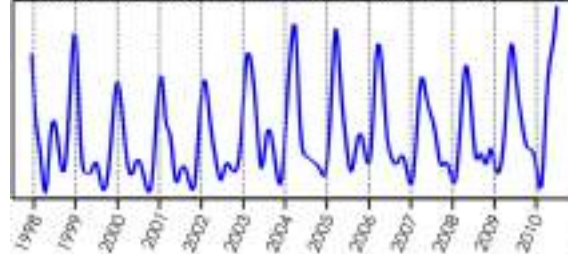
Fréquence relative d'occurrence de la classe  $Ra(\lambda)$  n° 87



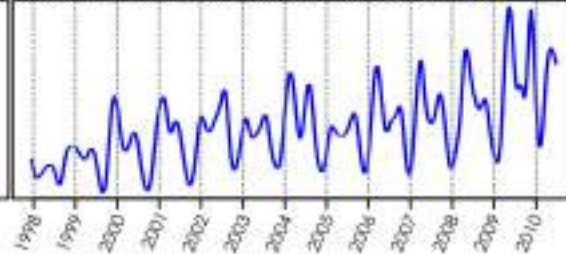
Fréquence relative d'occurrence de la classe  $Ra(\lambda)$  n° 88



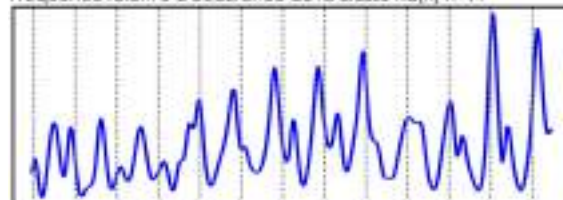
Fréquence relative d'occurrence de la classe  $Ra(\lambda)$  n° 89



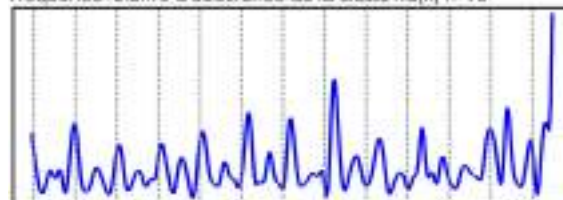
Fréquence relative d'occurrence de la classe  $Ra(\lambda)$  n° 90



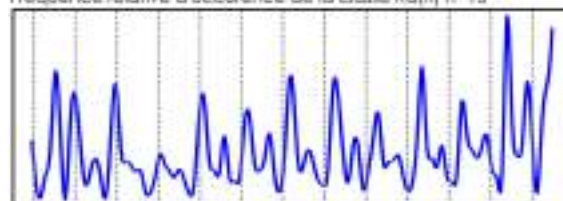
Fréquence relative d'occurrence de la classe Ra(A) n° 91



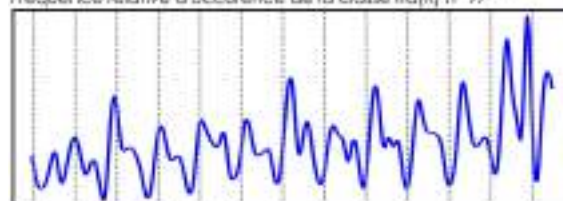
Fréquence relative d'occurrence de la classe Ra(A) n° 93



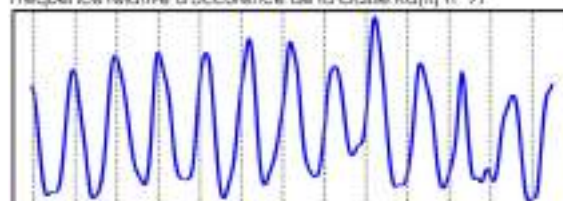
Fréquence relative d'occurrence de la classe Ra(A) n° 95



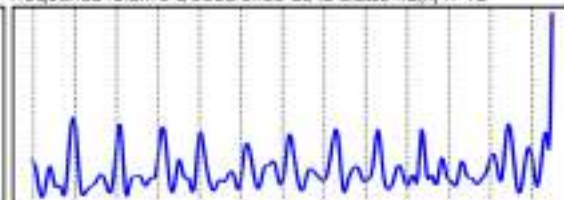
Fréquence relative d'occurrence de la classe Ra(A) n° 97



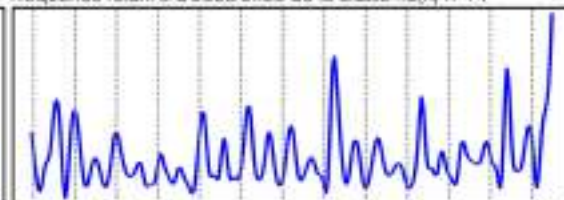
Fréquence relative d'occurrence de la classe Ra(A) n° 99



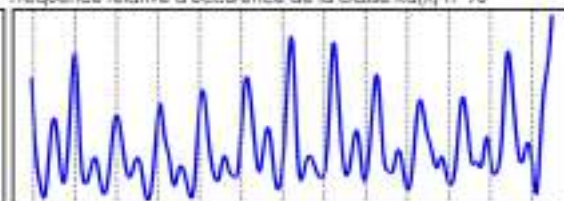
Fréquence relative d'occurrence de la classe Ra(A) n° 92



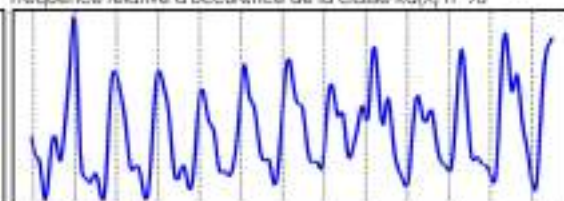
Fréquence relative d'occurrence de la classe Ra(A) n° 94



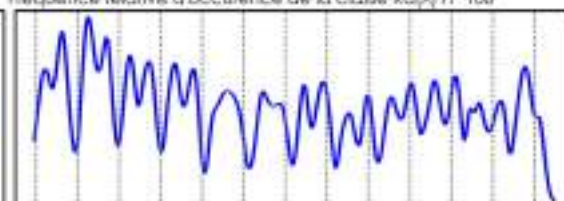
Fréquence relative d'occurrence de la classe Ra(A) n° 96



Fréquence relative d'occurrence de la classe Ra(A) n° 98



Fréquence relative d'occurrence de la classe Ra(A) n° 100





# Annexe 5 : Algorithme "FindPheno" développé en R, R Core Team (2015)

---

Où "vec" est la série temporelle et "med" le seuil "médiane +5%"

```
FindPheno <- fonction(vec,med,bw=1,x.coo=c(1:length(vec)))
{
  med<-median(vec[vec>0],na.rm=T)
  pos.x.max <- NULL ; pos.x.min <- NULL
  pos.y.max <- NULL ; pos.y.min <- NULL
  pos.x.init <- NULL ; pos.x.ter <- NULL
  pos.y.init <- NULL ; pos.y.ter <- NULL
  is.init<-NULL
  is.ter<-NULL
  for(i in 1:(length(vec)-1)) {
    if((i+1+bw)>length(vec)){
      sup.stop <- length(vec)} else {sup.stop <- i+1+bw
    }
    if((i-bw)<1){inf.stop <- 1} else {inf.stop <- i-bw-1}
    subset.sup <- vec[(i+1):sup.stop]
    subset.inf <- vec[inf.stop:(i-1)]
    subset.sup2 <- vec[i:length(vec)]
    subset.inf2 <- vec[i:1]
    is.max <- sum(subset.inf > vec[i]) == 0
    is.nomin <- sum(subset.sup > vec[i]) == 0
    no.max <- sum(subset.inf > vec[i]) == length(subset.inf)
    no.nomin <- sum(subset.sup > vec[i]) == length(subset.sup)
    if(is.max & is.nomin & vec[i] > med ){
      pos.x.max <- c(pos.x.max,x.coo[i])
      pos.y.max <- c(pos.y.max,vec[i])
      for(j in 1:length(subset.inf2)) {
        if(subset.inf2[j] > med) { is.init <- j
        } else break }
      if(length(is.init)>0) {
        if((i-is.init) >= 1 ){
```

```

        pos.y.init <- c(pos.y.init,vec[x.coo[(i-is.init)]])
        pos.x.init <- c(pos.x.init, x.coo[(i-is.init)]) } }
for(k in 1:length(subset.sup2)) {
  if(subset.sup2[k] > med) { is.ter <- k
  } else break }
if(length(is.init)>0) {
  if( (i+is.ter) < length(vec) ) {
    pos.y.ter <- c(pos.y.ter,vec[x.coo[(i+is.ter)]])
    pos.x.ter <- c(pos.x.ter, x.coo[(i+is.ter)]) } }
}
if(no.max & no.nomin & vec[i] < med ){
  pos.x.min <- c(pos.x.min,x.coo[i])
  pos.y.min <- c(pos.y.min,vec[i])
}
}
return(list(pos.x.max,pos.y.max,
           pos.x.min,pos.y.min,
           pos.x.init, pos.y.init,
           pos.x.ter,pos.y.ter))
}

```

Cette fonction R retourne une liste dont la structure est :

```

list[[1]] = moment du maximum
list[[2]] = valeur du maximum
list[[3]] = moment du minimum
list[[4]] = valeur du minimum
list[[5]] = moment de l'initiation
list[[6]] = valeur au moment d'initiation
list[[7]] = moment de terminaison
list[[8]] = valeur au moment de terminaison

```

**Annexe 6 : Cartes de fréquences  
des classes de  $Ra(\lambda)$  regroupées  
grâce à leurs ressemblances  
spectrales (super-classes *som*)**

---

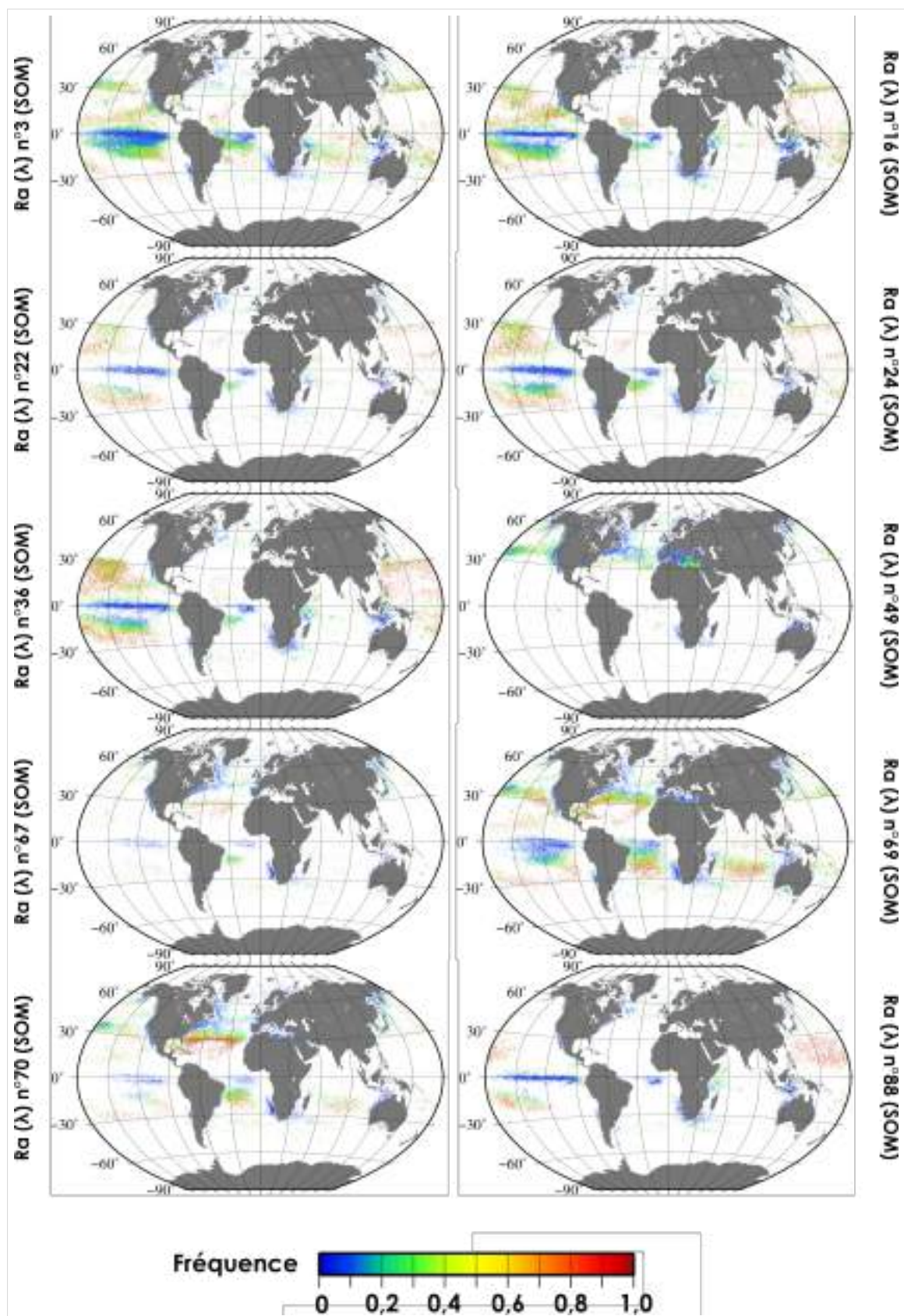
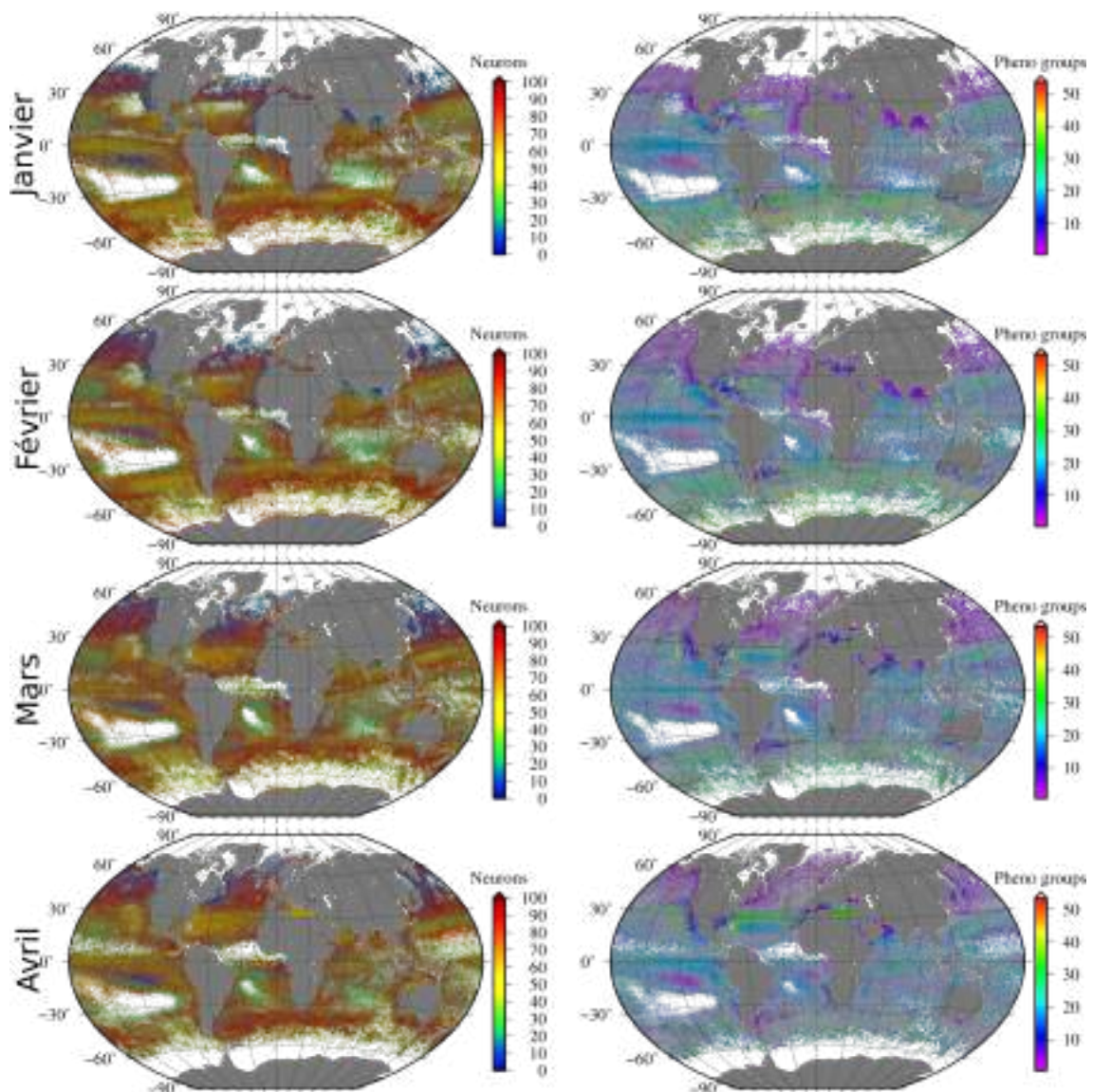


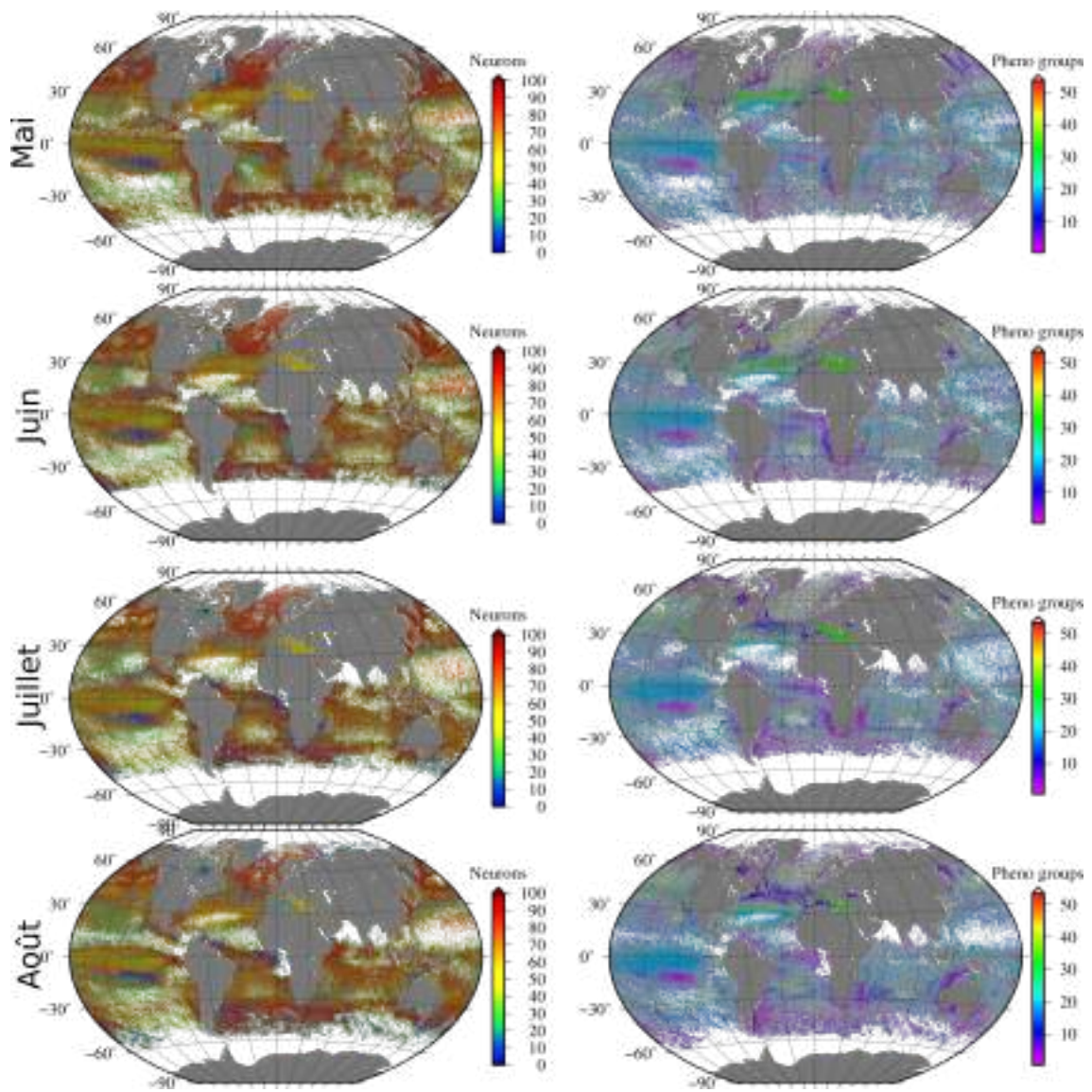
FIGURE 1, ANNEXE 6 – Carte des distributions spatiales des fréquences des classes de  $Ra(\lambda)$  SeaWiFS du mois de juillet (climatologie) pour les 10 classes de  $Ra(\lambda)$  issues de la carte auto-organisatrice regroupées en utilisant des super-classes spectrales. Cette méthode de regroupement des classes de  $Ra(\lambda)$  n'a pas été choisie, au vue des différences spatio-temporelles des occurrences de classes de  $Ra(\lambda)$  ayant des spectres similaires.

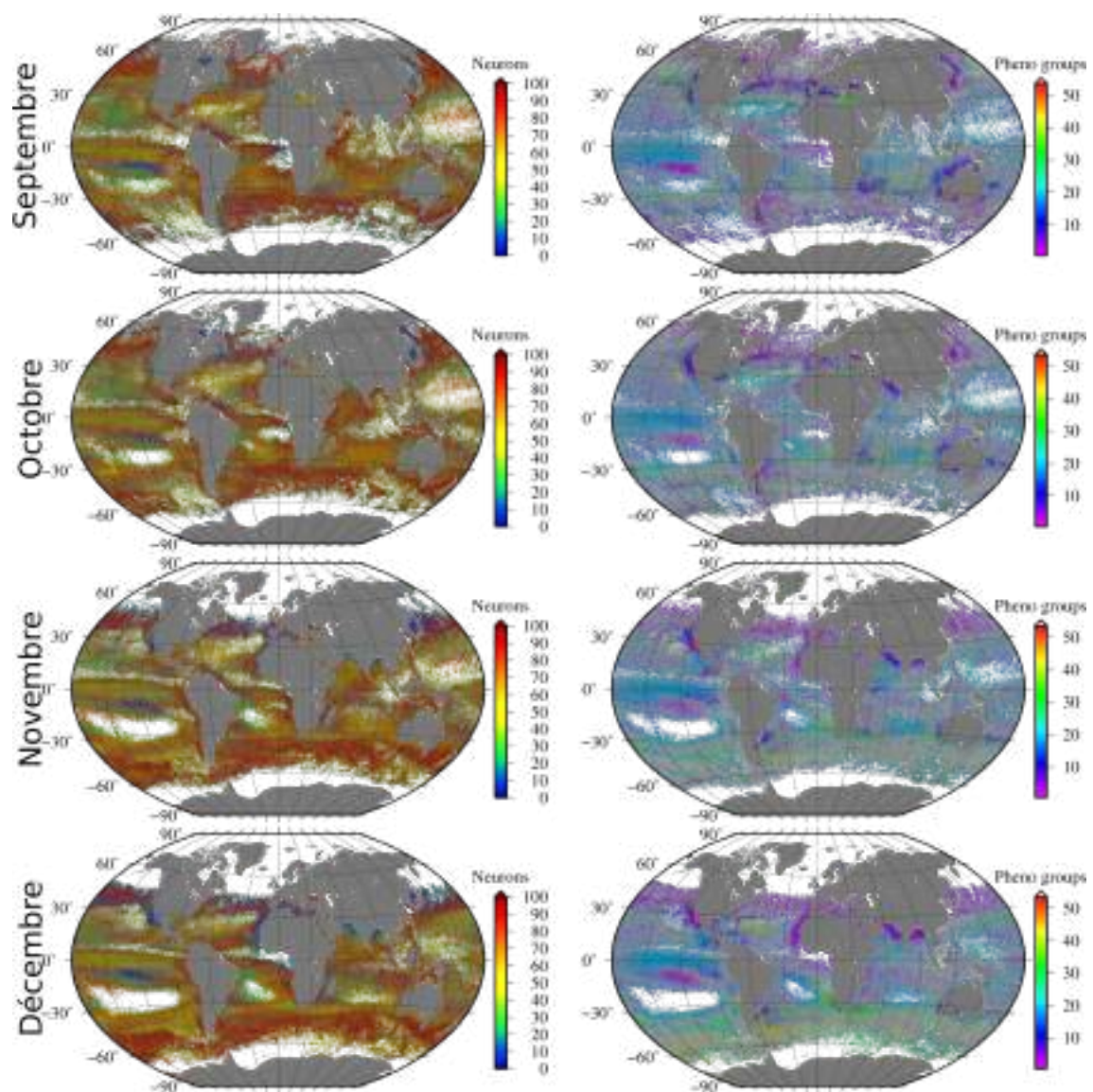




Annexe 7 : Carte d'occurrence des classes de  $Ra(\lambda)$  *som* (gauche) et des classes de  $Ra(\lambda)$  *phénologie* (droite) les plus fréquentes par mois (Climatologie SeaWiFS 1998-2010)







# Annexe 8 : Les campagnes en mer ayant collectés des échantillons pour analyses pigmentaires

---

## **Campagne OOMPH**

*OOMPH : Organics from the Oceans Modifying Particulates in both Hemispheres - Max-Planck Institute* (coordinateur)

Référent scientifique : Jonathan WILLIAMS (Max-Planck Institute)

27 Janvier au 4 Février 2007

Site web : <https://www.mpg.de/322136/forschungsSchwerpunkt1?c=119452>

## **Service d'observation - campagnes OISO**

*OISO : OCÉAN INDIEN SERVICE D'OBSERVATION - LOCEAN* (coordinateur)

Référent scientifique : Nicolas METZL et Claire LO MONACO (LOCEAN)

Nombreuses dates disponibles sur le site web

Site web : <http://campagnes.flotteoceanographique.fr/series/228/>

DOI : <http://dx.doi.org/10.18142/228>

## **Base de données MAREDAT**

*MAREDAT : MARine Ecosystem DATA - MAREMIP Project*

Référent scientifique : Chantal SWAN (IBP)

Site web : <http://archimer.ifremer.fr/doc/00128/23975/>

Base de données : <http://doi.pangaea.de/10.1594/PANGAEA.793246>

DOI : <http://doi.org/10.5194/essd-5-109-2013>

## **COLIBRI**

Référent scientifique : Yves DANDONNEAU (IRD/LODYC)

Atlantique-Nord, Avril et Mars

## **ICOTA**

*ICOTA : ICHTYOLOGIE COTIÈRE EN TERRE ADELIE - MNHM / UPMC* (coordinateur)

Référent scientifique : Philippe KOUBBI (MNHM / UPMC)

10/01/2005 (Dumont d'Urville) - 19/01/2005 (Dumont d'Urville)

Site web : <http://campagnes.flotteoceanographique.fr/campagnes/5210010/>

DOI : <http://dx.doi.org/10.17600/5210010>

### **CROZEX**

*CROZEX : The CROZet natural iron bloom and EXport experiment* - (coordinateur)

Référent scientifique : Raymond POLLARD (NOC Southampton)

3 Novembre 2004 - 21 Janvier 2005 / Océan Austral

Base de données : [https://www.bodc.ac.uk/data/bodc\\_database/nodb/data\\_collection/226/](https://www.bodc.ac.uk/data/bodc_database/nodb/data_collection/226/)

DOI : <http://doi.org/10.1016/j.dsr2.2007.06.011>

### **KEOPS (I et II)**

*KEOPS : KErguelen Ocean and Plateau compared Study - INSU-CNRS, IPEV and ANR*  
(financeurs)

Référent scientifique : Stéphane BLAIN (LOB/COM), Géraldine SARTHOU (IUEM/UBO)  
et Patrick LAAN (NIOZ)

Océan Austral - 11 Janvier au 23 Février 2005 (KEOPS I) et du 8 Octobre au 30 Novembre  
2011 (KEOPS II)

Site web : [http://keops2.obs-vlfr.fr/index.php?option=com\\_content&view=article&id=7&Itemid=12](http://keops2.obs-vlfr.fr/index.php?option=com_content&view=article&id=7&Itemid=12)

### **BOUM**

*BOUM : Biogeochemistry from the Oligotrophic to the Ultraoligotrophic Mediterranean - Institut Méditerranéen d'Océanologie* (coordinateur)

Référent scientifique : Thierry MOUTIN (MIO)

16/06/2008 (La Seyne-sur-Mer) - 20/07/2008 (La Seyne-sur-Mer)

Site web : <http://campagnes.flotteoceanographique.fr/campagnes/8010090/fr/>

Base de données : [http://www.obs-vlfr.fr/proof/php/x\\_datalist.php?xxop=boum&xxcamp=boum](http://www.obs-vlfr.fr/proof/php/x_datalist.php?xxop=boum&xxcamp=boum)

DOI : <http://dx.doi.org/10.17600/8010090>

### **HALIFAX**

Référent scientifique : Emmanuel DEVRED (Bedford Institute of Oceanography)

### **GEP&CO**

*GEP&CO : Geochemistry, Phytoplankton & Color of the Ocean - LODYC* (coordinateur)

Référent scientifique : Yves DANDONNEAU (IRD/LODYC)

La Havre - Nouméa *via* Panama - débuté en Octobre 1999 - 12 campagnes d'échantillonnage

Site web : [http://www.obs-vlfr.fr/proof/vt/op/ec/gep\\_co/gep\\_obj.htm](http://www.obs-vlfr.fr/proof/vt/op/ec/gep_co/gep_obj.htm)

# Annexe 9 : Nom des provinces de Longhurst

---

	<b>Abréviation</b>	<b>Type de biome <i>en anglais</i> - Nom complet <i>en anglais</i></b>
1	BPLR	Polar - Boreal Polar Province (POLR)
2	ARCT	Polar - Atlantic Arctic Province
3	SARC	Polar - Atlantic Subarctic Province
4	NADR	Westerlies - N. Atlantic Drift Province (WWDR)
5	GFST	Westerlies - Gulf Stream Province
6	NASW	Westerlies - N. Atlantic Subtropical Gyral Province (West) (STGW)
7	NATR	Trades - N. Atlantic Tropical Gyral Province (TRPG)
8	WTRA	Trades - Western Tropical Atlantic Province
9	ETRA	Trades - Eastern Tropical Atlantic Province
10	SATL	Trades - South Atlantic Gyral Province (SATG)
11	NECS	Coastal - NE Atlantic Shelves Province
12	CNRY	Coastal - Canary Coastal Province (EACB)
13	GUIN	Coastal - Guinea Current Coastal Province
14	GUIA	Coastal - Guianas Coastal Province
15	NWCS	Coastal - NW Atlantic Shelves Province
16	MEDI	Westerlies - Mediterranean Sea, Black Sea Province
17	CARB	Trades - Caribbean Province
18	NASE	Westerlies - N. Atlantic Subtropical Gyral Province (East) (STGE)
19	BRAZ	Coastal - Brazil Current Coastal Province
20	FKLD	Coastal - SW Atlantic Shelves Province
21	BENG	Coastal - Benguela Current Coastal Province
22	MONS	Trades - Indian Monsoon Gyres Province
23	ISSG	Trades - Indian S. Subtropical Gyre Province
24	EAFR	Coastal - E. Africa Coastal Province
25	REDS	Coastal - Red Sea, Persian Gulf Province

	<b>Abréviation</b>	<b>Type de biome <i>en anglais</i> - Nom complet <i>en anglais</i></b>
26	ARAB	Coastal - NW Arabian Upwelling Province
27	INDE	Coastal - E. India Coastal Province
28	INDW	Coastal - W. India Coastal Province
29	AUSW	Coastal - Australia-Indonesia Coastal Province
30	BERS	Polar - N. Pacific Epicontinental Province
31	PSAE	Westerlies - Pacific Subarctic Gyres Province (East)
32	PSAW	Westerlies - Pacific Subarctic Gyres Province (West)
33	KURO	Westerlies - Kuroshio Current Province
34	NPPF	Westerlies - N. Pacific Polar Front Province
35	NPSW	Westerlies - N. Pacific Subtropical Gyre Province (West)
36	TASM	Westerlies - Tasman Sea Province
37	SPSG	Westerlies - S. Pacific Subtropical Gyre Province
38	NPTG	Trades - N. Pacific Tropical Gyre Province
39	PNEC	Trades - N. Pacific Equatorial Countercurrent Province
40	PEQD	Trades - Pacific Equatorial Divergence Province
41	WARM	Trades - W. Pacific Warm Pool Province
42	ARCH	Trades - Archipelagic Deep Basins Province
43	ALSK	Coastal - Alaska Downwelling Coastal Province
44	CCAL	Coastal - California Upwelling Coastal Province
45	CAMR	Coastal - Central American Coastal Province
46	CHIL	Coastal - Chile-Peru Current Coastal Province
47	CHIN	Coastal - China Sea Coastal Province
48	SUND	Coastal - Sunda-Arafura Shelves Province
49	AUSE	Coastal - East Australian Coastal Province
50	NEWZ	Coastal - New Zealand Coastal Province
51	SSTC	Westerlies - S. Subtropical Convergence Province
52	SANT	Westerlies - Subantarctic Province
53	ANTA	Polar - Antarctic Province
54	APLR	Polar - Austral Polar Province

# Annexe 10 : Cartes de distribution des fréquences de détection mensuelles des assemblages v2017 en climatologie (PHYSAT version 2017, SeaWiFS, de 1998 à 2010)

---

Carte de distribution des fréquences de détection mensuelles en climatologie (PHYSAT version 2017, SeaWiFS, de 1998 à 2010) des classes de  $Ra(\lambda)$  labellisées (nom des assemblages sous les Figures suivantes) à partir de la calibration *in-situ* (Section 4.1.3, page 104).



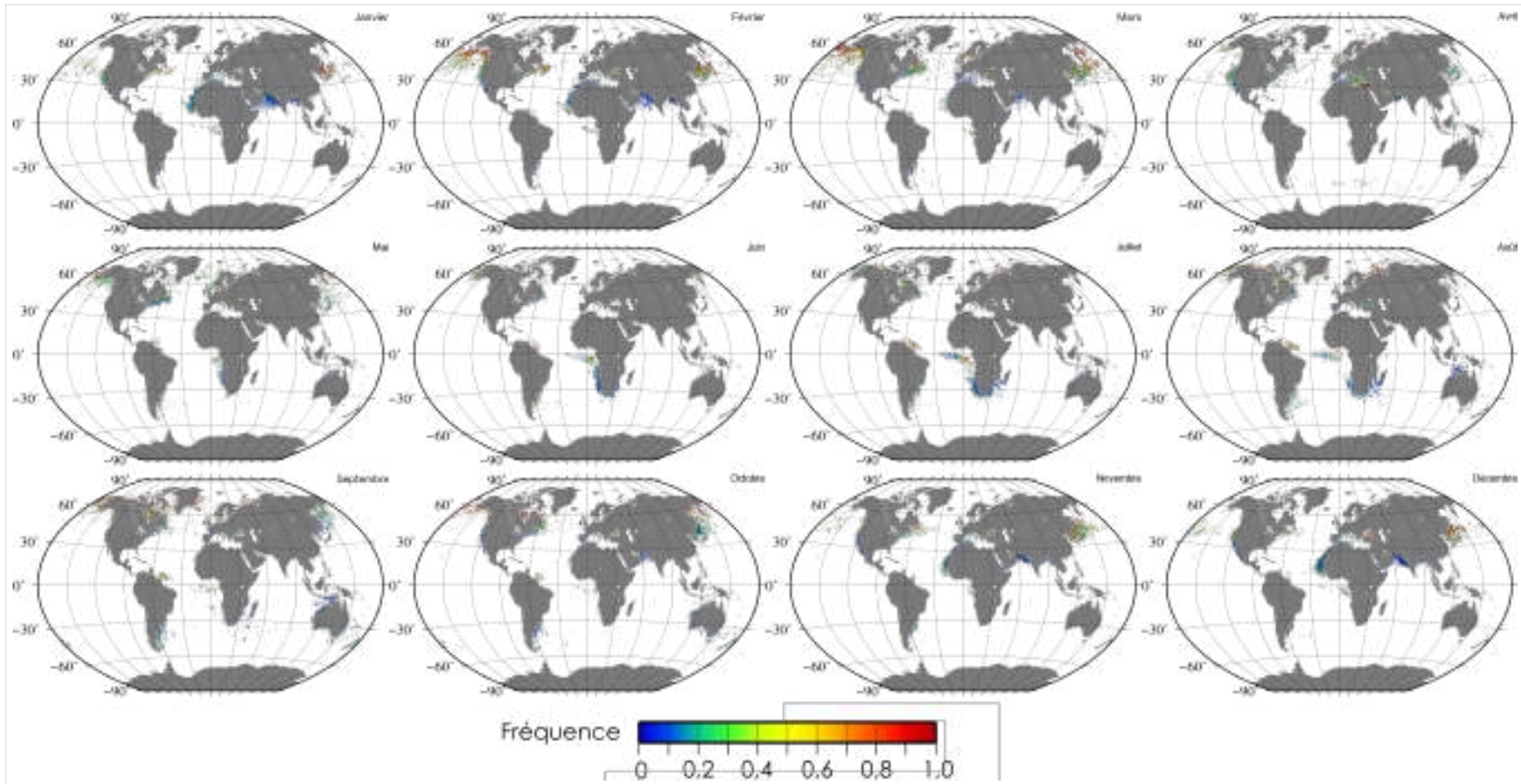


FIGURE 1, ANNEXE 10 – Distribution des fréquences de détection mensuelles des nanoeucaryotes dominants (v2017, climatologie 1998-2010, SeaWiFS).

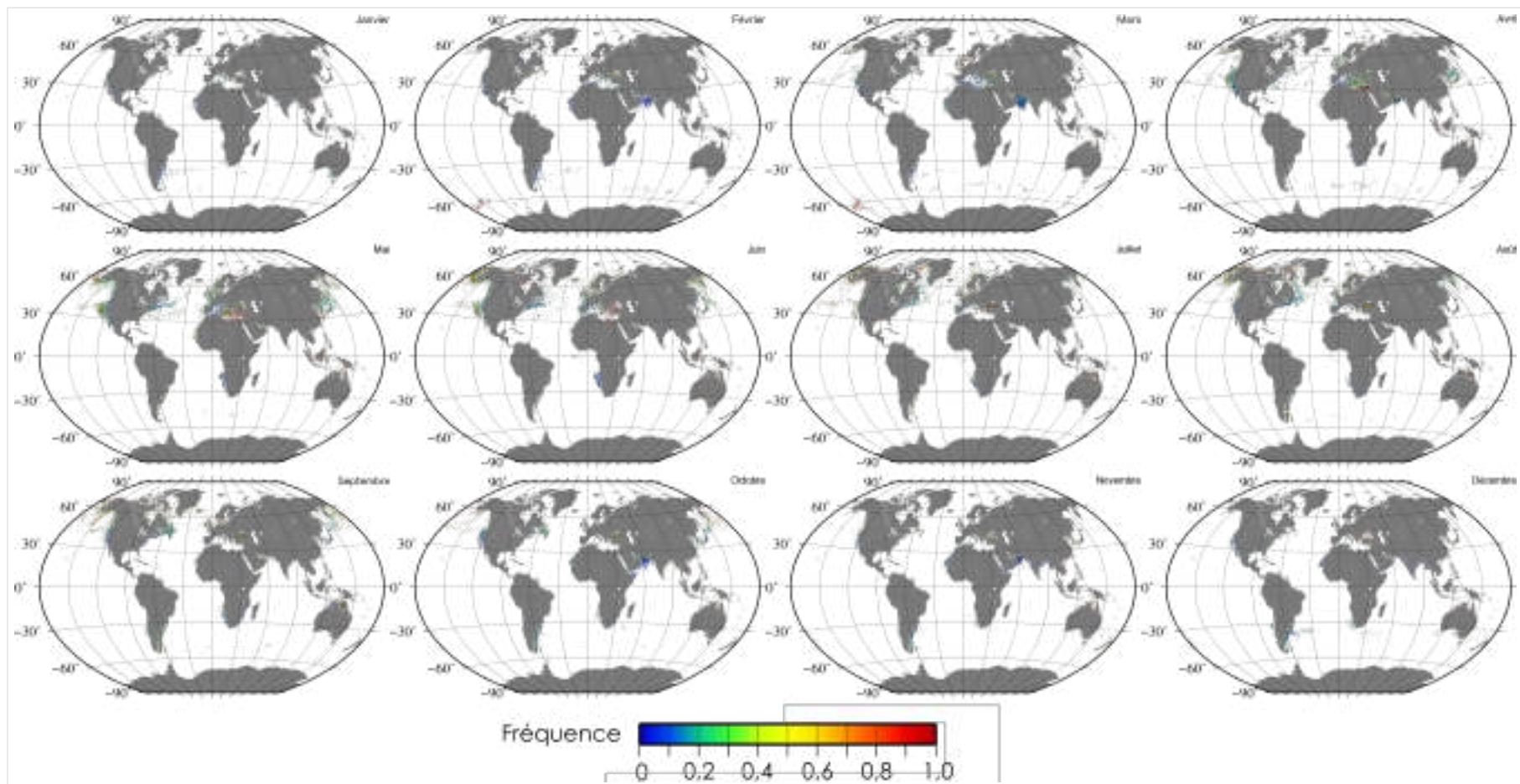


FIGURE 2, ANNEXE 10 – Distribution des fréquences de détection mensuelles des diatomées dominantes (v2017, climatologie 1998-2010, SeaWiFS).

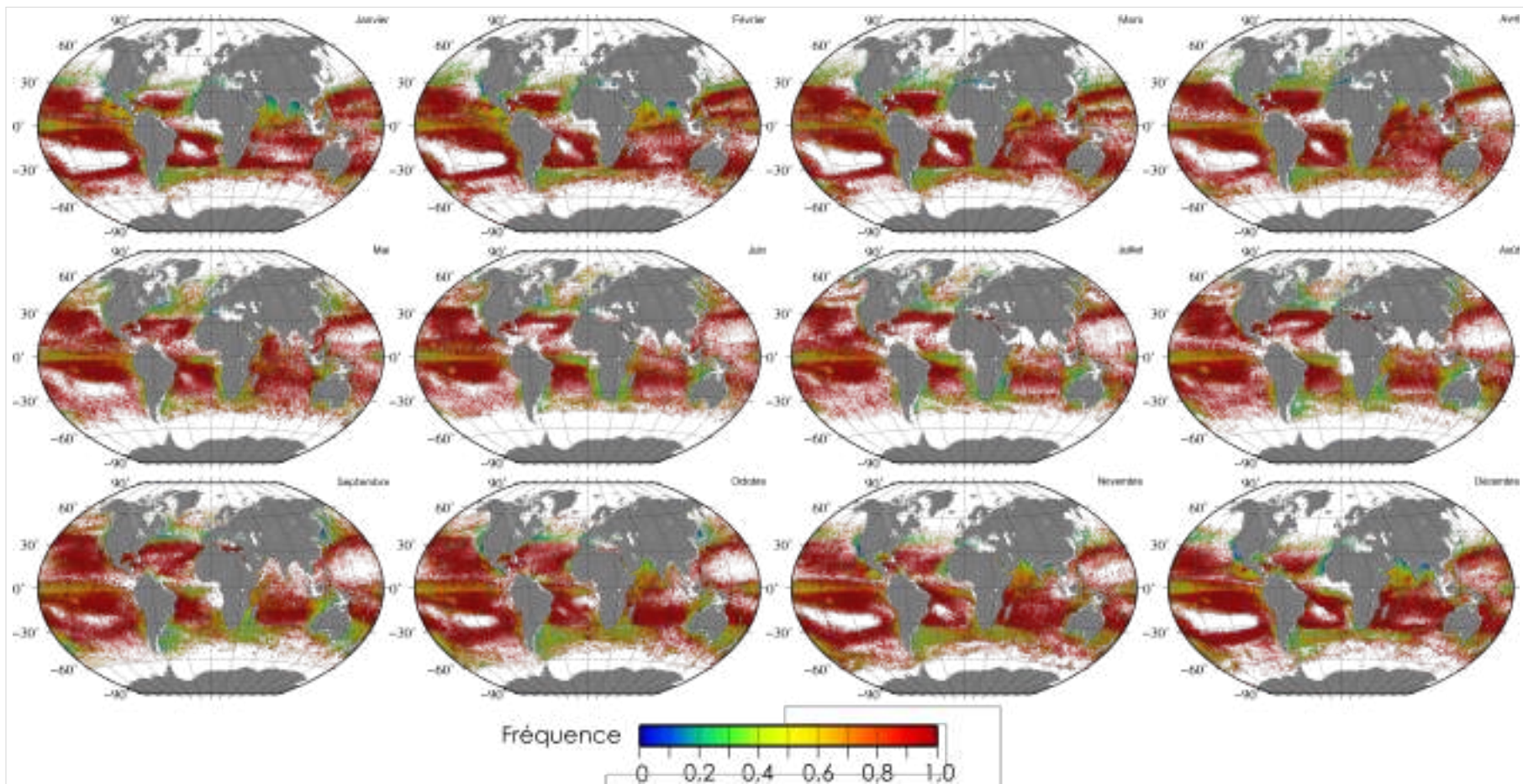


FIGURE 3, ANNEXE 10 – Distribution des fréquences de détection mensuelles des *Prochlorococcus* spp. non dominants (v2017, climatologie 1998-2010, SeaWiFS).

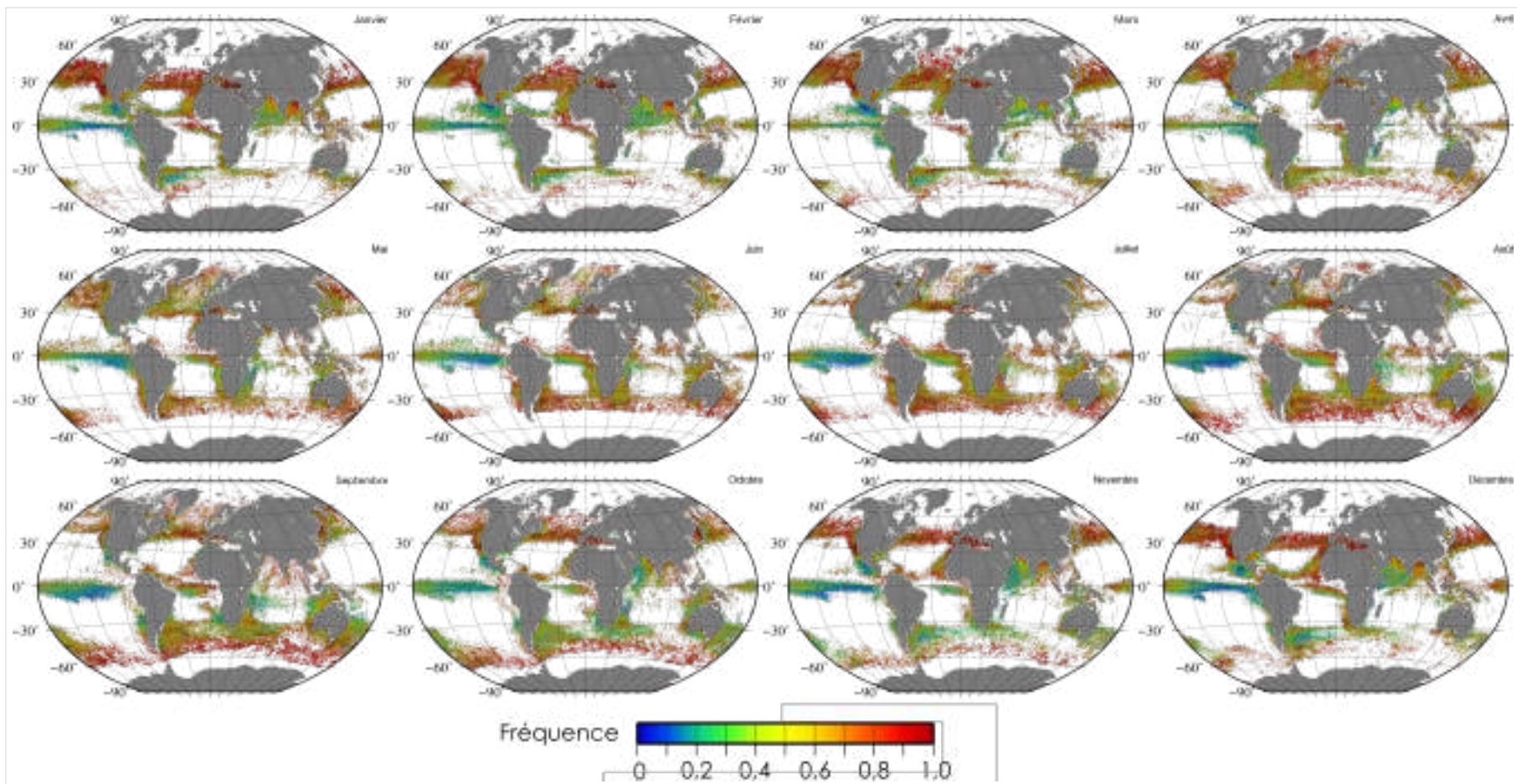


FIGURE 4, ANNEXE 10 – Distribution des fréquences de détection mensuelles de l’assemblage de diatomées et nanoeucaryotes (v2017, climatologie 1998-2010, SeaWiFS).

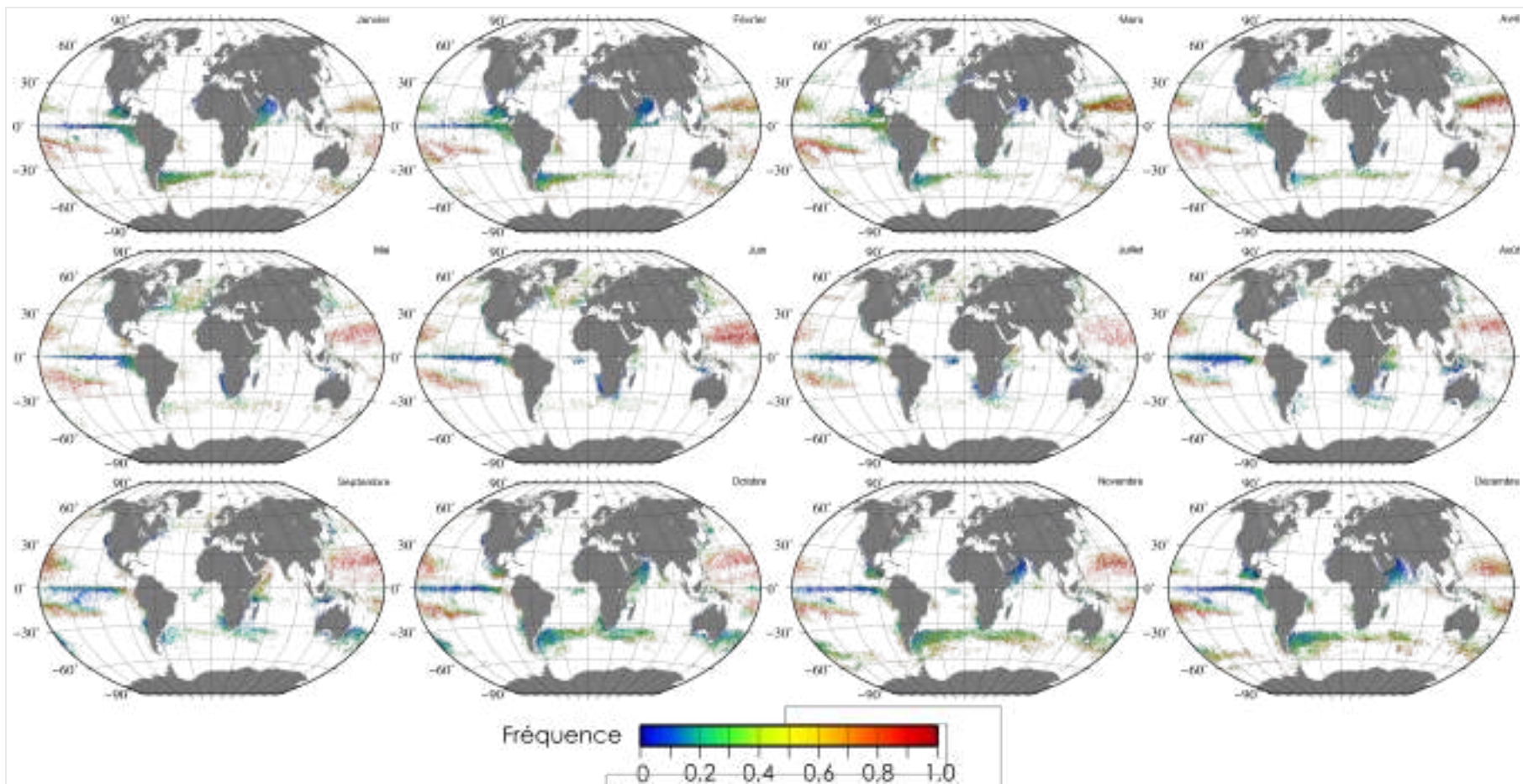


FIGURE 5, ANNEXE 10 – Distribution des fréquences de détection mensuelles de l’assemblage de *Synechococcus spp.-like* et *Prochlorococcus spp.* (v2017, climatologie 1998-2010, SeaWiFS).

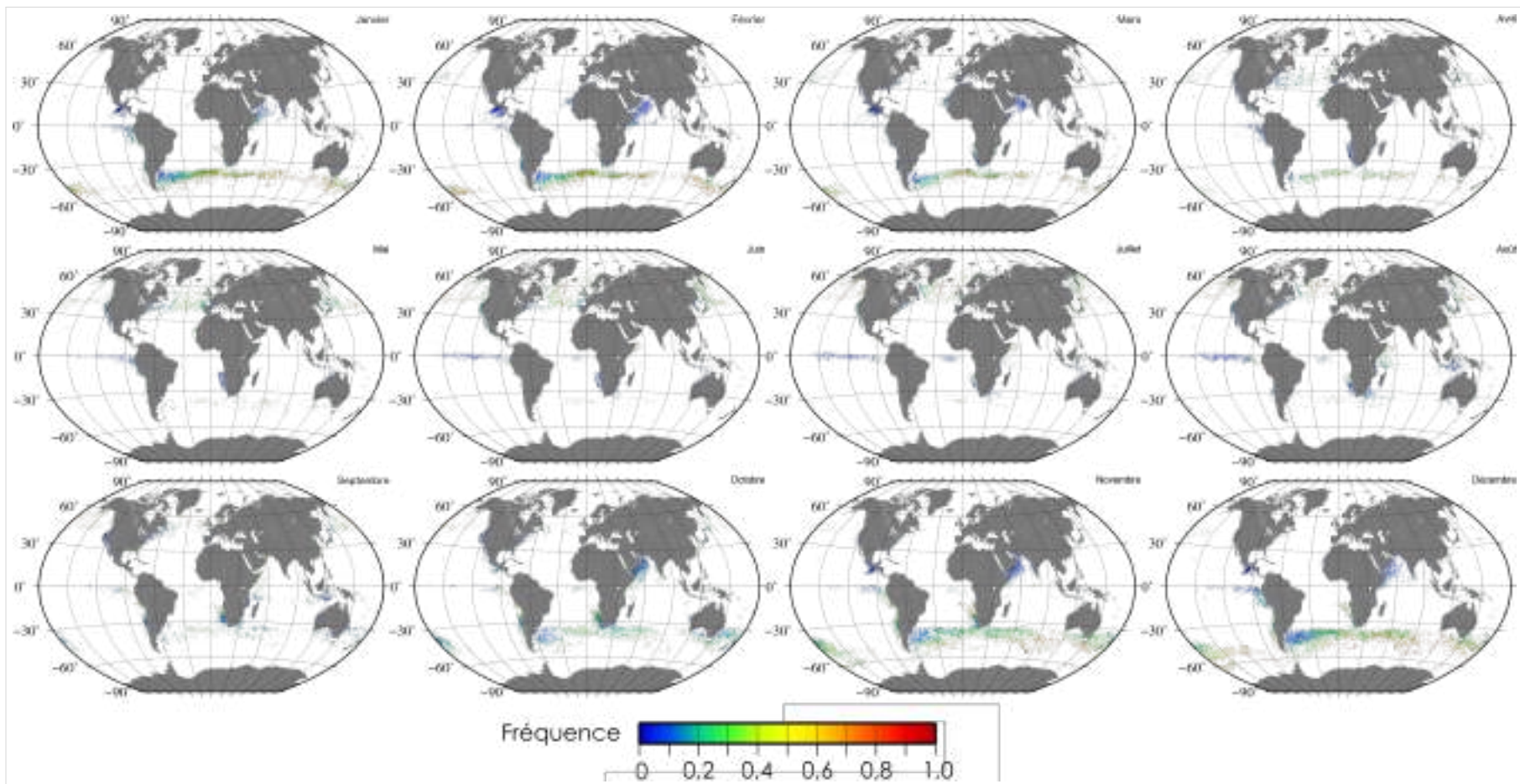


FIGURE 6, ANNEXE 10 – Distribution des fréquences de détection mensuelles des dinoflagellés dominants (v2017, climatologie 1998-2010, SeaWiFS).



# Annexe 11 : Cartes de distribution des fréquences de détection mensuelles des assemblages vMixte en climatologie (PHYSAT version Mixte (2008-2017), SeaWiFS, de 1998 à 2010)

---

Carte de distribution des fréquences de détection mensuelles en climatologie (PHYSAT version Mixte, SeaWiFS, de 1998 à 2010) des classes de  $Ra(\lambda)$  labellisées (nom des assemblages sous les Figures suivantes) à partir de la calibration *in-situ*, ou de la projection des  $Ra(\lambda)$  sur la carte auto-organisatrice (Section 4.1.3, page 104).



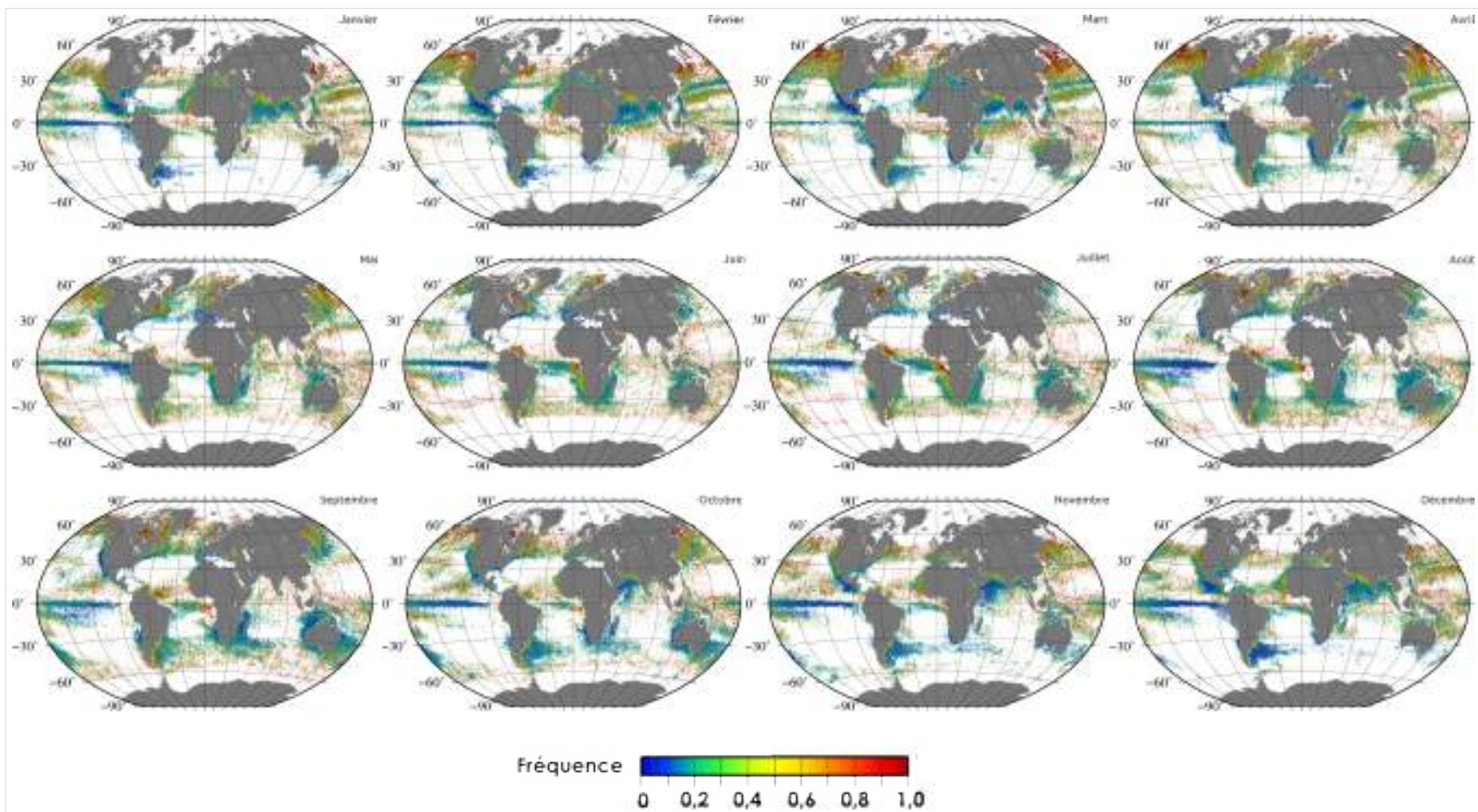


FIGURE 1, ANNEXE 11 – Distribution des fréquences de détection mensuelles des nanoeucaryotes dominants (vMixte, climatologie 1998-2010, SeaWiFS).

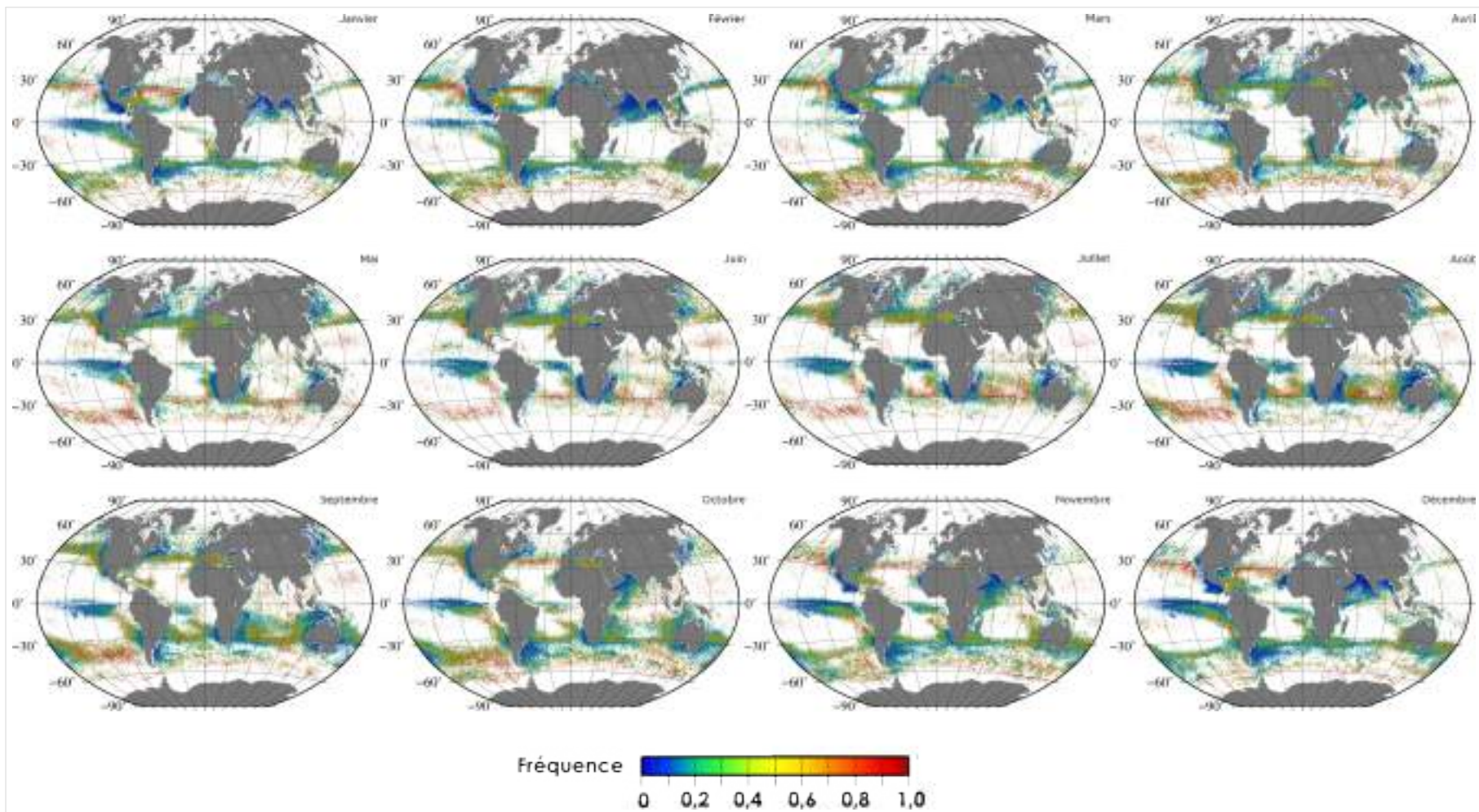


FIGURE 2, ANNEXE 11 – Distribution des fréquences de détection mensuelles des *Prochlorococcus* spp. dominants (vMixte, climatologie 1998-2010, SeaWiFS).

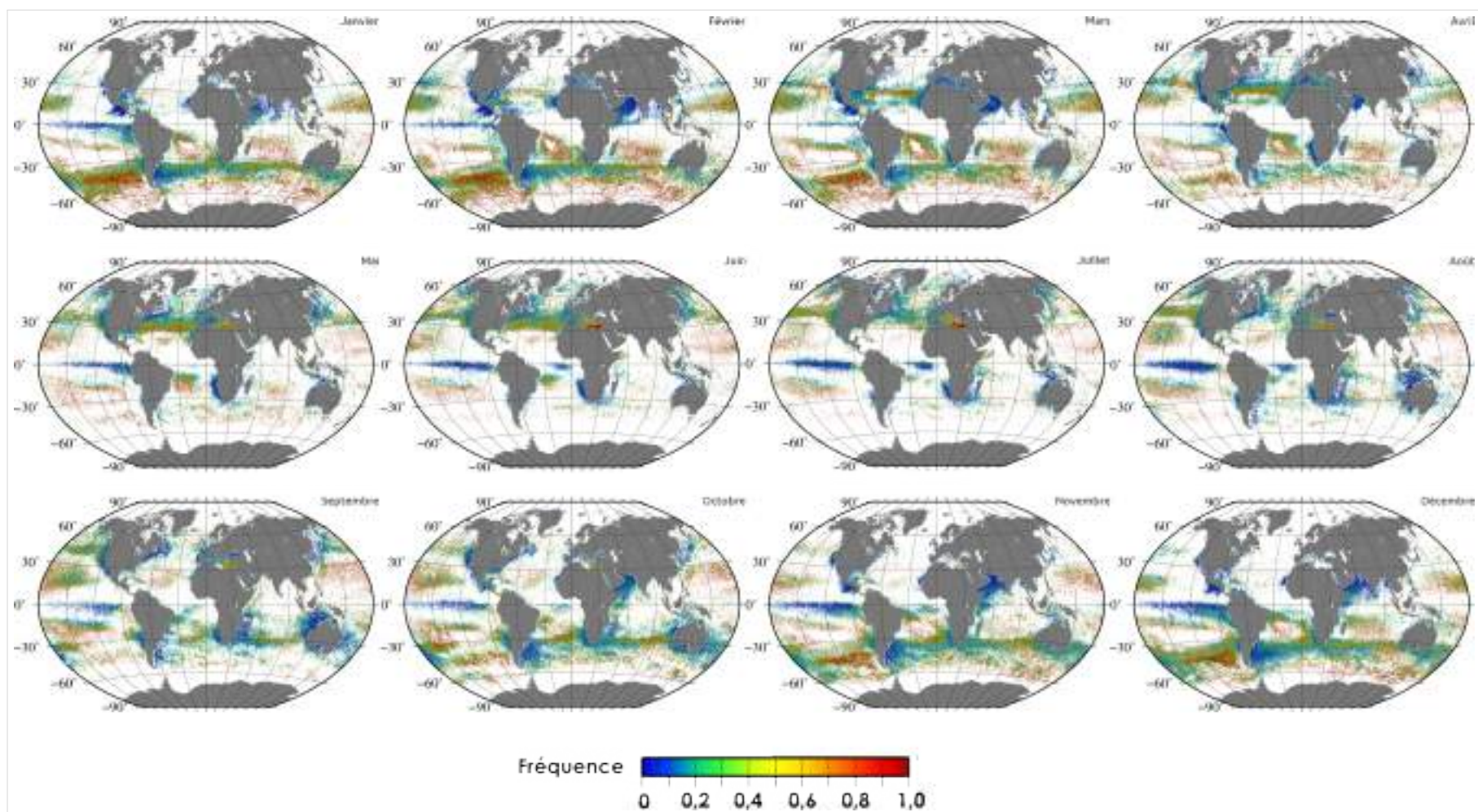


FIGURE 3, ANNEXE 11 – Distribution des fréquences de détection mensuelles des *Synechococcus spp.-like* dominants (vMixte, climatologie 1998-2010, SeaWiFS).

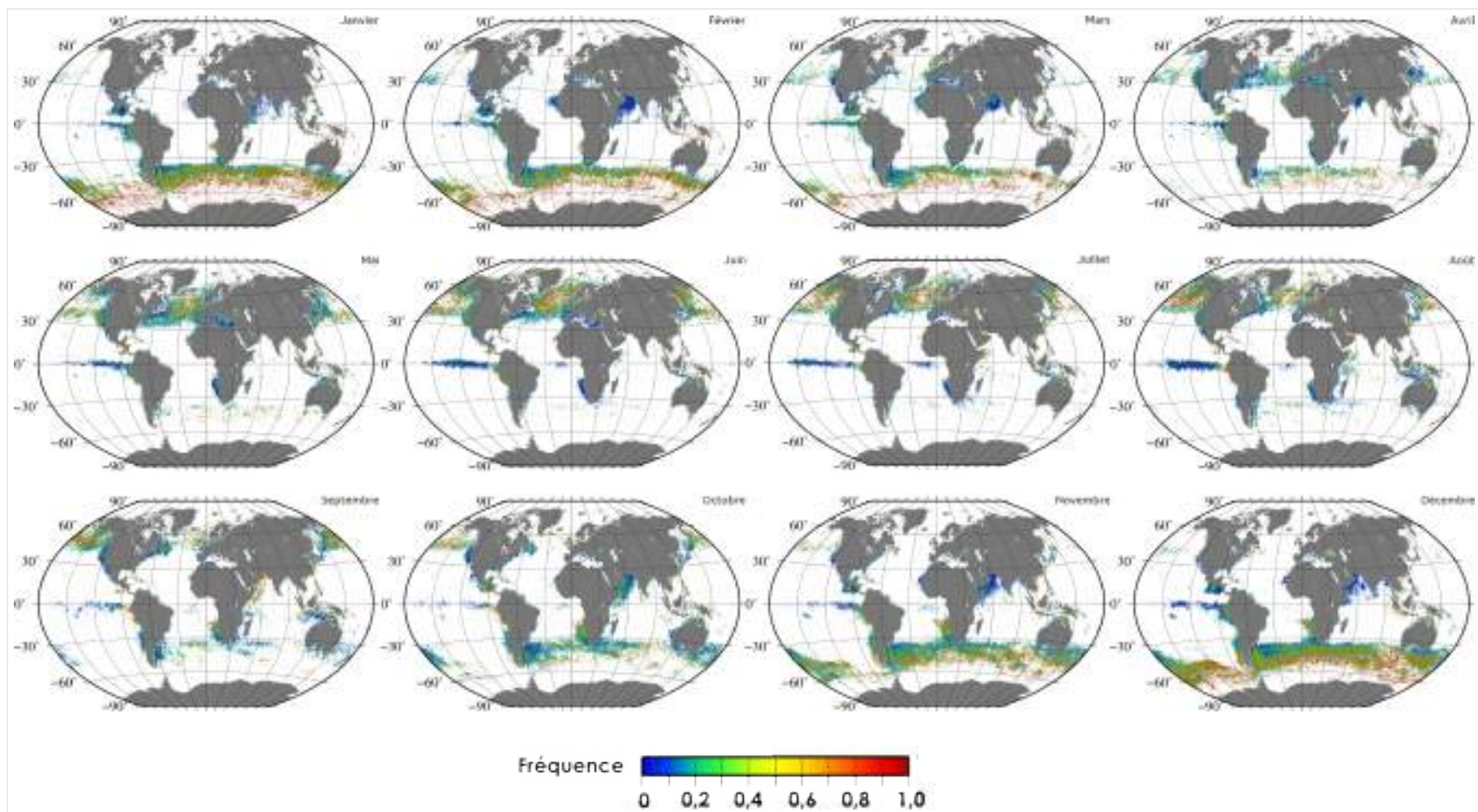


FIGURE 4, ANNEXE 11 – Distribution des fréquences de détection mensuelles des diatomées dominantes (vMixte, climatologie 1998-2010, SeaWiFS).

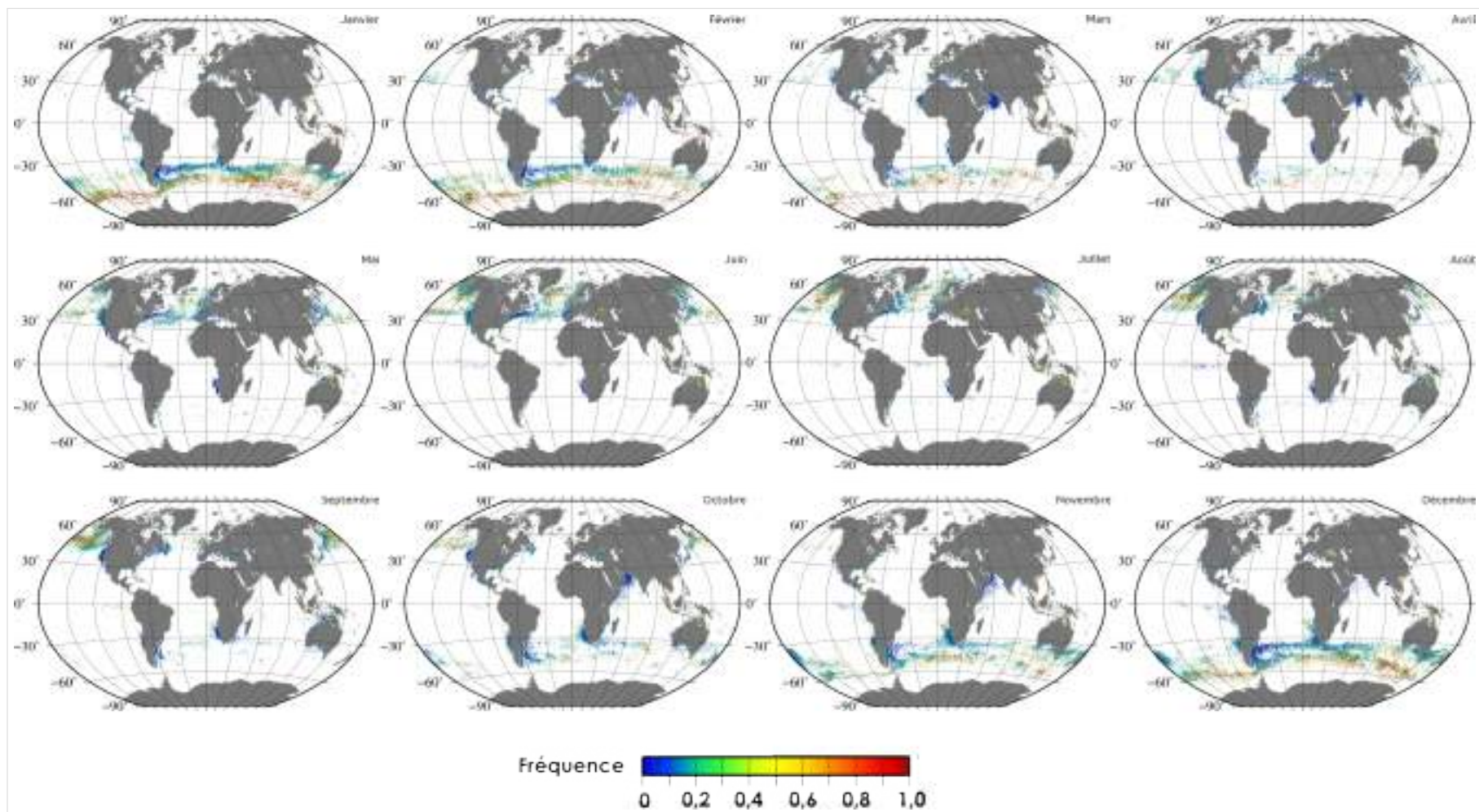


FIGURE 5, ANNEXE 11 – Distribution des fréquences de détection mensuelles des *Phaeocystis spp.-like* (vMixte, climatologie 1998-2010, SeaWiFS).

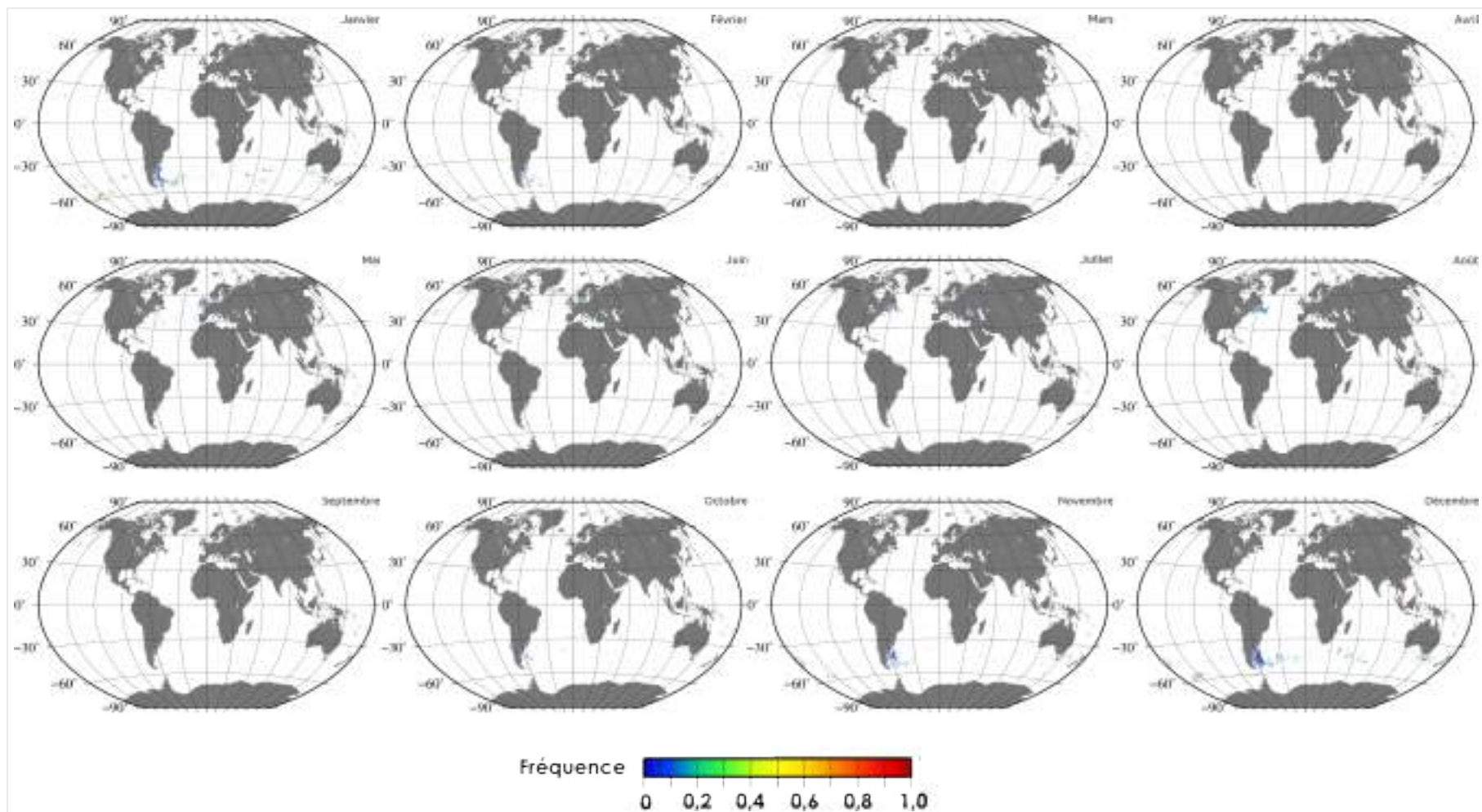


FIGURE 6, ANNEXE 11 – Distribution des fréquences de détection mensuelles des coccolithophoridés (vMixte, climatologie 1998-2010, SeaWiFS).

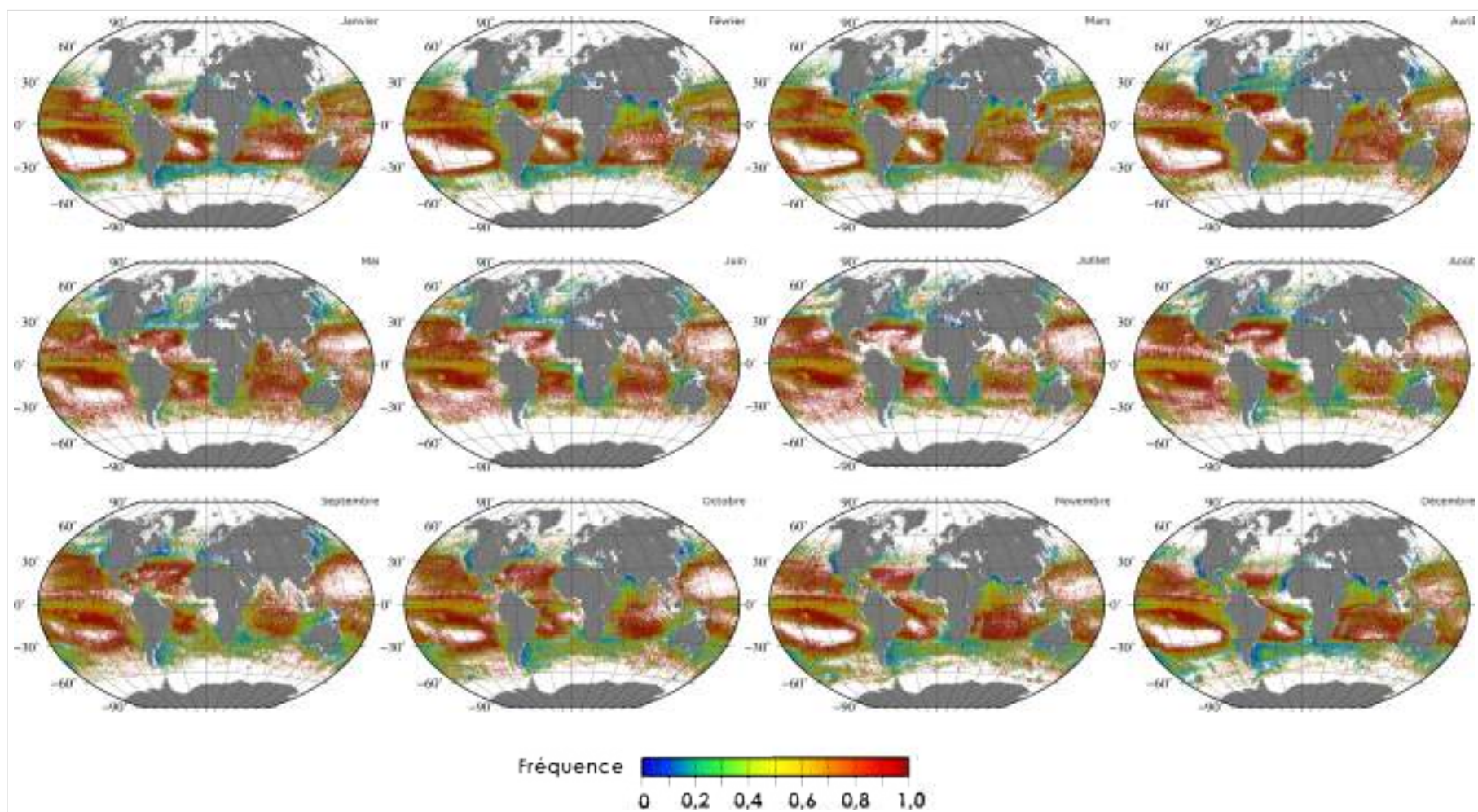


FIGURE 7, ANNEXE 11 – Distribution des fréquences de détection mensuelles des *Prochlorococcus spp.* non dominants (vMixte, climatologie 1998-2010, SeaWiFS).

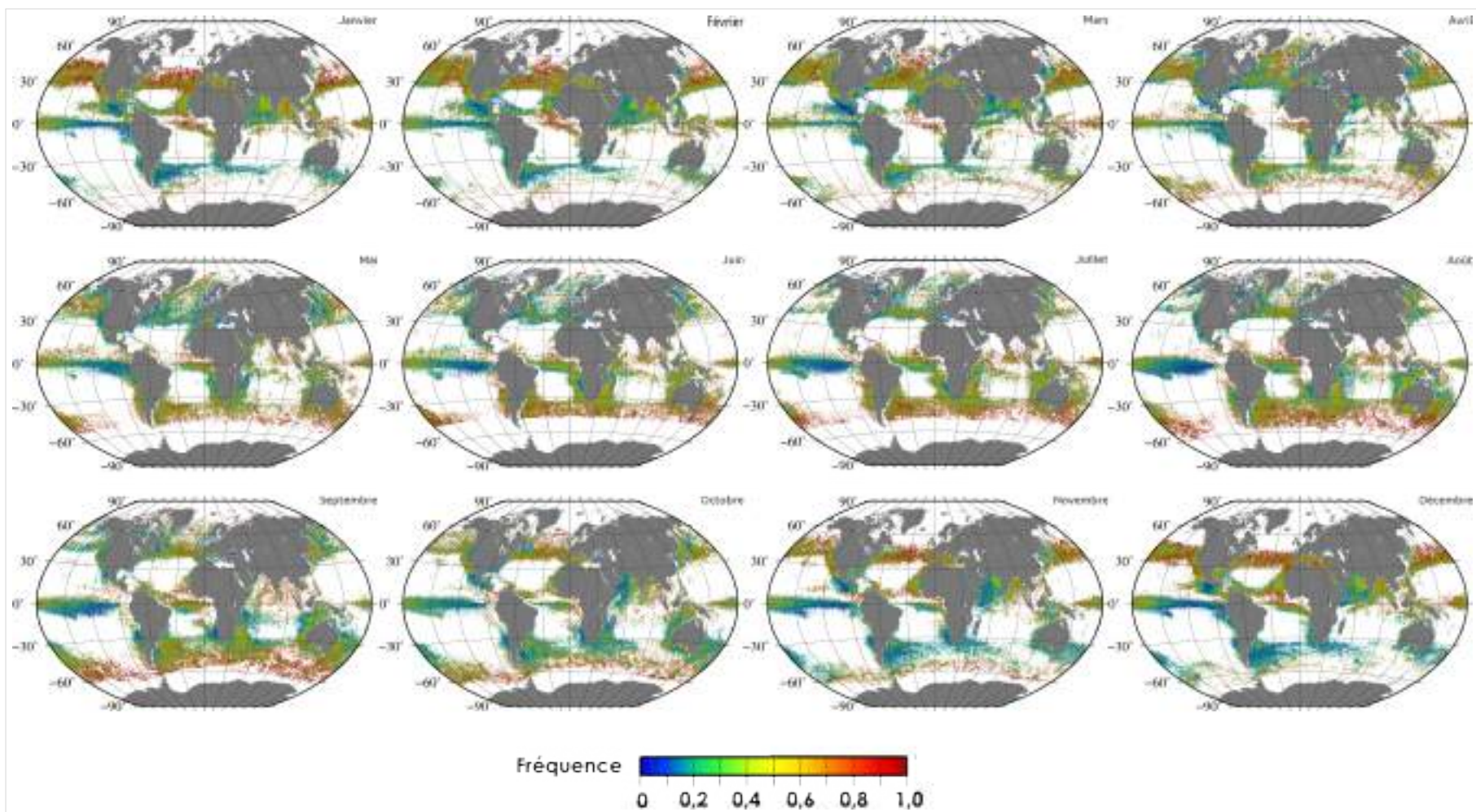


FIGURE 8, ANNEXE 11 – Distribution des fréquences de détection mensuelles de l’assemblage de diatomées et de nanoeucaryotes (vMixte, climatologie 1998-2010, SeaWiFS).



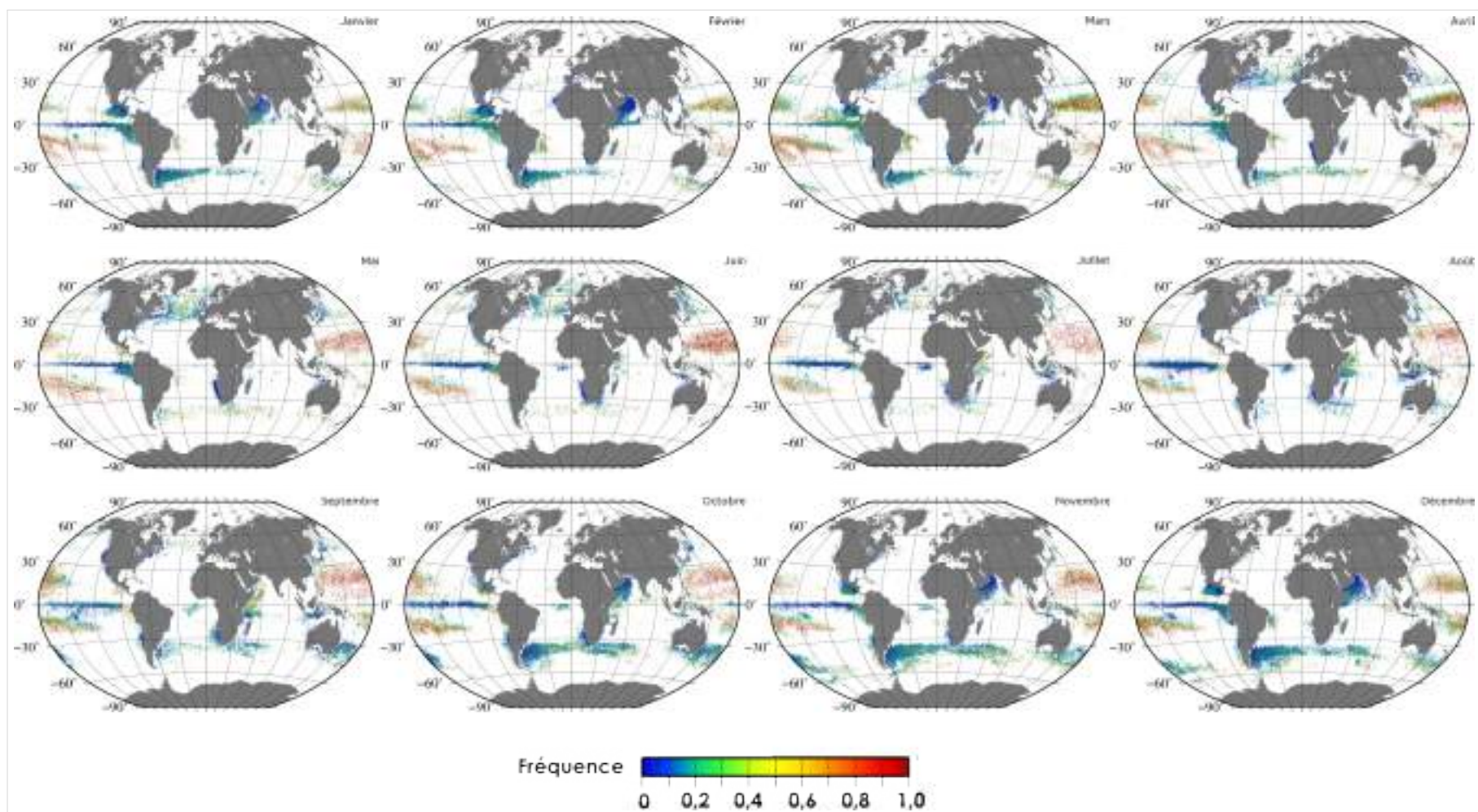


FIGURE 9, ANNEXE 11 – Distribution des fréquences de détection mensuelles de l'assemblage de *Synechococcus* spp.-like et *Prochlorococcus* spp. (vMixte, climatologie 1998-2010, SeaWiFS).

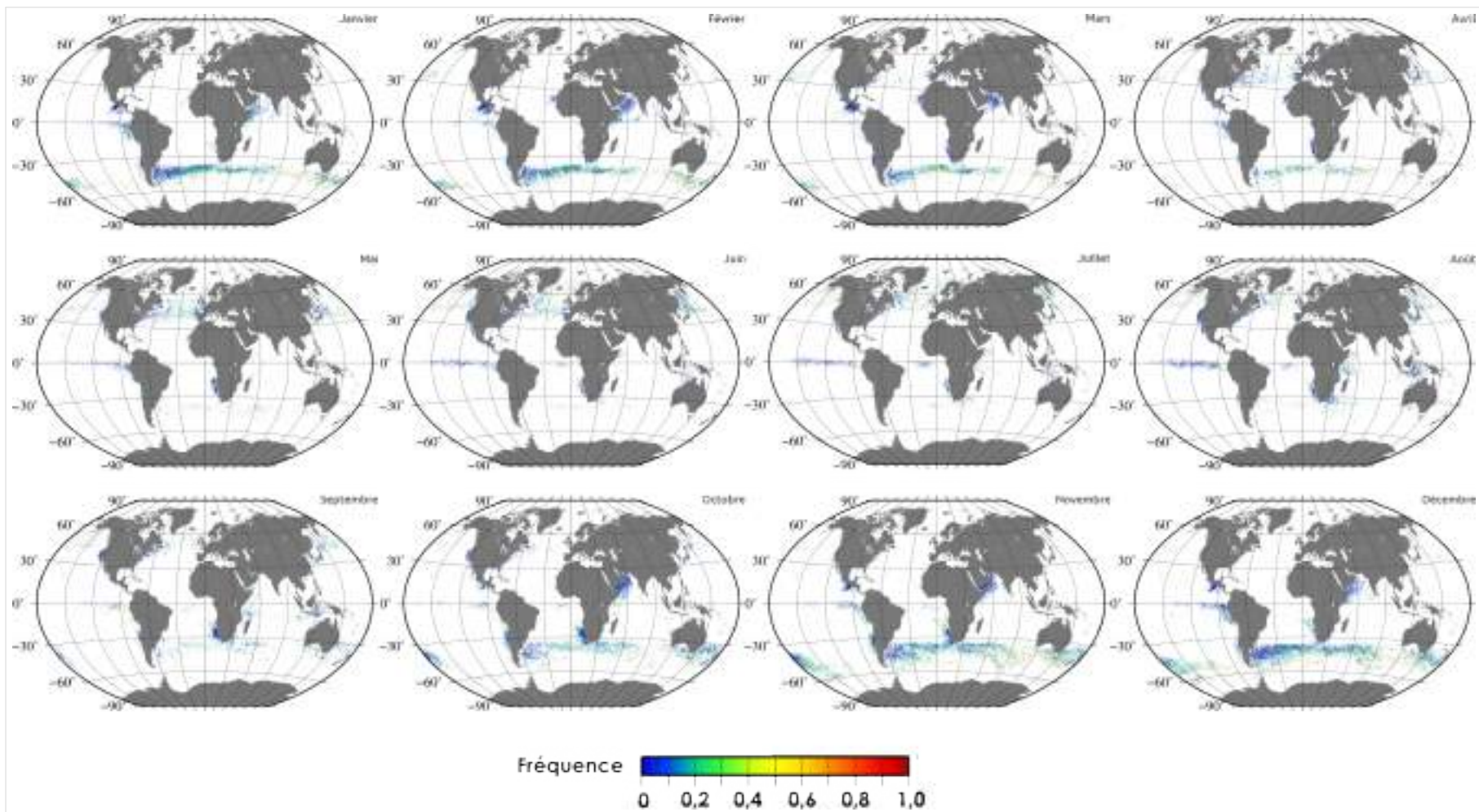


FIGURE 10, ANNEXE 11 – Distribution des fréquences de détection mensuelles des dinoflagellés dominants (vMixte, climatologie 1998-2010, SeaWiFS).



**Annexe 12 : *Estimation of the potential detection of diatom assemblages based on ocean color radiance anomalies in the North Sea - Frontiers in Marine Science, 2017***

---

A-H. Rêve, S. Alvain, M-F. Racault, D. Dessailly, N. Guiselin, C. Jamet, V. Vantrepotte, G. Beaugrand - **Estimation of the potential detection of diatom assemblages based on ocean color radiance anomalies in the North Sea** *Frontiers in Marine Science - Ocean Observation, Submitted on : 02 Aug 2017, accepted 29 Nov 2017.*



# Estimation of the Potential Detection of Diatom Assemblages Based on Ocean Color Radiance Anomalies in the North Sea

Anne-Hélène Réve-Lamarche<sup>1</sup>, Séverine Alvain<sup>1\*</sup>, Marie-Fanny Racault<sup>2</sup>, David Dessailly<sup>3</sup>, Natacha Guiselin<sup>3</sup>, Cédric Jamet<sup>3</sup>, Vincent Vantrepotte<sup>4</sup> and Grégory Beaugrand<sup>4</sup>

<sup>1</sup> CNRS, Univ. Lille, Univ. Littoral Côte d'Opale, UMR 8187, LOG, Laboratoire d'Océanologie et de Géosciences, Lille, France, <sup>2</sup> Plymouth Marine Laboratory, Plymouth, United Kingdom, <sup>3</sup> Univ. Littoral Côte d'Opale, Univ. Lille, CNRS, UMR 8187, LOG, Laboratoire d'Océanologie et de Géosciences, Wimereux, France, <sup>4</sup> CNRS, Univ. Lille, Univ. Littoral Côte d'Opale, UMR 8187, LOG, Laboratoire d'Océanologie et de Géosciences, Wimereux, France

## OPEN ACCESS

### Edited by:

Chris Bowler,  
École Normale Supérieure, Université  
de Sciences Lettres de Paris, France

### Reviewed by:

Sonia Batten,  
Sir Alister Hardy Foundation for Ocean  
Science, United Kingdom  
Alex C. Kraberg,  
Alfred Wegener Institut Helmholtz  
Zentrum für Polar und  
Meeresforschung (AWI), Germany

### \*Correspondence:

Séverine Alvain  
severine.alvain@cnrs.fr

### Specialty section:

This article was submitted to  
Ocean Observation,  
a section of the journal  
Frontiers in Marine Science

Received: 02 August 2017

Accepted: 29 November 2017

Published: 12 December 2017

### Citation:

Réve-Lamarche A-H, Alvain S,  
Racault M-F, Dessailly D, Guiselin N,  
Jamet C, Vantrepotte V and  
Beaugrand G (2017) Estimation of the  
Potential Detection of Diatom  
Assemblages Based on Ocean Color  
Radiance Anomalies in the North Sea.  
Front. Mar. Sci. 4:408.  
doi: 10.3389/fmars.2017.00408

Over the past years, a large number of new approaches in the domain of ocean-color have been developed, leading to a variety of innovative descriptors for phytoplankton communities. One of these methods, named PHYSAT, currently allows for the qualitative detection of five main phytoplankton groups from ocean-color measurements. Even though PHYSAT products are widely used in various applications and projects, the approach is limited by the fact it identifies only dominant phytoplankton groups. This current limitation is due to the use of biomarker pigment ratios for establishing empirical relationships between *in-situ* information and specific ocean-color radiance anomalies in open ocean waters. However, theoretical explanations of PHYSAT suggests that it could be possible to detect more than dominance cases but move more toward phytoplanktonic assemblage detection. Thus, to evaluate the potential of PHYSAT for the detection of phytoplankton assemblages, we took advantage of the Continuous Plankton Recorder (CPR) survey, collected in both the English Channel and the North Sea. The available CPR dataset contains information on diatom abundance in two large areas of the North Sea for the period 1998–2010. Using this unique dataset, recurrent diatom assemblages were retrieved based on classification of CPR samples. Six diatom assemblages were identified *in-situ*, each having indicators taxa or species. Once this first step was completed, the *in-situ* analysis was used to empirically associate the diatom assemblages with specific PHYSAT spectral anomalies. This step was facilitated by the use of previous classifications of regional radiance anomalies in terms of shape and amplitude, coupled with phenological tools. Through a matchup exercise, three CPR assemblages were associated with specific radiance anomalies. The maps of detection of these specific radiance anomalies are in close agreement with current *in-situ* ecological knowledge.

**Keywords:** ocean-color, diatom assemblages, North Sea, CPR, PHYSAT

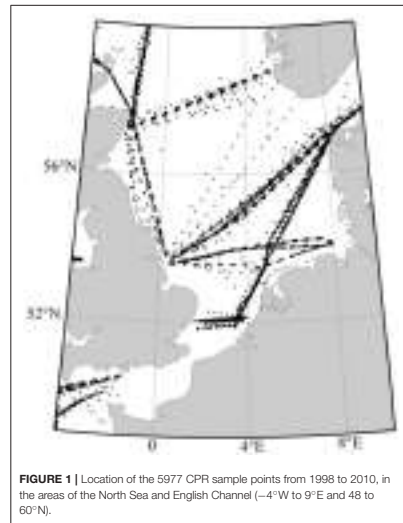
## 1. INTRODUCTION

Phytoplankton play a key-role in oceanic biogeochemical cycles (Falkowski, 1994; Beaugrand, 2015). For example, by fixing inorganic carbon through photosynthesis, phytoplankton act as a carbon biological pump by exporting carbon to the deep ocean (Kump et al., 2010). Moreover, phytoplankton is at the base of the marine food web—transferring carbon and energy from grazers to higher trophic levels. Phytoplankton species scatter and absorb light differently according to their concentration, pigments composition, size, morphology, intracellular structure, cells arrangement, and associated dissolved organic matter (Morel and Bricaud, 1981; Bricaud and Morel, 1986; Dubelaar et al., 1987; Stramski and Kiefer, 1991; Siegel et al., 2005; Clavano et al., 2007; Boss et al., 2009; Whitmire et al., 2010). Optical properties can be studied from *in-situ* and remote-sensing measurements, and are currently used in ocean-color applications to investigate phytoplankton in surface waters. Thus, chlorophyll-*a* pigment concentration (Chl-*a* hereafter), considered as a proxy for phytoplankton biomass (e.g., Kirk, 1994), has now been estimated from space for more than 30 years (O'Reilly et al., 1998). More recently, using particulate organic carbon, or chlorophyll-Chromophoric Dissolved Organic Matter (CDOM), remote-sensing products have been released (e.g., Stramski et al., 2008; Dufoiré-Gaurier et al., 2010). These ocean-color products provide medium-resolution and synoptic scale data that are required to study the main characteristics of phytoplankton distribution.

Beyond phytoplankton concentration estimation, an accurate knowledge of phytoplankton community composition and distribution is important to understand the diversity, structure and functioning of marine food-web, and associated ecosystem services (Nair et al., 2008; Sathyendranath et al., 2014). Phytoplankton Functional Types (PFTs) are defined as phytoplankton species with similar functions in terms of geochemical roles and physiological traits (Hood et al., 2006). Several PFT classifications have been proposed, depending on scientific interest (Nair et al., 2008; Sathyendranath et al., 2014). For example, cell size is considered as a first level approach to define PFTs (Nair et al., 2008). These size classes are useful in understanding biochemical functions such as nutrient uptake efficiencies in relation to surface-area-to-volume-ratio (Platt and Jassby, 1976). However, a size-based approach would not be suitable when phytoplankton characterized by different functions fall under the same size class. For example, dimethyl sulphide (DMS) producers and calcifiers are often grouped in the nanophytoplankton size class (Nair et al., 2008). Another approach to define PFTs is to classify phytoplankton according to their biogeochemical functions (calcifiers, silicifiers, nitrogen-fixers, DMS producers) (Le Quééré et al., 2005). In the last decade, ocean-color derived products have been developed to allow for the detection of phytoplankton communities in surface waters (Nair et al., 2008; Sathyendranath et al., 2014; Mouw et al., 2017). Some remote-sensing methods are able to detect phytoplankton size classes (e.g., Uitz et al., 2006; Mouw and Yoder, 2010; Brewin et al., 2011; Devred et al., 2011; Bricaud et al., 2012; Li et al., 2013 or particle size distribution, Kostadinov et al., 2009,

2010), giving a global distribution of phytoplankton size in the ocean. In addition, several algorithms have been developed to identify: (i) one specific PFT from space (e.g., Smyth et al., 2002; Subramaniam et al., 2002; Sathyendranath et al., 2004), and (ii) several PFTs (e.g., Aiken et al., 2007; Alvain et al., 2008; Raitos et al., 2008; Bracher et al., 2009; Hirata et al., 2011; Sadeghi et al., 2012). A review of the different PFTs detection methods, based either on direct analysis of remote-sensing data or on empirical and semi-empirical approaches, can be found in recent papers (Nair et al., 2008; Brewin et al., 2011; Sathyendranath et al., 2014; Bracher et al., 2015b; Mouw et al., 2017).

Among the available approaches, PHYSAT was developed to detect phytoplankton groups on a global scale by using radiance anomalies (Alvain et al., 2008). This method retrieves the empirical labeling of specific ocean-color radiance anomalies based on *in-situ* measurements. During its first implementation, four phytoplankton groups were detected, when dominant, based on biomarker *in-situ* pigment inventories (Alvain et al., 2005). These four groups were detected in global oceanic waters, from prior analysis of their biomarker pigments: diatoms, nanoeucaryotes, *Synechococcus*-like, and *Prochlorococcus*. Even though PHYSAT maps have been used in a large variety of applications (e.g., D'Ovidio et al., 2010; De Monte et al., 2013; Navarro et al., 2014; Thyssen et al., 2015), its limitation could hold back future developments. In fact, to date, the algorithm only detects dominant phytoplankton groups. This limitation is due to the use of ratios on biomarker pigment concentrations during the empirical calibration steps of the PHYSAT algorithm, which do not provide information about phytoplankton assemblages (Alvain et al., 2005, 2008). However, the study of Alvain et al. (2012) showed that characteristics of radiance anomalies are explained theoretically by phytoplankton characteristics such as cell size, composition, intracellular structure, cells arrangement, and absorption. This study made the suggestion that specific radiance anomalies could be empirically extended beyond dominance cases. Thus, analysis restricted to "dominant" groups based on biomarker pigments may represent an under-exploitation of radiance anomalies database, which could also contain a large diversity of information. Other types of *in-situ* data about phytoplankton are common, such as the Continuous Plankton Recorder (CPR thereafter, SAHFOS, 2015). In this context, the possibility to enhance the radiance anomaly labeling capabilities, when *in-situ* information about specific phytoplankton abundance are available, was evaluated. For this purpose, the CPR diatom database from the North Sea and the English Channel were used (Figure 1). Broadly, diatoms represent a dominant taxon in the North Sea, at least during part of the year, with an abundance contribution from 40 to 90% of total microphytoplankton (Reid et al., 1990; Leterme et al., 2006). Please note that here the term "dominant" has not the same meaning that the one used in PHYSAT (see definition n° 6 in Supplementary Material). Although North Sea waters are not only characterized by diatoms, the fairly detailed information on diatom species and taxa provides a good starting point in this evaluation. However, a new approach was necessary to promote the use of abundance data to empirically associate radiance anomalies with specific



**FIGURE 1** | Location of the 5977 CPR sample points from 1998 to 2010, in the areas of the North Sea and English Channel (−4°W to 9°E and 48 to 60°N).

phytoplankton assemblages *in-situ*; and therefore, enable future improvements of PHYSAT.

In this paper, a method was designed to label radiance anomalies ( $Ra(\lambda)$  thereafter) according to *in-situ* information provided by the CPR survey. The goal was to answer the following question: is there an empirical association between the presence of a diatom assemblage and specific radiance anomaly signals (as supported by the theoretical study of Alvain et al., 2012)? First, tools were developed and applied to extract suitable ecological information (i.e., diatom assemblages) from the CPR *in-situ* database (section 2.4.2). Second, the  $Ra(\lambda)$  was classified according to their form and amplitude (Ben Mustapha et al., 2014), and to their phenological characteristics (section 2.2). The objective of this second step was to fully characterize the  $Ra(\lambda)$  ranges in terms of shapes, amplitudes, and phenological behaviors in order to increase the chances of finding an association between specific *in-situ* diatom assemblages and a specific signal from remote-sensing. Finally, simultaneous *in-situ* and remote-sensing measurements were used to evaluate the relationships between the presence of diatom assemblages and the detection of specific radiance anomalies (sections 2.4.3 and 2.4.4).

## 2. MATERIALS AND METHODS

**Figure 2** displays the diagram of the process followed in this study. A list of definitions is available in the Supplementary

Material (Supplementary Material Section 1, Definitions). These words are number indexed.

The method used in this paper is based on the PHYSAT approach. Hence, a description of the general principle of the PHYSAT (Alvain et al., 2005, 2008) is given first, followed then by the actual process applied in this study.

### 2.1. General Principle of PHYSAT

The PHYSAT method is based on remote-sensed normalized water leaving radiances anomalies ( $Ra(\lambda)$ )<sup>15</sup>. In fact, for a given chlorophyll-*a* concentration, it is possible to compute the mean value of radiances at each wavelength for the entire globe (2005 and 2014 versions) or specific regions (in present study, the North Sea was used). However, when considering one given pixel, associated with a chlorophyll-*a* concentration, the corresponding radiance values are slightly different from the global or regional mean. These differences can be characterized by the following Equation:

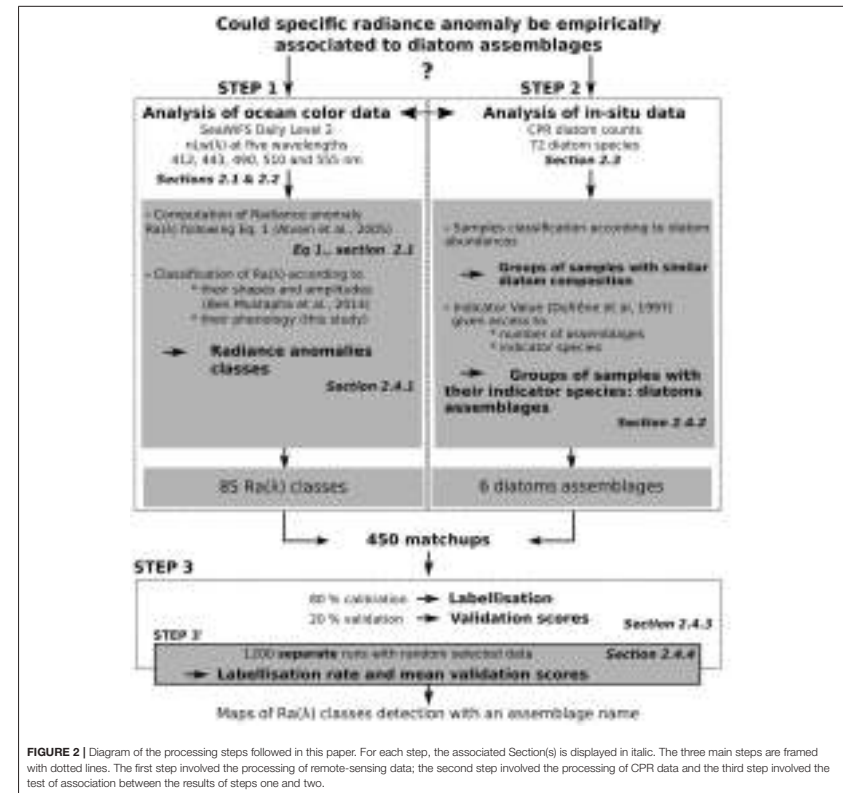
$$Ra(\lambda) = \frac{nLw(\lambda)}{nLw_{ref}(\lambda, Chl-a)} \quad (1)$$

Where  $Ra(\lambda)$  is the radiance anomaly at the sensor wavelengths, computed for each pixel;  $nLw(\lambda)$  is the observed normalized water-leaving radiance<sup>12</sup> at the same wavelength for a given pixel, and  $nLw_{ref}(\lambda, Chl-a)$  is the mean value of radiances at each wavelength and for a given chlorophyll-*a*.

Thus, when the value of  $Ra(\lambda)$  is greater than one, this indicates that the pixel has  $nLw(\lambda)$  values higher than the mean for the corresponding chlorophyll-*a* value. Whereas, when the value of  $Ra(\lambda)$  is less than one, this indicates that the pixel has  $nLw(\lambda)$  values lower than the mean. By considering the range of wavelength of a given remote sensor, it is possible to obtain  $Ra(\lambda)$  spectra with specific shape and amplitude, depending on the differences at each wavelength between the pixel values and the mean for the corresponding chlorophyll-*a*.

PHYSAT takes advantage of these small anomalies by empirically associating specific  $Ra(\lambda)$  shapes and amplitudes with a phytoplankton situation observed *in-situ*, through matchups<sup>10</sup> exercise and *in-situ* investigations. Theoretical studies first brought an explanation for the specific shapes and amplitudes, showing that they are related to a combination of phytoplankton absorption, backscattering, and dissolved organic matter absorption (Alvain et al., 2012). All these parameters, when acting simultaneously on the light, produce specific anomalies around the mean radiance signal for a given chlorophyll-*a* concentration, associated with specific phytoplankton situation. To develop PHYSAT it is therefore necessary to have:

- Mean values of  $nLw(\lambda)$  for small ranges of chlorophyll-*a*. These mean values have to be computed for a given sensor in order to calculate a mean signal, including the characteristics of the sensor and processes
- $Ra(\lambda)$  computed from the observed values and the above mean values following Equation 1; and then classified according to their shapes, amplitudes (Alvain et al., 2005; Ben Mustapha et al., 2014), and phenology (this study).



- A concurrent *in-situ* database with information on phytoplankton to empirically give a name to radiance anomalies
- Clear sky conditions (as define in Alvain et al., 2005) for matchups between the remote-sensing  $Ra$  and the *in-situ* data

Different approaches can be used to analyze  $Ra(\lambda)$  shapes and amplitudes. In the first version of PHYSAT (Alvain et al., 2005), large envelopes of minimum and maximum  $Ra(\lambda)$  values were used. This was in fact, the first time radiance anomalies had been analyzed. In Ben Mustapha et al. (2014), an improvement consisted of using a Self-Organizing Map (SOM hereafter) to classify the  $Ra(\lambda)$  more precisely and independently. Here, this

last approach was used, in addition to some phenological tools (see section 2.4.1). Phenological tools allow, for instance, the separation of pixels with similar radiance anomalies (shape and amplitude), but with very different phenological characteristics (see the Figure S1 in the Supplementary Material for an illustration). These steps allow the grouping of the  $Ra(\lambda)$  into refined  $Ra(\lambda)$  classes<sup>16</sup> with their specific shape, amplitude, and seasonal characteristics. Once this two-step classification is made, it is possible to empirically associate the  $Ra(\lambda)$  classes to a specific phytoplankton situation. In the 2005 (Alvain et al., 2005) and 2014 (Ben Mustapha et al., 2014) versions, biomarkers pigments were used for this step, allowing only the detection of

dominance<sup>6</sup> cases. Here, Ra( $\lambda$ ) spectra are associated with a more detailed *in-situ* database (CPR) in order to evaluate the potential of specific Ra( $\lambda$ ) for specific diatoms assemblages<sup>1</sup> in the surface waters. This last step aims to give a name to specific Ra( $\lambda$ ), names which could be completed each time new *in-situ* samples with Ra( $\lambda$ ) matchups are available.

## 2.2. Computation of PHYSAT Radiance Anomalies for This Study

In this study, the daily level 2 SeaWiFS remote-sensed normalized water leaving radiance (nLw, in the up-to-date version) between 1998 and 2004 in the English Channel and the North Sea ( $-4^{\circ}$ W to  $9^{\circ}$ E and  $48$  to  $60^{\circ}$ N), with a resolution of  $1 \text{ km}^2$ , was used (NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group, 2014). The SeaWiFS data were selected instead of other ocean-color sensors due to the greater number of matchups with CPR data. Standard flags were applied to mask out pixels from clouds, ice, land, and sun-glint. In addition, pixels with aerosol optical thickness (at  $865 \text{ nm}$ ) above  $0.15$  were excluded to ensure high atmospheric correction quality (Alvain et al., 2005). Furthermore, the classification proposed by Vantrepotte et al. (2012) was applied in this area to delineate waters with optical properties influenced by phytoplankton and associated CDOM<sup>3</sup> (from phytoplankton degradation), and to flag waters influenced by other mineral particulate matters and potential remaining waters influenced by terrestrial origin CDOM. Radiance anomalies were subsequently computed by applying Equation 1 to selected data. The use of SeaWiFS data till 2004 ensured a homogeneous spatial coverage in the study area. Equation 1 was applied based on regional level 2 data in order to refine the study in the area of interest in this first attempt to highlight empirical links between phytoplankton assemblages and Ra( $\lambda$ ) classes. Subsequently, a specific mean signal for small ranges of chlorophyll-*a* ( $nL_{w,ref}(\lambda, \text{Chl-}a)$  in Equation 1) was computed for the area under consideration.

## 2.3. In-Situ CPR Data

To evaluate the potential of PHYSAT in the detection of phytoplankton assemblages, it is necessary to make use of *in-situ* information about phytoplankton assemblages that may be potentially associated with a specific remote-sensing signal. An ideal *in-situ* database would be an inventory of all phytoplankton assemblages at a synoptic scale and with a high-repeat frequency over the course of one full calendar year (or longer). Unfortunately, such a database is currently not available. However, CPR data (SAHFOS, 2015) collected in the English Channel and North Sea, are particularly relevant in this first attempt to evaluate the potential of PHYSAT for the detection of diatom assemblages.

The Continuous Plankton Recorder survey is an upper layer plankton monitoring program that regularly collects samples in the North Atlantic and the North Sea, since 1946 (Reid et al., 2003). Studies have shown that this machine gives a satisfactory picture of the epipelagic zone (Lindley and Williams, 1980; Williams and Lindley, 1980). Inside the CPR machine, plankton are filtered by a moving band of silk—of mesh size  $270 \mu\text{m}$ . These samples (which represent approximately  $18 \text{ km}$  of tow)

are then analyzed later in the laboratory (Warner and Hays, 1994; Batten et al., 2003). Consequently, more than 400 species or taxa are identified and/or counted each month within the scale of the North Atlantic and its adjacent seas (Beaugrand, 2004). CPR phytoplankton abundance is a semi-quantitative estimate (see methods of abundance determination in Batten et al., 2003; Richardson et al., 2006, and Warner and Hays, 1994). However, the proportion of captured cells by the machine reflects major changes in abundance, distribution, and composition of phytoplankton (Robinson, 1970; Colebrook, 1982; Batten et al., 2003).

This *in-situ* dataset is unequaled in terms of size, spatio-temporal coverage, and representativeness of microphytoplankton abundance; it is classified into four groups: dinoflagellates, diatoms, silicoflagellates, and prymnesiophyceae. In particular, the CPR dataset (5,977 samples, Figure 1) recorded an abundance of 72 diatom species or taxa over the period 1998–2010, representing from 40 to 90% of the total microphytoplankton abundance in Summer (June–August) and Spring (March–June), respectively (Leterme et al., 2006). These diatoms are identified at a species or taxonomic level (one taxon could represent more than one species). Although diatoms are not alone in North Sea waters, and are associated with other types of phytoplankton (e.g., pico- and nano-phytoplankton), dealing with diatoms as an indicator of assemblages is the first step in describing phytoplankton communities. Thus, diatom assemblages are suitable to give a first answer to the following question: can specific radiance anomaly be empirically associated with the presence of a certain phytoplankton (here diatom) assemblages, as suggested by a previous theoretical study (Alvain et al., 2012)?

## 2.4. Processing Steps

### 2.4.1. Characterization of PHYSAT Radiance Anomalies: Step 1

The following processing steps correspond to Step 1 in Figure 2. The main objective was to characterize and classify the set of radiance anomalies. First, the regional radiance anomalies (computed by Equation 1) were classified according to their shape and amplitude by a Self-Organizing Map (SOM) algorithm (Kohonen, 2013) as previously defined in Ben Mustapha et al. (2014). Second, outgoing radiance anomaly classes were classified according to their phenological<sup>14</sup> signature. This latter classification is supported by the hypotheses that two similar spectra (shape and amplitude)—(i) having the same phenology could be considered as identical; and (ii) having a different phenology could be associated with different *in-situ* situations.

So, to describe the characteristics of PHYSAT radiance anomalies, regional Ra( $\lambda$ ) spectra were classified by SOM into 100 sets of radiance anomalies following the approach of Ben Mustapha et al. (2014). However, this large number of Ra( $\lambda$ ) classes was potentially above the variability attainable by the phytoplankton. Consequently, the outputs of the SOM classification were adjusted to reflect the sought-after variability. To that purpose, a phenology-based classification was applied to group some of the 100 Ra( $\lambda$ ) classes, which presented almost identical seasonal characteristics based on

occurrence frequencies (Figure 2, Step 1), by using conventional delineations of phenological characteristics (Platt et al., 2009). These metrics have been shown to be relevant at global and regional scales to characterize seasonal signals (e.g., Thackeray et al., 2008; Zhai et al., 2011; Racault et al., 2012). Timing of initiation<sup>14</sup>, maximum, termination<sup>14</sup>, and minimum, together with duration<sup>14</sup> between (i) initiation and maximum, and (ii) maximum and termination, were used as multivariate characteristics for each Ra( $\lambda$ ) class; then used to regroup the Ra( $\lambda$ ) classes with the traditional UPGMA<sup>19</sup> (Unweighted Pair Group Method with Arithmetic mean, Jain and Dubes, 1988) algorithm, based on an Euclidean distance. Analyses of classifications, based on different distances (Euclidean and Chord) and classification algorithms, did not alter the classifications (not shown here). Consequently, we selected the distance (here Euclidean) and the classification algorithm (here UPGMA) that yielded the best cophenetic correlation<sup>4</sup> (Halkidi et al., 2001). The UPGMA classification was performed using a bootstrap re-sampling method to obtain robust groups of Ra( $\lambda$ ) classes, and verified by the Approximately Unbiased *p*-value ( $p < 0.05$ ) (Shimodaira and Hasegawa, 2001; Shimodaira, 2002). The total number of Ra( $\lambda$ ) classes was downsized from 100 to 85, each representing a different radiance anomaly spectrum (shape and magnitude of the spectrum) and phenology (seasonal characteristics). At this stage, the 85 Ra( $\lambda$ ) classes are not associated with the presence of specific diatom assemblages in the surface open waters<sup>13</sup>. This last piece of information can be extracted from the CPR data analysis explained hereafter.

### 2.4.2. Characterization of Diatom Assemblages from In-Situ CPR Data: Step 2

The following processing steps correspond to Step 2 in Figure 2. The main objective was to analyze the large CPR database in order to obtain condensed ecological information suitable to be matched with Ra( $\lambda$ ) classes in Step 3. The CPR diatom counts recorded the occurrence of 72 diatom species and taxa. As mentioned above, diatoms were considered in this study as an indicator of the phytoplankton community composition; as such a species or taxa level description was not available from the other phytoplankton groups identified *via* CPR. The diatom species or taxa formed assemblages<sup>1</sup>; some occurring in the same location at the same time according to their physical, chemical, and/or biological preferences (Sverdrup, 1953; Platt and Jassby, 1976). Before the calibration<sup>2</sup> step of PHYSAT, it was essential to assess recurrent *in-situ* diatom assemblages and their composition. For that purpose, the Indicator Value method (IndVal method, Dufréne and Legendre, 1997) was chosen. This method has been successfully applied in various disciplines (e.g., Earth ecology, Chen et al., 2010; Ocean ecology, Darnis et al., 2008; medicine, Meadow et al., 2014; biochemistry, Schröder et al., 2015), and for various problems, e.g., climate change (Connor et al., 2013), and system pollution (Nahmani et al., 2006). It is a two-step approach: first, the samples are classified according to the composition and abundance of taxa; and second, an indicator value is computed according to the samples classification, allowing the indicator taxa or species<sup>8</sup> to be highlighted. The IndVal (Indicator Value) method was selected to define the diatom assemblages because

it deals with relative abundance and occurrence. It is based on a simple asymmetrical approach combining two criteria (the fidelity<sup>7</sup> and the specificity<sup>18</sup>) to define a composite measure of the indicator value of each taxon with respect to a group of samples. Essentially, the measure of fidelity<sup>7</sup> is at its maximum when a taxon occurs in only one group of samples; and specificity<sup>18</sup>, when a particular taxon is present in all samples of this group.

Assemblages characterization from the IndVal method first required the classification of CPR samples according to their diatom composition and abundances. The samples typology was obtained using an UPGMA classification algorithm<sup>19</sup>, based here on a dissimilarity distance computed on samples-standardized abundance of diatom taxa. The rare diatom taxa<sup>17</sup> were removed from this classification step, leading to a database of 11 non-rare diatom taxa. Then, the next step was to determine the number of assemblages that had both a statistical and an ecological meaning as defined in Legendre and Legendre (2012), through the computation of the Indicator Value (Equation 2). The IndVal is computed for each possible cut-off level in typological samples (i.e., for each possible diatom assemblage). For example, the first cut-off level delineated two diatom assemblages, and the second one delineated three diatom assemblages—with their respective Indicator Value.

$$\begin{aligned} A_{ij} &= \text{Nindividuals}_{ij} / \text{Nindividuals}_i \\ B_{ij} &= \text{Nsamples}_{ij} / \text{Nsamples}_j \\ \text{IndVal}_{ij} &= A_{ij} \times B_{ij} \end{aligned} \quad (2)$$

Where  $A_{ij}$  is a measure of specificity:  $\text{Nindividuals}_{ij}$  is the mean number of individuals of taxon  $i$  across group of samples  $j$ ; and  $\text{Nindividuals}_i$  is the sum of the mean number of individuals of taxon  $i$  over all groups of samples.

And  $B_{ij}$  is a measure of fidelity:  $\text{Nsamples}_{ij}$  is the number of samples in cluster  $j$  where taxon  $i$  occurs; and  $\text{Nsamples}_j$  is the number of samples in cluster  $j$ .

And  $\text{IndVal}_{ij}$  is the Indicator Value of the taxon  $i$  in the sample group  $j$  (Equations from Dufréne and Legendre, 1997).

An ideal number of assemblages is reached when the Indicator Value does not increase significantly for a higher number of assemblages (see details in Dufréne and Legendre, 1997; Aho et al., 2008). Please note that, the Indicator Value method was computed for the whole CPR diatom dataset (from 1998 to 2010) in order to reflect, as close as possible, the *in-situ* situations.

### 2.4.3. Calibration and Validation Procedure: Step 3

At this stage, *in-situ* diatom assemblages and classes of remote-sensing Ra( $\lambda$ ) (with different shape, magnitude, and phenology) were obtained. So, it was now possible to study potential matches between diatom assemblages and Ra( $\lambda$ ) classes. To achieve this, the *in-situ* CPR samples were matched to their associated satellite pixels (in a 3-by-3-pixel box corresponding to  $9 \text{ km}^2$ ) with a margin of  $\pm 3 \text{ h}$  between the two measures (classical matchup<sup>10</sup> procedures, referred in Bailey and Werdell, 2006, Figure 2, Step 3). The set of coincident measures ( $n_{\text{matchups}} = 450$ ) was randomly divided in two subsets: a calibration subset ( $n_{\text{calibration}} = 360$ ), and a validation subset ( $n_{\text{validation}} = 90$ ); corresponding

to 80 and 20% of the total number of matchups, respectively. The calibration step labeled<sup>9</sup> empirically the radiance anomalies if, at least 50% of Ra( $\lambda$ ) class observations were assigned to the same *in-situ* assemblage in the calibration subset of matchups (this is the classical threshold used in PHYSTAT, Alvain et al., 2005; Ben Mustapha et al., 2014).

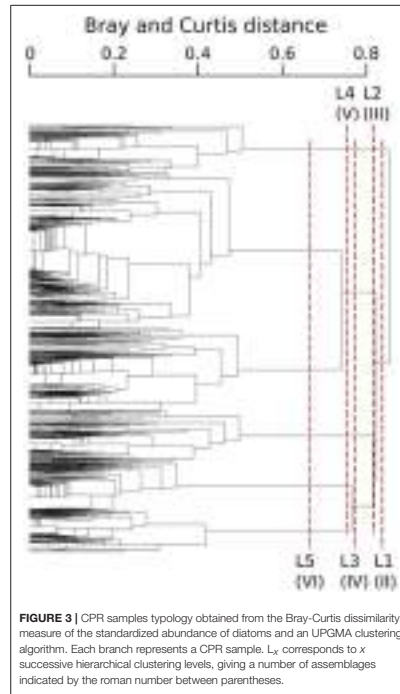
Then, the radiance anomalies labeled with a diatom assemblage were subsequently validated with an independent subset (20% of CPR coincident measures, not used for calibration). The validation procedure considered only the labeled-Ra( $\lambda$ ). Here, the occurrence of labeled-Ra( $\lambda$ ) (one or more radiance anomaly classes) over the validation subset was checked. Ideally, in the validation subset, the labeled-Ra( $\lambda$ ) class has to be associated with the same diatom assemblage as the one which it was associated during the calibration (corresponding to a validation of 100%). Misclassification occurs when a Ra( $\lambda$ ) class is associated with a specific assemblage in the calibration step, but is found to correspond to another assemblage in the validation step. Furthermore, a misclassification occurs too when the specific labeled Ra( $\lambda$ ) class is not found in the validation subset.

**2.4.4. Repetition of the Calibration Procedure: Step 3'** Because of the variability of the *in-situ* CPR data (in time and space), this calibration validation procedure was performed several times (number of separate procedures = 1,200) for different randomly-defined subsets of the calibration and validation to optimize the representativeness of the validation subset and strengthen the results (following a concept applied in ocean-color e.g., Cralg et al., 2012; Bracher et al., 2015a; Brewin et al., 2015). This number of separate repetitions ( $n = 1,200$ ) ensured that each randomly-selected matchup measurement was included at least once in the validation subset (Figure 2, Step 3). For each run of a calibration validation procedure, some Ra( $\lambda$ ) were labeled as a specific assemblage, followed by a validation percentage computation. This enabled the most robust Ra( $\lambda$ ) labels to be retained according to the rate of calibration, which meant that these labels had emerged regardless of the calibration and validation subsets.

### 3. RESULTS

#### 3.1. Indicator Taxa and Diatom Assemblages

The typology of CPR samples is displayed in Figure 3. According to this hierarchical clustering, the Indicator Value (IndVal, see section 2.4.2) was computed from level one (L1) to level six (L6)—see Table 1. For the first hierarchical subdivision in two groups (level 1, first Column of Table 1), only two diatom taxa were indicators (*Chaetoceros* spp2 and *Thalassiosira* spp.). From level 1 to 5 in Table 1, the number of indicator taxa increased, reaching six indicator taxa in level 5. The IndVal increased significantly among the first six hierarchical levels (or slightly decreased, insignificantly). There was an exception for the *Thalassiosira* spp., for which the IndVal values decreased significantly from level 1 to 5. This can be explained by the decrease in the specificity measure, whereas the fidelity measure was always at a maximum. The fifth level,



**FIGURE 3 |** CPR samples typology obtained from the Bray-Curtis dissimilarity measure of the standardized abundance of diatoms and an UPGMA clustering algorithm. Each branch represents a CPR sample. L<sub>x</sub> corresponds to x successive hierarchical clustering levels, giving a number of assemblages indicated by the roman number between parentheses.

differentiating six diatom assemblages, was the best one for which a compromise was found between the number of assemblages, the indicator taxa, and the IndVal values. Consequently, six diatom assemblages were retained, represented by an indicator diatom taxa and a set of secondary taxa. These assemblages are described in Table 2.

#### 3.2. Repetitive Calibration-Procedure

The number of matchups for each *in-situ* assemblage is presented in Column M (number of matchups) in Table 2. The calibration validation procedure was run 1,200 times to take into account the natural variability of the diatom assemblages. Figure 4 displays the number of times Ra( $\lambda$ ) was labeled for the 1,200 random calibration subsets. About 30% of Ra( $\lambda$ ) classes were never associated with an *in-situ* assemblage according to the criteria defined in the previous sections [for instance, Ra( $\lambda$ ) n°7 in Figure 4]. First, these Ra( $\lambda$ ) were either absent from the matchups or not present significantly to be labeled. Second,

**TABLE 1 |** This table represents a combination of the results from the classification of CPR samples and the computation of Indicator Values.

	1 (I)	2 (II)	3 (IV)	4 (V)	5 (VI)	6 (VII)
<i>Chaetoceros</i> spp1			27.8 (III)	38.3 (II)	<b>61.6 (II)</b>	59.3 (II)
<i>Chaetoceros</i> spp2	31.4 (I)		<b>70.2 (I)</b>	65.1 (I)	60.6 (I)	59.4 (I)
<i>T. nitzschoides</i>		30 (II)	<b>83.6 (II)</b>	81.1 (III)	77.8 (III)	77.1 (III)
<i>P. seriata</i>					<b>74.8 (IV)</b>	71.6 (IV)
<i>Thalassiosira</i> spp.	<b>92.4 (II)</b>	83.9 (III)	76.7 (III)	72.5 (IV)	67.9 (V)	67 (V)
<i>R. imbricata</i>				<b>86.1 (V)</b>	82.9 (VI)	81.9 (VI)

Each column represents a clustering level (number), given a number of assemblages (roman number). Indicator values are obtained for the successive hierarchical clustering levels from level one (two assemblages, I and II) to level six (seven assemblages, I to VII). For a species at a specific clustering level (a particular case of the table), IndVal is displayed as well as their attributed assemblage (indicated by the roman number between parentheses). Only indicator values greater than 25% and significant ( $p$ -value < 0.05) were displayed (according to Dufrêne and Legendre, 1997). The results for the six most abundant species are displayed in the Table for clarity, which are the six species for which the Indicator Value was significant. Bold values identify the highest IndVal observed among clustering levels for each dominant species.

**TABLE 2 |** This table represents information on *in-situ* data on the left part and on coincident measures on the right part.

Species in assemblage	M	C	V
<b><i>Chaetoceros</i> spp.2</b> (60.6); <i>Rhizosolenia styliformis</i> (3.6); <i>Corethron hystrix</i> (1.7)	52	41	11
<b><i>Chaetoceros</i> spp.1</b> (61.6); <i>Bacteriastrium</i> spp. (1.8); <i>Leptocylindrus</i> spp. (1.7)	72	57	15
<b><i>Thalassionema nitzschoides</i></b> (78.1); <i>Odontella sinensis</i> (11.8); <i>Fragilaria</i> spp. (4.8)	93	74	19
<b><i>Pseudonitzschia seriata</i></b> (74.8); <i>Proboscia alata</i> (8.3); <i>Coscinodiscus</i> spp. (2.7)	65	51	14
<b><i>Thalassiosira</i> spp.</b> (67.9); <i>Skeletonema costatum</i> (15.6); <i>Asterionellopsis glacialis</i> (7.0)	116	92	24
<b><i>Rhizosolenia imbricata</i></b> (82.9); <i>Rhizosolenia hebetata</i> (7.8); <i>Paralia sulcata</i> (7.4)	52	41	11

On the left part, the *in-situ* assemblages are described with indicator species (in bold) and secondary diatom species with their attributed Indicator Value according to Dufrêne and Legendre (1997) from CPR diatom data in parentheses. On the right part, M is the number of coincident measures associated with this assemblage; and C and V are respectively the number of coincident measures for calibration (80% of M) and validation (20% of M).

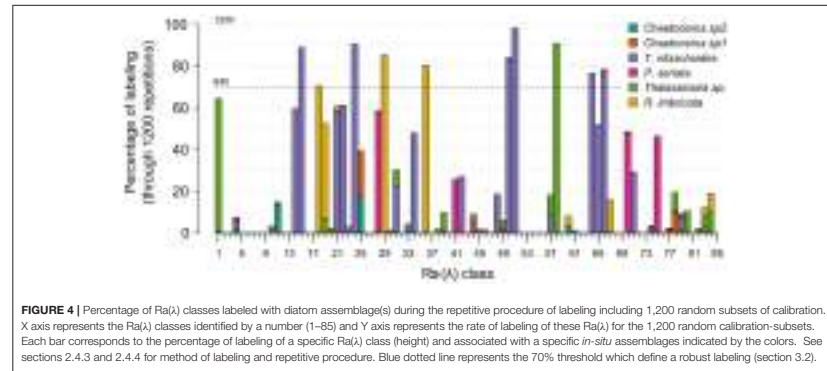
around 33% of Ra( $\lambda$ ) classes were associated with only one specific assemblage over the entire repetition procedure of calibration validation [for instance, Ra( $\lambda$ ) n°15 in Figure 4]. Finally, the remaining 37% of Ra( $\lambda$ ) classes were associated with multiple assemblages during the repetitive procedure [for instance, Ra( $\lambda$ ) n°25 in Figure 4]. These cases of multiple labeling were assumed to be possible considering that different phytoplanktonic communities can be associated with the presence of the diatom assemblage described here. For example, Ra( $\lambda$ ) n°57 in Figure 4, was labeled in less than 20% of calibration subsets; and labeled for half of the *Thalassionema nitzschoides* assemblage and half of the *Thalassiosira* spp. assemblage. For this Ra( $\lambda$ ), the *in-situ* information available at this stage was not sufficient as this Ra( $\lambda$ ) was probably associated with a unidentified phytoplankton assemblage at this point. For a few number of Ra( $\lambda$ ) classes [example n°50 in Figure 4], the Ra( $\lambda$ ) were labeled in a large majority with the same assemblage. This depended upon the random separation of CPR samples in calibration and validation subsets; this result shows how necessary the repetition of the calibration was using a repetitive procedure. Only the assemblage labels which appeared in at least 70% of calibration repetitions (840 random calibration subsets) were retained. This threshold was chosen to ensure that the approach was robust enough for this first attempt in identifying specific radiance anomalies associated with the presence of a specific diatom assemblage. The threshold helped to retain only labeling that could be observed regardless of the calibration

subset. Considering this threshold, the procedure led to the labeling of 11.7% of the 85 available Ra( $\lambda$ ) classes (total of Ra( $\lambda$ ) classes). Amongst the six *in-situ* diatom assemblages identified in the CPR dataset, three assemblages were robustly associated with ten Ra( $\lambda$ ) classes. In fact, during the 1200 repetitions of the calibration, assemblages characterized by *Thalassionema nitzschoides*, *Thalassiosira* sp., and *Rhizosolenia imbricata*, were associated with the one of the Ra( $\lambda$ ) classes more than 840 times (70% threshold). On the contrary, assemblages characterized by two taxa of *Chaetoceros* and by *Pseudonitzschia seriata* were not sufficiently associated with Ra( $\lambda$ ) classes in the multiple steps of calibration.

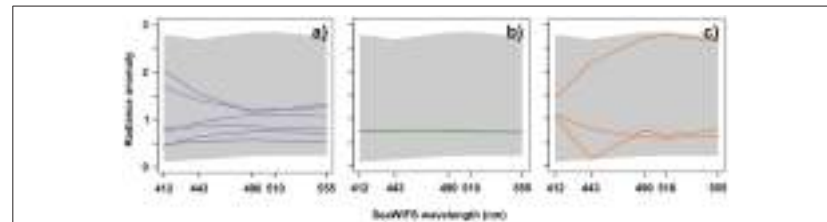
#### 3.3. Radiance Anomaly Spectra Associated with Diatom Assemblages

Figure 5 displays the spectra of labeled-Ra( $\lambda$ ) classes associated with the three diatom assemblages. These Ra( $\lambda$ ) spectra are characterized by different amplitudes, shapes, and seasonalities. Note that these results were expected for intra-assemblage as well as inter-assemblage spectra variations in terms of shape and amplitude (see further information in section 4). Please also consider that, some Ra( $\lambda$ ) could appear very similar in terms of shape and amplitude, but their seasonal characteristics were different (an example is shown in Figure S1). The assemblage characterized by *Thalassionema nitzschoides* (Table 2) was associated with six different Ra( $\lambda$ ) classes (representing 7.1% of available Ra( $\lambda$ ) classes, Figure 5a for referent spectra) based





**FIGURE 4 |** Percentage of Ra( $\lambda$ ) classes labeled with diatom assemblage(s) during the repetitive procedure of labeling including 1,200 random subsets of calibration. X axis represents the Ra( $\lambda$ ) classes identified by a number (1–85) and Y axis represents the rate of labeling of these Ra( $\lambda$ ) for the 1,200 random calibration-subsets. Each bar corresponds to the percentage of labeling of a specific Ra( $\lambda$ ) class (height) and associated with a specific *in-situ* assemblages indicated by the colors. See sections 2.4.3 and 2.4.4 for method of labeling and repetitive procedure. Blue dotted line represents the 70% threshold which define a robust labeling (section 3.2).



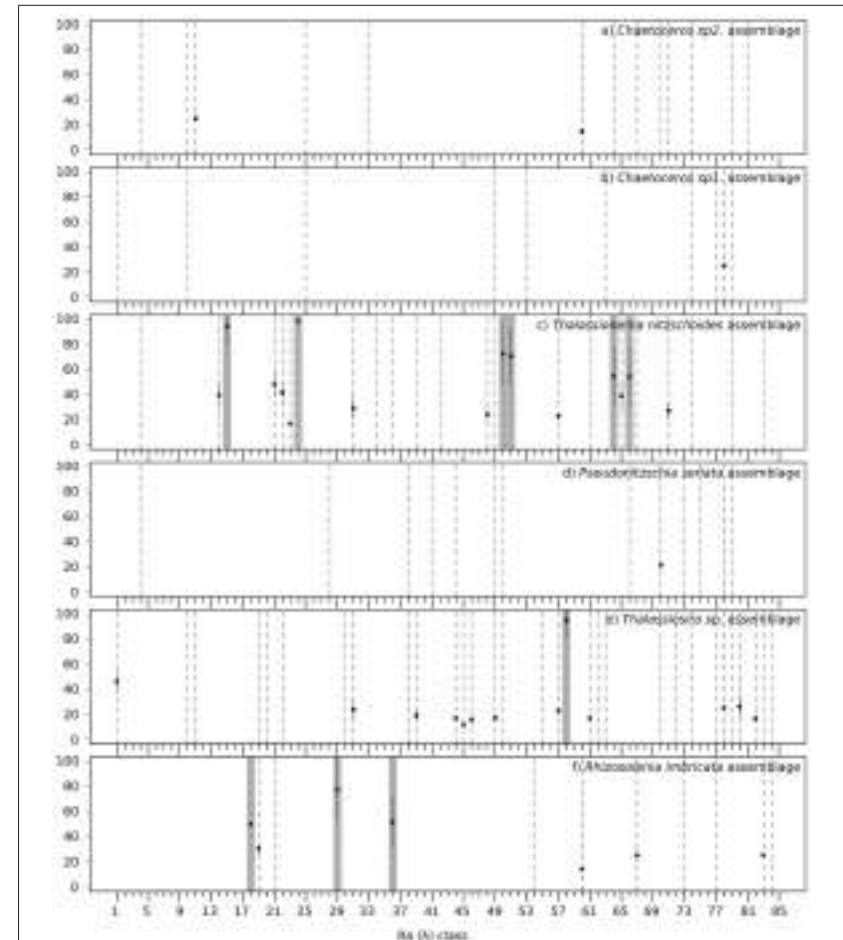
**FIGURE 5 |** Labeled-radiance anomaly spectra (reference spectra of Ra( $\lambda$ ) classes\*) for (a) assemblage characterized by *Thalassiosira nitzschoides*; (b) assemblage characterized by *Thalassiosira* spp.; and (c) assemblage characterized by *Rhizosolenia imbricata*. The gray background represents the entire diversity<sup>5</sup> (minimum and maximum) of North Sea radiance anomalies spectra. X axis represents the five SeaWiFS wavelengths and the Y axis represents the radiance anomaly according to Equation (1). \*The reference spectrum of a Ra( $\lambda$ ) class is computed from the Self-Organizing Map algorithm (SOM). It represents a compression of the information contained in all initial Ra( $\lambda$ ) spectra assigned in this specific Ra( $\lambda$ ) class by the SOM algorithm (Ben Mustapha et al., 2014).

on the available *in-situ* dataset. According to their shapes, these labeled Ra( $\lambda$ ) could be divided into two subgroups: high and low Ra( $\lambda$ ), at 412 and 443 nm, respectively. The assemblage characterized by *Thalassiosira* spp. was associated with one Ra( $\lambda$ ) class (1.2% of available Ra( $\lambda$ ), Figure 5b for referent spectrum). This reference spectrum had a straight shape. The *Rhizosolenia imbricata* assemblage was associated with three Ra( $\lambda$ ) classes (3.3% of available Ra( $\lambda$ ), Figure 5c for referent spectra). These spectra were different in shape and amplitude. At this stage, 10 Ra( $\lambda$ ) classes, with specific shapes, amplitudes, and seasonalities, were associated with three diatom assemblages, which were selected after the repetitive calibration validation procedure applied to the available CPR dataset.

### 3.4. Validation

Validation results are presented in Figure 6 and a summary of the validation scores is presented in Table 3. Figure 6 displays

the mean validation scores (and standard deviation) for all Ra( $\lambda$ ) classes over the successive repetition of the calibration procedure. This highlights the variability of the validation scores, according to the variation of the calibration and validation subsets of matchups. The three assemblages for which a strong association could not be found with Ra( $\lambda$ ) had validation scores around 20% (see Total validation in Table 3). The validation for assemblages with a strong association with Ra( $\lambda$ ), as defined in section 2.4.4, is displayed in two ways. First, the total validation scores, which takes into account all labeled Ra( $\lambda$ ) (robustly-labeled or not), reached 61.8, 60.6, and 60.5%, respectively for *Thalassiosira nitzschoides* assemblage, *Thalassiosira* spp. assemblage and *Rhizosolenia imbricata* assemblage (Table 3, column Total validation). These validation scores are increased by selecting the Ra( $\lambda$ ) according to the rate of labeling (70% threshold on repetitive calibration, see Figure 4 and section 2.4.4). Second, the validation scores reached 74.7,



**FIGURE 6 |** Mean validation scores (square points in percentage, Y axis) across the repetition of labeling procedure for the 85 Ra( $\lambda$ ) classes (X axis), with the standard deviation (black lines) for the six diatom assemblages: (a) *Chaetoceros* spp.2 assemblage, (b) *Chaetoceros* spp.1 assemblage, (c) *Thalassiosira nitzschoides* assemblage, (d) *Pseudonitzschia seriata* assemblage, (e) *Thalassiosira* spp. assemblage, and (f) *Rhizosolenia imbricata* assemblage. A dotted line is drawn when the Ra( $\lambda$ ) class has been labeled as the assemblage at least one time. A gray background is drawn when the Ra( $\lambda$ ) class has been robustly labeled according to the rate of labeling (see Figure 4 and section 2.4.4). When the *in-situ* information is not sufficient to label and validate the association, only a dotted line is drawn (for example a), Ra( $\lambda$ ) = 4).

**TABLE 3** | This table represents the validation scores in two ways: (i) the total validation scores, which take into account all labeled-Ra( $\lambda$ ) in each 1,200 repetition of the calibration/validation procedure (average of all validation scores displayed in **Figure 6**) and (ii) the selected validation score according to the rate of labeling (see **Figure 4** and section 2.4.4), which take into account only the robustly labeled-Ra( $\lambda$ ) (average of validation scores with a gray background in **Figure 6**).

Assemblage name	Labeled-Ra( $\lambda$ )	Total validation (%)	Selected validation (%)
<i>Chaetoceros</i> spp.2	0	17.06	0
<i>Chaetoceros</i> spp.1	0	25.0	0
<i>Thalassionema nitzschoides</i>	n° 15, 24, 50, 51, 64, 66	61.82	74.73
<i>Pseudonitzschia seriata</i>	0	21.25	0
<i>Thalassiosira</i> spp.	n° 58	60.58	95.24
<i>Rhizosolenia imbricata</i>	n° 18, 29, 36	60.47	63.62

When a robust association is not found, the symbol 0 is displayed.

95.2, and 63.8%, respectively, for *Thalassionema nitzschoides* assemblage, *Thalassiosira* spp. assemblage, and *Rhizosolenia imbricata* assemblage (**Table 3**, column Selected validation).

### 3.5. Spatial and Seasonal Distribution of Ra( $\lambda$ ) Classes Associated with Diatom Assemblages

Specific labeled Ra( $\lambda$ ) were sought in the daily standard ocean-color maps with the aim of producing frequency detection maps. Monthly mean spatial frequencies of the labeled Ra( $\lambda$ ) are displayed in **Figure 7**. The frequency of occurrence of the labeled Ra( $\lambda$ ) for each pixel was averaged for each month over the period 1998–2004 (based on SeaWiFS data). It is worth to note that the frequency rarely rose above 0.5, except for a few pixels, suggesting that extension by additional Ra( $\lambda$ ) classes not labeled yet, opens the possibility for future investigations. The Ra( $\lambda$ ) identified as the *Thalassionema nitzschoides* assemblage (**Figure 7A**) showed their highest frequencies in spring (March and April) and fall (September and October). These classes were more frequently detected in the southern part of the North-Sea, the English Channel, and along the western coasts of the United Kingdom (UK). In the summer, the frequency of Ra( $\lambda$ ) classes associated with *Thalassionema nitzschoides* was low, and was located in the northern part of the North Sea and along the southern coasts of the UK. The frequency of the detection of the Ra( $\lambda$ ) class labeled as the *Thalassiosira* spp. assemblage showed a very low frequency (due to only one labeled Ra( $\lambda$ ), based on available *in-situ* dataset at this stage). This Ra( $\lambda$ ) class had its highest frequency in the middle of the year (June and July), and was mainly located in the middle of the North Sea (**Figure 7B**). The Ra( $\lambda$ ) classes identified as the *Rhizosolenia imbricata* assemblage exhibited a higher frequency in spring (March and April), which subsequently diminished during the following months, remaining low for the rest of the year. These labeled-Ra( $\lambda$ ) were located in the English Channel, and along the southern coast of the UK. In summer (June and July), a

higher frequency pattern appeared near the Norwegian coast, which progressively disappeared during the following months (**Figure 7C**).

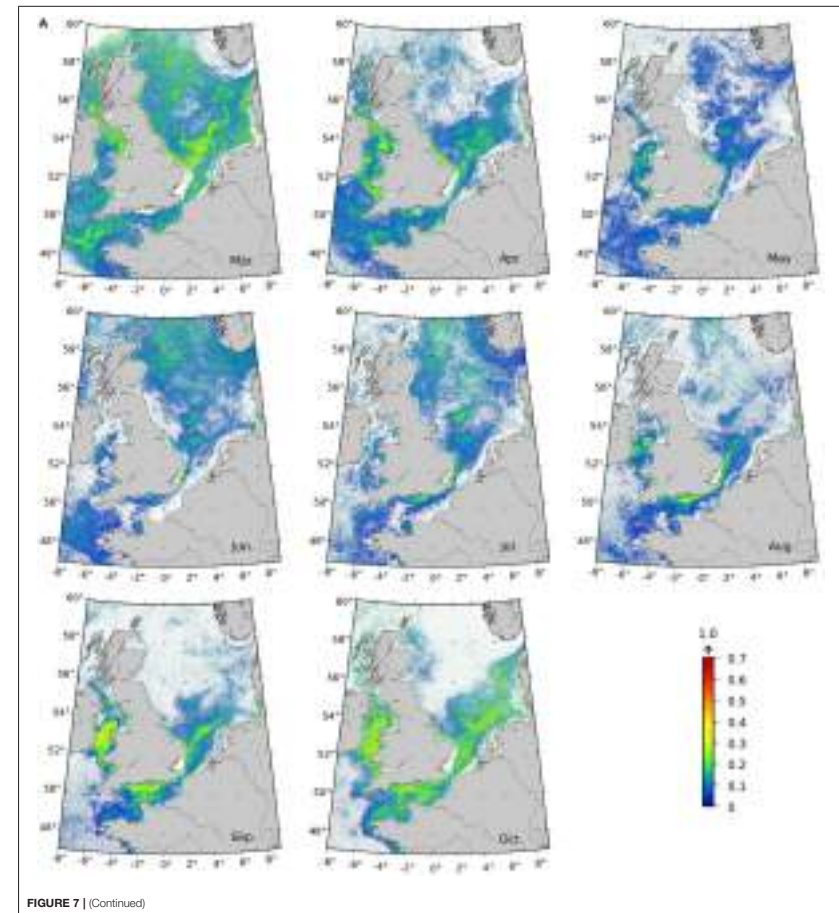
## 4. DISCUSSION

### 4.1. Concerning the Choice of Indicator Value Method to Analyze CPR Diatom Counts

The indicator value method, combined with samples classification, allowed the identification of six diatom assemblages, each having their own seasonal cycle and geographical characteristics. These assemblages were stable over the period of 1998–2010 (in qualitative terms). For the five indicator taxa (*Chaetoceros* spp1, spp2, *T. nitzschoides*, *P. seriata*, and *R. imbricata*), the IndVal increased until the fifth clustering level of the CPR samples classification (reaching six assemblages), and then decreased, indicating that these taxa are stenotopic, i.e., they have a small niche<sup>11</sup> breadth (Dufrene and Legendre, 1997). They also have a strong seasonality and a relatively limited spread, which has been commonly reported in the North Sea and English Channel (e.g., Reid et al., 1990; Rousseau et al., 2002; Hoppenrath, 2004; Bresnan et al., 2009; Hinder et al., 2012). On the contrary, *Thalassiosira* spp. IndVal decreased significantly among the successive hierarchical levels, indicating that this taxa is eurytopic, i.e., it has a large "niche breadth." During the IndVal procedure, *Thalassiosira* spp. was the only one which was indicator at the taxa level (potentially more than one species), which probably explains the eurytopic behavior of this taxon. However, these results are supported by a number of studies, which have reported an increase in *Thalassiosira* spp. abundance and its spatial distribution (e.g., Bresnan et al., 2009; Hinder et al., 2012).

In this study, diatom assemblages were used as indicator of phytoplankton community composition. However, the three diatom assemblages highlighted here vary in terms of associated phytoplankton community, which has not been totally assessed at this point. From the CPR data, the dinoflagellates could be present with diatoms. Indeed, *Thalassionema nitzschoides* assemblage contributes an average of 93% of total microphytoplankton cells (7% of dinoflagellates cells), 83.2% for *Thalassiosira* spp. assemblage (16.8% of dinoflagellates cells), and 65.1% for *Rhizosolenia imbricata* assemblage (34.9% of dinoflagellates cells). So, the dinoflagellate community could partly explain the variability in terms of labeled-Ra( $\lambda$ ) characteristics (section 4.3), together with the other non-characterized community. Although dinoflagellates are present in the assemblages alongside diatoms, they are not classified as indicators according to Dufrene and Legendre (1997).

Even if diatom assemblages are appropriate for this first evaluation of PHYSAT in the detection of assemblages, a more detailed *in-situ* database will be required in the future to take into account additional phytoplankton cases (more assemblages or taxonomic groups).



**FIGURE 7** | (Continued)

### 4.2. Spatio-Temporal Distribution of Labeled-Ra( $\lambda$ )

Based on the CPR dataset available for this study, ten Ra( $\lambda$ ) classes (out of 85) were associated with three specific diatom assemblages in North Sea waters, with a high degree of confidence. The three other assemblages could not be robustly

associated with Ra( $\lambda$ ) classes, according to our criteria (sections 2.4.3 and 2.4.4). For the ten Ra( $\lambda$ ) classes associated with diatom assemblages, monthly maps of Ra( $\lambda$ ) frequency show, for the first time, spatial and temporal variability in agreement with previously published literature in the North Sea (Reid et al., 1990; Beaugrand et al., 2001). In addition, the spatial

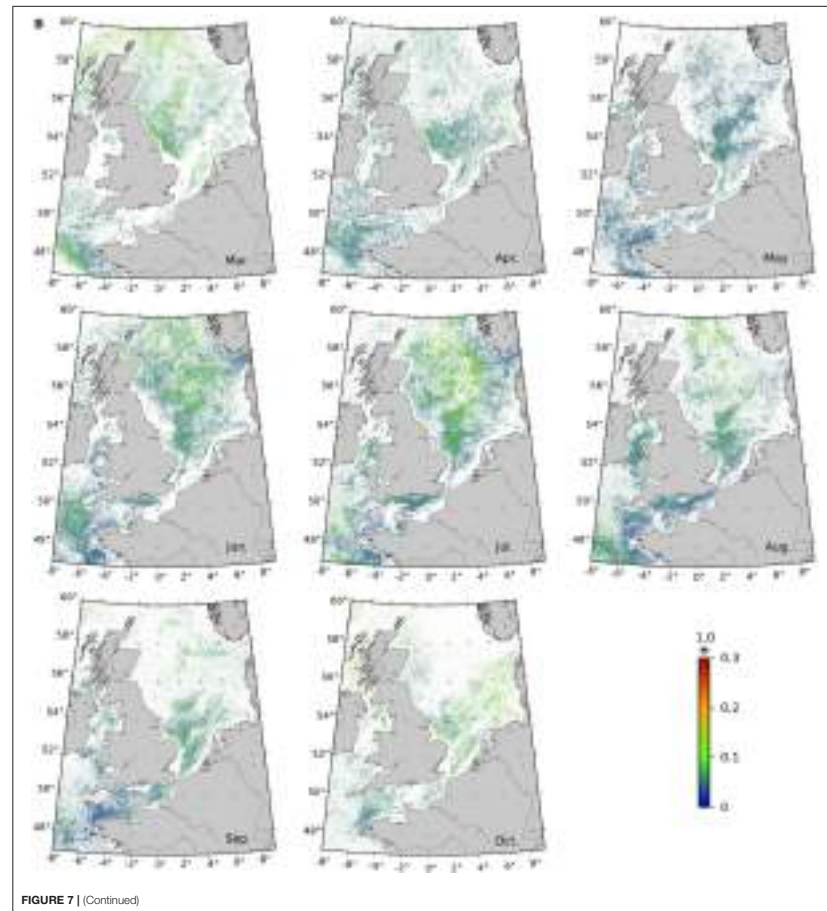


FIGURE 7 | (Continued)

distribution of all labeled- $Ra(\lambda)$  classes followed known patterns relating to physical and hydrological conditions in the North Sea during the year that drive the phytoplankton distribution (Reid et al., 1990). In fact, four ecological regions have been commonly recorded based on phytoplankton, zooplankton, and fish studies in the North Sea (Daan et al., 1990; Reid et al., 1990; Franz et al., 1991; Beaugrand et al., 2001). The Flamborough

Front is a transitional region between the north of the front seasonally stratified and the south of the front (well-mixed area) (Otto et al., 1990; Ducrottoy et al., 2000). The fourth region is the Skagerrak, related to the Kattegat-Skagerrak flow (Otto et al., 1990). These regions have different conditions in nutrients, temperature, and available light, explaining that different phytoplankton assemblages could bloom according

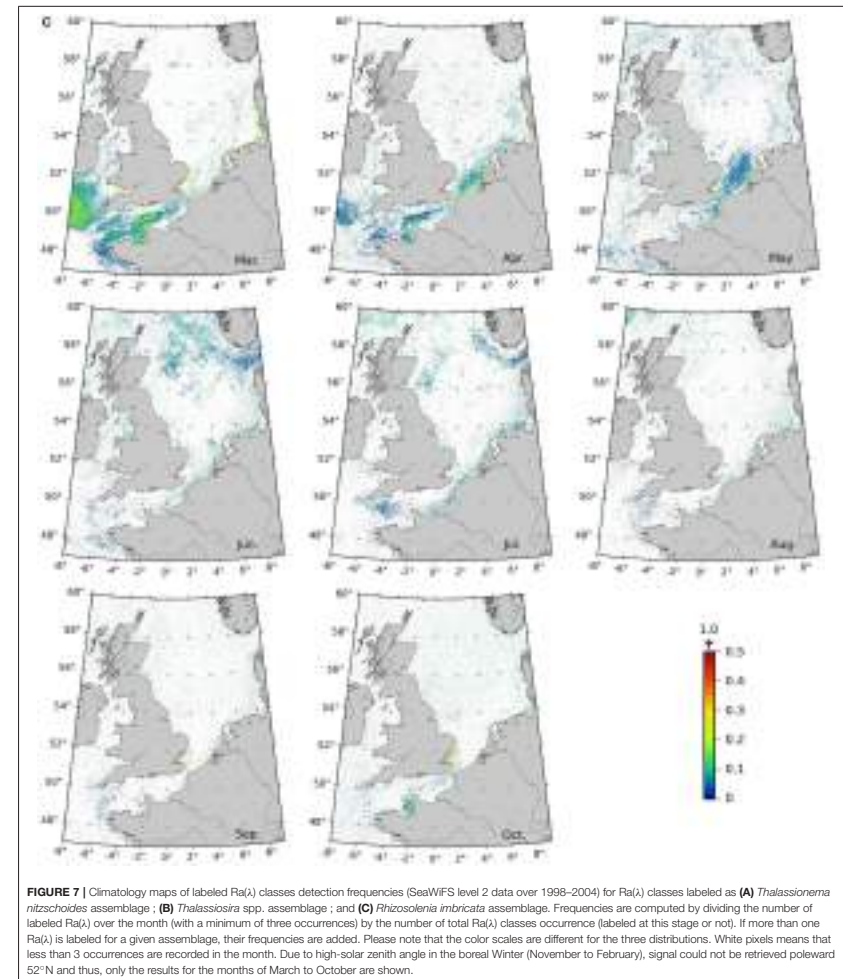
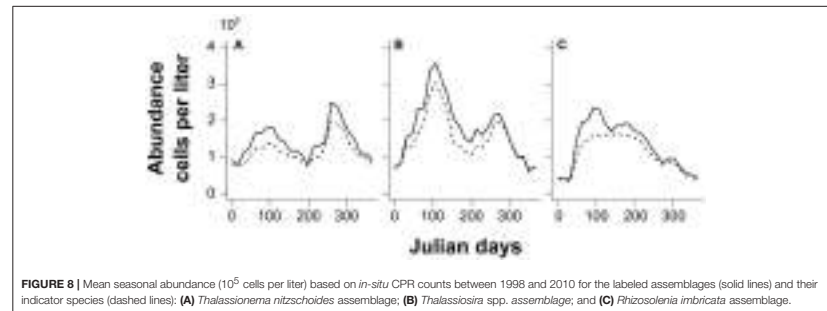


FIGURE 7 | Climatology maps of labeled  $Ra(\lambda)$  classes detection frequencies (SeaWiFS level 2 data over 1998–2004) for  $Ra(\lambda)$  classes labeled as (A) *Thalassionema nitzschoides* assemblage; (B) *Thalassiosira* spp. assemblage; and (C) *Rhizosolenia imbricata* assemblage. Frequencies are computed by dividing the number of labeled  $Ra(\lambda)$  over the month (with a minimum of three occurrences) by the number of total  $Ra(\lambda)$  classes occurrence (labeled at this stage or not). If more than one  $Ra(\lambda)$  is labeled for a given assemblage, their frequencies are added. Please note that the color scales are different for the three distributions. White pixels means that less than 3 occurrences are recorded in the month. Due to high-solar zenith angle in the boreal Winter (November to February), signal could not be retrieved poleward 52°N and thus, only the results for the months of March to October are shown.

to their preferences (Margalef, 1978). The spatial distribution of labeled  $Ra(\lambda)$  frequencies delineated these four ecological regions.

*Thalassionema nitzschoides* and its associated taxa are dominant in winter, and bloom both in early spring and autumn in the North Sea (Figure 8A). The literature records



**FIGURE 8** | Mean seasonal abundance ( $10^5$  cells per liter) based on *in-situ* CPR counts between 1998 and 2010 for the labeled assemblages (solid lines) and their indicator species (dashed lines): **(A)** *Thalassionema nitzschoides* assemblage; **(B)** *Thalassiosira* spp. assemblage; and **(C)** *Rhizosolenia imbricata* assemblage.

show that their occurrence in some locations of the southern North Sea relates to specific environmental conditions (e.g., Edwards et al., 2001; Rousseau et al., 2002; Wiltshire et al., 2010). In fact, the assemblage characterized by *Thalassionema nitzschoides* is common in high-nutrient conditions in neritic waters (Van Iperen et al., 1993) as is the case in spring and autumn in the southern North Sea (Commission, 2010). The distribution of the labeled  $Ra(\lambda)$  was very similar to the literature on the occurrence and bloom of this assemblage.

The *Rhizosolenia imbricata* assemblage blooms in early spring with mixed waters, and maintains itself until summer (Figure 8C); taking place most of the time along the coastline (e.g., Hinder et al., 2012). The labeled  $Ra(\lambda)$  had a similar spatio-temporal distribution.

Concerning the assemblage composed by *Thalassiosira* spp., the literature has recorded a winter dominance, a spring blooming period, and a second autumn bloom since the 2000s (Hinder et al., 2012 and Figure 8B). Our results, based on the labeled  $Ra(\lambda)$  at this stage, show, in contrast, a higher frequency in summer in the core area of the North Sea. However, this difference could be due to the limited amount of information available to correctly detect this group. In fact, coincident measurements are sufficient for calibration in summer (71.5% of the *Thalassiosira* spp. assemblage matchups from May to August); while there were only a few matchups for the other months (28.5%, split in other months). Even if this  $Ra(\lambda)$  was correctly labeled, based on our first validation exercise (60% of correct identification), there are probably other winter, spring, and/or autumn radiance anomalies that may correspond to this assemblage. Thus, for the moment, the frequency of labeled anomaly signal remains low due to the small number of  $Ra(\lambda)$  classes with coincident *in-situ* measures.

### 4.3. Variability in Labeled- $Ra(\lambda)$ Spectrum Characteristics in Terms of Shape, Amplitude and Seasonality

The main characteristics of the  $Ra(\lambda)$  used in the past PHYSAT versions (Alvain et al., 2005, 2008; Ben Mustapha et al., 2014)

are the shape and amplitude of the spectra (Alvain et al., 2012). In this study, the  $Ra(\lambda)$  have also been classified according to their phenology. This classification has two main advantages: (i) it avoids splitting the  $Ra(\lambda)$  classes (from the SOM algorithm), and therefore increases the chances of getting a larger number of coincident measurements; and (ii) it allows the distinction between two spectra that look similar in terms of shape and amplitude (as shown in Figure S1).

#### 4.3.1. Intra- and Inter-assemblage Variation of $Ra(\lambda)$ Characteristics

In some instances, one diatom assemblage could be associated with more than one  $Ra(\lambda)$  class (e.g., *Thalassionema nitzschoides* assemblage, which was associated with six  $Ra(\lambda)$  classes). These  $Ra(\lambda)$  spectra could be very different in terms of shape and amplitude. This is not surprising, as the shape and amplitude of  $Ra(\lambda)$  could vary with inherent optical properties due to phytoplankton characteristics and composition. In a theoretical study (Alvain et al., 2012) showed that radiance anomalies in PHYSAT results from a combination of (i) phytoplankton absorption, which affects both  $Ra(\lambda)$  shape and amplitude (especially at 443 nm); (ii) particulate backscattering, which strongly affects amplitude, but has a limited effect on the shape of  $Ra(\lambda)$ ; and (iii) CDOM (due to phytoplankton) absorption, which affects shape (mainly at 412 nm) and amplitude of the  $Ra(\lambda)$ . This may explain why several  $Ra(\lambda)$  could be associated with different assemblages depending on: (a) the phytoplankton community composition associated with the diatom assemblage; (b) the growth stage, inducing different size, morphology, cells arrangement (including free cells, chains forming and aggregates), and pigment concentration; and (c) the CDOM concentration, sometimes reflecting the decline of a bloom.

The following hypotheses are proposed for the explanation of the inter-assemblage variations of radiance anomalies: (i) the morphology of cells: pennate (*Thalassionema nitzschoides* and *Rhizosolenia imbricata*), and oblate diatoms (*Thalassiosira* spp.); (ii) the arrangements of cells: the three characteristic diatoms

could form chains (star, zigzag, and linear chains); (iii) cell sizes (from 2 to 230  $\mu\text{m}$ , depending on growth stage); and (iv) associated phytoplankton communities. Intra-assemblage variations of radiance anomaly could be explained by: (i) the shift between free and chained cells; (ii) cell size variations; (iii) different species (in the case they are identified at taxa level); (iv) the shift in associated communities (e.g., from associated other diatoms to dinoflagellates); or (iv) the adaptation of pigments (quality and quantity), depending on available light and on the stage of the bloom.

All of the above hypotheses will need further investigations, considering *in-situ* optical parameters and a full characterization of phytoplankton communities, CDOM, and non-algal particles. Presently, it is impossible to isolate the contribution of inherent optical properties on the variation of radiance anomaly without *in-situ* or laboratory experiments (Alvain et al., 2012). However, the robustness of the labels and validation encourages us to complete more analyses following these approaches. Furthermore, the phenological tools applied to the  $Ra(\lambda)$  classification seems a promising tool for the calibration of PHYSAT.

#### 4.3.2. "Multiple" Labeling Cases

The labeling-procedure shows that one  $Ra(\lambda)$  class could be successively linked to different assemblages depending on the calibration subset (cases of "multiple" labeling, see Figure 4, and section 3.2). Two different phytoplankton communities inducing the same radiance anomalies (depending on combinations of inherent optical properties) are rare cases according to Alvain et al. (2012), but are possible. In addition to an insufficient optical signal, this situation could also be explained by the presence of other phytoplankton communities associated with our diatom assemblages; not yet identified due to CPR data not available for other phytoplankton groups (pico-, and nano-phytoplankton). However, the validation scores observed in this study are encouraging, and also highlight the fact that a large number of  $Ra(\lambda)$  classes remain unattributed to any specific assemblage, indicating that our approach could be further completed. Future *in-situ* measurements coupling phytoplankton analyses and optical properties are strongly needed to ameliorate our procedure and identify whether specific  $Ra(\lambda)$  classes associated with many phytoplankton cases exist. The addition of phenology to classify the  $Ra(\lambda)$  reduces misclassification. Indeed, different optical properties leading to a similar spectra and phenology are considered as rare, based on current knowledge.

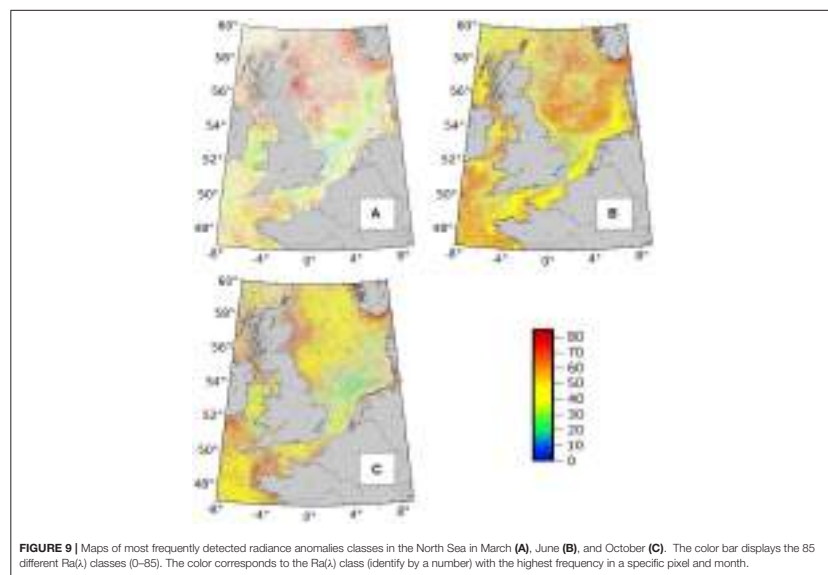
### 4.4. Remaining $Ra(\lambda)$ Database, Limitations, and Perspectives

For this first attempt to evaluate PHYSAT in non-dominance cases detection, very stringent steps were chosen, with a high degree selection (such as the 70% threshold on the rate of labeling). Potential future extension of this method, with additional *in-situ* data (hopefully not just limited to diatoms) will have to find a balance between strict thresholds, dataset availability, and score of validation. If the proposed method allows the identification of 11.7% of the available  $Ra(\lambda)$  classes

at this stage, the remaining anomaly classes are available for future investigation and empirical labeling. Figure 9 illustrates the potential of PHYSAT development by showing spatial distribution of the most frequently detected  $Ra(\lambda)$  classes in the North Sea. The  $Ra(\lambda)$  classes in Figure 9 show a clear seasonal succession, as the most frequent  $Ra(\lambda)$  classes are not the same between months. Furthermore, their spatial distributions delineate specific patterns in agreement with the environmental conditions described above (ecological regions in the North Sea). The 88.3% remaining unlabeled  $Ra(\lambda)$  classes could represent: (i) one of the diatom assemblages described here, at a different step of development, or associated with different non-diatom phytoplankton species (with no-matchup at the moment); or (ii) other unidentified assemblages of phytoplankton. In fact, the North Sea is a highly variable environment for phytoplankton, mostly dominated by diatoms (winter, spring, and autumn) and dinoflagellates (summer), but also by *Prymnesiophyceae* (such *Phaeocystis* sp.), *Chlorophyceae*, Cyanobacteria, and sometimes Coccolithophorids (Gieskes and Kraay, 1983; Reid et al., 1990; Hoppenrath, 2004; Hoppenrath et al., 2007; Sapp et al., 2007; Masquelier et al., 2011; Bonato et al., 2015).

If this study shows the potential of detection of phytoplankton situations beyond the dominance, there are still developments to be carried out with other *in-situ* information to characterize phytoplankton community. Although non-quantitative, these observations would be useful to estimate the contribution of assemblages to the chlorophyll-*a* estimation (Alvain et al., 2006, 2008) and then completing approaches such as OC-PFT (Hirata et al., 2011), PhytoDAS (Bracher et al., 2009). In this way, we support that development of *in-situ* measurements will allow us to identify more phytoplankton assemblages and better understand the limitation of this empirical approach. Some preliminary investigations, based on *in-situ* data, have shown that the  $Ra(\lambda)$  classes associated with diatoms assemblages previously characterized would not be associated with other groups (especially microphytoplankton as dinoflagellates) (not shown here). This observation is valid only for the  $Ra(\lambda)$  classes considered as robustly labeled, which would confirm that cases of multiple labeling (in the repetition of the calibration procedure) are due to other phytoplankton assemblages. These investigations will be considered in future developments.

Further studies will be required in the future to assess three main concerns about PHYSAT. First, a full characterization of the phytoplankton community and their optical properties during a bloom could permit us to analyze the variations of the radiance anomaly in terms of shape and amplitude, and to understand the optical justification of the empirical linkage (Alvain et al., 2012). Then, the CPR measurements are based on moving samples representing 18 km of tow, which may lead to a biased location of the sample and thereby on the matchups measurements. This has been considered in this study by observing the homogeneity of the remote-sensing observations within 3 days. This point needs to be considered for future studies, especially for a study of the  $Ra(\lambda)$  characteristics. Second, the requirements for regional algorithms have to be estimated by studying other regions. Previous studies have already demonstrated that regional ocean-color algorithms are sometimes necessary (for example, Volpe



**FIGURE 9** | Maps of most frequently detected radiance anomalies classes in the North Sea in March (A), June (B), and October (C). The color bar displays the 85 different  $Ra(x)$  classes (0–85). The color corresponds to the  $Ra(x)$  class (identify by a number) with the highest frequency in a specific pixel and month.

et al., 2007 for the Mediterranean Sea). The regional approach of PHYSAT developed for this first evaluation is not comparable to the global approach due to the use of specific reference tables in these two cases. Third, PHYSAT requires a specific reference table (see Equation 1) for each ocean-color sensor. For now, it is difficult to evaluate the uncertainties during the transition of the PHYSAT labels from one sensor to another one. Thus, impacts of multi-sensors inter-calibration will have to be carefully addressed before PHYSAT application to standardized remote-sensing database. That is one reason why we greatly support the development of *in-situ* measurements to carry out a new calibration and/or to validate the transposition of labels between ocean-color sensors.

## 5. CONCLUSION

PHYSAT (Alvain et al., 2005, 2008) is an ocean-color algorithm that has allowed us to identify five phytoplankton groups based on ocean-color radiance anomalies analysis in terms of shape and amplitude. However, this empirical algorithm was limited to dominant cases due to the use of biomarker pigment ratios in the empirical calibration step. In this study, we took advantage of both the addition of phenological tools to analyze the radiance anomalies and the diatom species

abundance data to evaluate the PHYSAT potential in terms of phytoplankton assemblages detection. The use of abundance data in the PHYSAT calibration step presents a promising avenue for future investigations and development of PFT algorithms. Although this study is based on a limited number of assemblages at this stage, it demonstrates the potential of development of the PHYSAT method beyond the dominance case, toward the detection of phytoplankton assemblages. Three diatom assemblages were successfully associated with specific ocean-color radiance anomalies with a robust validation. For now, there are no scientific reasons to prevent the use of PHYSAT radiance anomalies in other non-coastal waters and to associate remaining radiance anomalies with other assemblages. This first attempt to compute maps of signals associated with specific diatom assemblages highlights distinct spatial and temporal patterns in agreement with previous knowledge. Thyssen et al. (2015) showed that it is possible to describe the community structure at the basin-scale by comparing high resolution techniques (cytometry) and remote-sensing radiance anomalies. These recent results support both the development of PHYSAT beyond dominance-case detection, as well as the strong need of additional *in-situ* information about phytoplankton assemblages. This last objective could not only be performed based on CPR

observations, but also on automated *in-situ* analysis such as cytometry (Thyssen et al., 2015) or DNA sequencing approaches (Ottesen et al., 2011). Indeed, PHYSAT calibration is not limited to a specific type of data and only requires homogeneous datasets. Approaches based on ocean-color signal analysis, empirically coupled with exhaustive *in-situ* measurements (phytoplankton and optical properties) will help us to further understand which and how environmental conditions drive the phytoplankton composition and its succession, together with potential applications in the investigation of long-term changes.

## AUTHOR CONTRIBUTIONS

All authors contributed to the conception of the work, data analysis and interpretations. A-HR-L and SA drafted the article and all authors contributed to the critical revision of the article and made improvements. All authors approved the final version.

## FUNDING

This work was founded by the CNES/TOSCA program, CNRS and Hauts-de-France region. MR was partially supported by the

European Space Agency Living Planet Fellowship program grant number 4000112798/15/1/Bo.

## ACKNOWLEDGMENTS

The authors would like to thank NASA/GSFC/DAAC for providing access to daily SeaWiFS products and the Sir Alistair Hardy Foundation for Ocean Science (SAHFOS) for providing *in-situ* observations from CPR. [www.englisheditor.webs.com](http://www.englisheditor.webs.com) is thanked for its English proofing. David Antoine, Annick Bricaud, Lucile Duforêt-Gaurier, Hubert Loisel, Bertrand Lubac, and Julia Uitz are thanked for their valuable comments on an earlier draft of this paper. Experiments presented in this paper were carried out using the CALCULO computing platform, supported by SCOSI/ULCO (Service Commun du Système d'Information de l'Université du Littoral Côte d'Opale).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmars.2017.00408/full#supplementary-material>

## REFERENCES

- Aho, K., Roberts, D. W., and Weaver, T. (2008). Using geometric and non-geometric internal evaluators to compare eight vegetation classification methods. *J. Veg. Sci.* 19, 549–562. doi: 10.3170/2008-8-18406
- 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. *Int. J. Remote Sens.* 28, 497–516. doi: 10.1080/01431160600821036
- Alvain, S., Loisel, H., and Dessailly, D. (2012). Theoretical analysis of ocean color radiance anomalies and implications for phytoplankton groups detection in case 1 waters. *Opt. Exp.* 20, 1070–1083. doi: 10.1364/OE.20.01070
- Alvain, S., Moulin, C., Dandonneau, Y., and Bréon, F. (2005). Remote sensing of phytoplankton groups in case 1 waters from global seaWiFS imagery. *Deep Sea Res.* 52, 1989–2004. doi: 10.1016/j.dsr.2005.06.015
- Alvain, S., Moulin, C., Dandonneau, Y., and Loisel, H. (2008). Seasonal distribution and success of dominant phytoplankton groups in the global ocean: a satellite view. *Glob. Biogeochem. Cycles* 22, 1–15. doi: 10.1029/2007GB003154
- Alvain, S., Moulin, C., Dandonneau, Y., Loisel, H., and Bréon, F.-M. (2006). A species-dependant bio-optical model of case 1 waters for global ocean color processing. *Deep Sea Res.* 53, 917–925. doi: 10.1016/j.dsr.2006.01.011
- Bailey, S. W., and Werdell, P. J. (2006). A multi-sensor approach for the on-orbit validation of ocean color satellite data products. *Remote Sens. Environ.* 102, 12–23. doi: 10.1016/j.rse.2006.01.015
- Batten, S., Clark, R., Flinkman, J., Hays, G., John, E., John, A., et al. (2003). CPR sampling: the technical background, materials and methods, consistency and comparability. *Prog. Oceanogr.* 58, 193–215. doi: 10.1016/j.pocean.2003.08.004
- Beaugrand, G. (2004). Continuous plankton records: plankton atlas of the north atlantic ocean (1958–1999). I. Introduction and methodology. *Mar. Ecol. Prog. Ser.* 268, 3–10. doi: 10.3354/mepspr003
- Beaugrand, G. (2015). *Marine Biodiversity, Climatic Variability and Global Change*. Oxford, New York, NY: Routledge.
- Beaugrand, G., Ibañez, F., and Lindley, J. A. (2001). Geographical distribution and seasonal and diel changes of the diversity of calanoid copepods in the north atlantic and north sea. *Mar. Ecol. Prog. Ser.* 219, 189–203. doi: 10.3354/meps219189
- Ben Mustapha, Z., Alvain, S., Jamet, C., Loisel, H., and Dessailly, D. (2014). Automatic classification of water-leaving radiance anomalies from global seaWiFS imagery: application to the detection of phytoplankton groups in open ocean waters. *Remote Sens. Environ.* 146, 97–112. doi: 10.1016/j.rse.2013.08.046
- Bonato, S., Christaki, U., Lefebvre, A., Lizon, F., Thyssen, M., and Artigas, L. F. (2015). High spatial variability of phytoplankton assessed by flow cytometry, in a dynamic productive coastal area, in spring: the eastern english channel. *Estuar. Coast. Shelf Sci.* 154, 214–223. doi: 10.1016/j.ecss.2014.12.037
- Boss, E., Slade, W., and Hill, P. (2009). Effect of particulate aggregation in aquatic environments on the beam attenuation and its utility as a proxy for particulate mass. *Opt. Exp.* 17, 9408–9420. doi: 10.1364/OE.17.009408
- Bracher, A., Hardman-Mountford, N., Hirata, T., Bernard, S., Boss, E., Brewin, R., et al. (2015a). "Report on icgg workshop: phytoplankton composition from space: towards a validation strategy for satellite algorithms" in *The International Ocean-Colour Coordinating Group (IOCCG) 25-26 October 2014*(NASA/TM-2015-217528 (Portland, ME).
- Bracher, A., Taylor, M., Taylor, B., Dinter, T., Roettgers, R., and Steinmetz, F. (2015b). Using empirical orthogonal functions derived from remote sensing reflectance for the prediction of phytoplankton pigments concentrations. *Ocean Sci.* 11, 139–158. doi: 10.5194/os-11-139-2015
- Bracher, A., Vountas, M., Dinter, T., Burrows, J., Röttgers, R., and Peeken, I. (2009). Quantitative observation of cyanobacteria and diatoms from space using PhytoDOAS on SCIAMACHY data. *Biogeosciences* 6, 751–764. doi: 10.5194/bg-6-751-2009
- Bresnan, E., Hay, S., Hughes, S., Fraser, S., Rasmussen, J., Webster, L., et al. (2009). Seasonal and interannual variation in the phytoplankton community in the north east of scotland. *J. Sea Res.* 61, 17–25. doi: 10.1016/j.seares.2008.05.007
- Brewin, R. J., Hardman-Mountford, N. J., Lavender, S. J., Raisos, D. E., Hirata, T., Uitz, J., et al. (2011). An intercomparison of bio-optical techniques for detecting dominant phytoplankton size class from satellite remote sensing. *Remote Sens. Environ.* 115, 325–339. doi: 10.1016/j.rse.2010.09.004
- Brewin, R. J., Sathyendranath, S., Jackson, T., Barlow, R., Brotas, V., Ains, R., et al. (2015). Influence of light in the mixed-layer on the parameters of a three-component model of phytoplankton size class. *Remote Sens. Environ.* 168, 437–450. doi: 10.1016/j.rse.2015.07.004
- Bricaud, A., Clotti, A., and Gentili, B. (2012). Spatial-temporal variations in phytoplankton size and colored detrital matter absorption at global and

- regional scales, as derived from twelve years of seawifs data (1998–2009). *Glob. Biogeochem. Cycles* 26:GB1010. doi: 10.1029/2010GB003952
- Bricaud, A., and Morel, A. (1986). Light attenuation and scattering by phytoplanktonic cells: a theoretical modeling. *Appl. Opt.* 25, 571–580. doi: 10.1364/AO.25.000571
- Chen, L., Mi, X., Comita, L. S., Zhang, L., Ren, H., and Ma, K. (2010). Community-level consequences of density dependence and habitat association in a subtropical broad-leaved forest. *Ecol. Lett.* 13, 695–704. doi: 10.1111/j.1461-0248.2010.01468.x
- Clavano, W., Boss, E., and Karp-Boss, L. (2007). Inherent optical properties of non-spherical marine-like particles from the north to observation. *Oceanogr. Mar. Biol.* 45, 1–38. doi: 10.1201/9781420050943.ch1
- Colebrook, J. (1982). Continuous plankton records-phytoplankton, zooplankton and environment, northeast atlantic and north sea, 1958–1980. *Oceanol. Acta* 5, 473–480.
- Commission, O. (2010). *Bilan de Santé 2010*. London: Commission OSPAR.
- Connor, S. E., Ross, S. A., Sobotkova, A., Hetties, A. I., Mooney, S. D., Longford, C., et al. (2013). Environmental conditions in the se balkans since the last glacial maximum and their influence on the spread of agriculture into europe. *Q. Sci. Rev.* 68, 200–215. doi: 10.1016/j.quascirev.2013.02.011
- Craig, S. E., Jones, C. T., Li, W. K., Lazin, G., Horne, E., Caverhill, C., et al. (2012). Deriving optical metrics of coastal phytoplankton biomass from ocean colour. *Remote Sens. Environ.* 119, 72–83. doi: 10.1016/j.rse.2011.12.007
- Daan, N., Bromley, P., Hislop, J., and Nielsen, N. (1990). Ecology of north sea fish. *Netherlands J. Sea Res.* 26, 343–386. doi: 10.1016/0077-7579(90)90096-Y
- Darnis, G., Barber, D. G., and Fortier, L. (2008). Sea ice and the onshore-offshore gradient in pre-winter zooplankton assemblages in southeastern beaufort sea. *J. Mar. Syst.* 74, 994–1011. doi: 10.1016/j.jmarsys.2007.09.003
- De Monte, S., Soccodato, A., Alvain, S., and d'Ovidio, F. (2013). Can we detect oceanic biodiversity hotspots from space? *ISME J.* 7, 2054–2056. doi: 10.1038/ismej.2013.72
- Devred, E., Sathyendranath, S., Stuart, V., and Platt, T. (2011). A three component classification of phytoplankton absorption spectra: application to ocean color data. *Remote Sens. Environ.* 115, 2255–2266. doi: 10.1016/j.rse.2011.04.025
- D'Ovidio, F., De Monte, S., Alvain, S., Dandonneau, Y., and Lévy, M. (2010). Fluid dynamical niches of phytoplankton types. *Proc. Natl. Acad. Sci. U.S.A.* 107, 18366–18370. doi: 10.1073/pnas.1004620107
- Dubelaar, G. B., Visser, J. W., and Donze, M. (1987). Anomalous behaviour of forward and perpendicular light scattering of a cyanobacterium owing to intracellular gas vacuoles. *Cytometry* 8, 405–412. doi: 10.1002/cyto.990080410
- Ducrotot, J.-P., Elliott, M., and de Jonge, V. N. (2000). The north sea. *Mar. Pollut. Bull.* 41, 5–23. doi: 10.1016/S0025-326X(00)00099-0
- Duforêt-Gaurier, L., Loisel, H., Dessailly, D., Nordkvist, K., and Alvain, S. (2010). Estimates of particulate organic carbon over the euphotic depth from *in situ* measurements. Application to satellite data over the global ocean. *Deep Sea Res. I Oceanogr. Res. Papers* 57, 351–367. doi: 10.1016/j.dsr.2009.12.007
- Dufrène, M., and Legendre, P. (1997). Species assemblages and indicator species: the need for a flexible asymmetrical approach. *Ecol. Monogr.* 67, 345–366. doi: 10.2307/2963459
- Edwards, M., Reid, P., and Planque, B. (2001). Long-term and regional variability of phytoplankton biomass in the northeast atlantic (1960–1995). *ICES J. Mar. Sci.* 58, 39–49. doi: 10.1006/jmsc.2000.0987
- Falkowski, P. G. (1994). The role of phytoplankton photosynthesis in global biogeochemical cycles. *Photosynth. Res.* 39, 235–258. doi: 10.1007/BF00014586
- Fransz, H., Colebrook, J., Gamble, J., and Krause, M. (1991). The zooplankton of the north sea. *Netherlands J. Sea Res.* 28, 1–52. doi: 10.1016/0077-7579(91)90003-J
- Gieskes, W., and Kraay, G. (1983). Dominance of cryptophyceae during the phytoplankton spring bloom in the central north sea detected by hplc analysis of pigments. *Mar. Biol.* 75, 179–185. doi: 10.1007/BF00406000
- Halkidis, M., Batistakis, Y., and Vazirgiannis, M. (2001). On clustering validation techniques. *J. Intel. Inform. Syst.* 17, 107–145. doi: 10.1023/A:1012801612483
- Hinder, S. L., Hays, G. C., Edwards, M., Roberts, E. C., Walne, A. W., and Gravenor, M. B. (2012). Changes in marine diinoflagellate and diatom abundance under climate change. *Nat. Clim. Change* 2, 271–275. doi: 10.1038/nclimate1388
- Hirata, T., Hardman-Mountford, N. J., Brewin, R. J. W., Aiken, J., Barlow, R., Suzuki, K., et al. (2011). Synoptic relationships between surface Chlorophyll-a and diagnostic pigments specific to phytoplankton functional types. *Biogeochemistry* 8, 311–327. doi: 10.5194/bg-8-311-2011
- Hood, R. R., Laws, E. A., Armstrong, R. A., Bates, N. R., Brown, C. W., Carlson, C. A., et al. (2006). Pelagic functional group modeling: progress, challenges and prospects. *Deep Sea Res. II Top. Stud. Oceanogr.* 53, 459–512. doi: 10.1016/j.dsr2.2006.01.025
- Hoppenrath, M. (2004). A revised checklist of planktonic diatoms and dinoflagellates from helgoland (north sea, german bight). *Helgoländ. Mar. Res.* 58, 243–251. doi: 10.1007/s10152-004-0190-6
- Hoppenrath, M., Beszteri, B., Drees, G., Halliger, H., Van Beusekom, J. E., Janisch, S., et al. (2007). Thalassiosira species (bacillariophyceae, thalassiosirales) in the north sea at helgoland (german bight) and silt (north frish wadden sea)—a first approach to assessing diversity. *Eur. J. Phycol.* 42, 271–288. doi: 10.1080/09670260701352288
- Jain, A. K., and Dubes, R. C. (1988). *Algorithms for Clustering Data*. Upper Saddle River, NJ: Prentice-Hall, Inc.
- Kirk, J. T. (1994). *Light and Photosynthesis in Aquatic Ecosystems*. Cambridge: Cambridge University Press.
- Kohonen, T. (2013). Essentials of the self-organizing map. *Neural Netw.* 37, 52–65. doi: 10.1016/j.neunet.2012.09.018
- Kostadinov, T., Siegel, D., and Maritorea, S. (2009). Retrieval of the particle size distribution from satellite ocean color observations. *J. Geophys. Res. Oceans* 114, 1–22. doi: 10.1029/2009JC005303
- Kostadinov, T., Siegel, D., and Maritorea, S. (2010). Global variability of phytoplankton functional types from space: assessment via the particle size distribution. *Biogeochemistry* 8, 3239–3257. doi: 10.5194/bg-7-3239-2010
- Kump, L. R., Kasting, J. F., and Crane, R. G. (2010). *The Earth System*. San Francisco, CA: Prentice Hall.
- Le Quéré, C., Harrison, S. P., Collin Prentice, I., Buitenhuis, E. T., Aumont, O., Bopp, L., et al. (2005). Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. *Glob. Change Biol.* 11, 2016–2040. doi: 10.1111/j.1365-2486.2005.1004.x
- Legendre, P., and Legendre, L. F. (2012). *Numerical Ecology*, Vol. 24. Oxford: Elsevier.
- Leterme, S. C., Seuront, L., and Edwards, M. (2006). Differential contribution of diatoms and dinoflagellates to phytoplankton biomass in the ne atlantic ocean and the north sea. *Mar. Ecol. Prog. Ser.* 312, 57–65. doi: 10.3354/meps312057
- Li, Z., Li, L., Song, K., and Cassar, N. (2013). Estimation of phytoplankton size fractions based on spectral features of remote sensing ocean color data. *J. Geophys. Res. Oceans* 118, 1445–1458. doi: 10.1002/jgrc.20137
- Lindley, J., and Williams, R. (1980). Plankton of the fladen ground during flex 76 II. Population dynamics and production of *Thysanessa inermis* (crustacea: Euphausiacea). *Mar. Biol.* 57, 79–86. doi: 10.1007/BF00387373
- Margalef, R. (1978). Life-forms of phytoplankton as survival alternatives in an unstable environment. *Oceanol. Acta* 1, 493–509.
- Masquelier, S., Foulon, E., Jouenne, F., Ferréol, M., Brossard, C. P., and Vault, D. (2011). Distribution of eukaryotic plankton in the english channel and the north sea in summer. *J. Sea Res.* 66, 112–122. doi: 10.1016/j.seares.2011.05.004
- Meadow, J. F., Altrichter, A. E., Kembel, S. W., Moriyma, M., O'Connor, T. K., Womack, A. M., et al. (2014). Bacterial communities on classroom surfaces vary with human contact. *Microbiome* 2, 1–7. doi: 10.1186/2049-2618-2-7
- Morel, A., and Bricaud, A. (1981). Theoretical results concerning light absorption in a discrete medium, and application to specific absorption of phytoplankton. *Deep Sea Res. A Oceanogr. Res. Papers* 28, 1375–1393. doi: 10.1016/0198-0149(81)90039-X
- Mouw, C. B., Hardman-Mountford, N. J., Alvain, S., Bracher, A., Brewin, R. J., Bricaud, A., et al. (2017). A consumer's guide to satellite remote sensing of multiple phytoplankton groups in the global ocean. *Front. Mar. Sci.* 4:41. doi: 10.3389/fmars.2017.00041
- Mouw, C. B., and Yoder, J. A. (2010). Optical determination of phytoplankton size composition from global seaWiFS imagery. *J. Geophys. Res. Oceans* 115:C12108. doi: 10.1029/2010JC006337
- Nahmani, J., Lavelle, P., and Rossi, J.-P. (2006). Does changing the taxonomical resolution alter the value of soil macroinvertebrates as bioindicators of metal pollution? *Soil Biol. Biochem.* 38, 385–396. doi: 10.1016/j.soilbio.2005.04.037
- Nair, A., Sathyendranath, S., Platt, T., Morales, J., Stuart, V., Forget, M.-H., et al. (2008). Remote sensing of phytoplankton functional types. *Remote Sens. Environ.* 112, 3366–3375. doi: 10.1016/j.rse.2008.01.021
- NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group. (2014). *Sea-Viewing Wide Field-of-view Sensor (SeaWiFS) Data*. Greenbelt, MD: NASA OB.DAAC. doi: 10.5067/ORBVIEW-2/SEAWiFS/L3B/CHL/2014
- Navarro, G., Alvain, S., Vantrepotte, V., and Huertas, I. E. (2014). Identification of dominant phytoplankton functional types in the mediterranean sea based on a regionalized remote sensing approach. *Remote Sens. Environ.* 152, 557–575. doi: 10.1016/j.rse.2014.06.029
- O'Reilly, J. E., Maritorena, S., Mitchell, B. G., Siegel, D. A., Carder, K. L., Garver, S. A., et al. (1998). Ocean color chlorophyll algorithms for seawiFS. *J. Geophys. Res. Oceans* 103, 24937–24953. doi: 10.1029/98JC02160
- Ottesen, E. A., Marin, R., Preston, C. M., Young, C. R., Ryan, J. P., Scholin, C. A., et al. (2011). Metatranscriptomic analysis of autonomously collected and preserved marine bacterioplankton. *ISME J.* 5, 1881–1895. doi: 10.1038/ismej.2011.70
- Otto, L., Zimmerman, J., Furnes, G., Mork, M., Saetre, R., and Becker, G. (1990). Review of the physical oceanography of the north sea. *Netherlands J. Sea Res.* 26, 161–238. doi: 10.1016/0077-7579(90)90091-T
- Platt, T., and Jasby, A. D. (1976). The relationship between photosynthesis and light for natural assemblages of coastal marine phytoplankton. *J. Phycol.* 12, 421–430. doi: 10.1111/j.1529-8817.1976.tb02866.x
- Platt, T., White, G. N., Zhai, L., Sathyendranath, S., and Roy, S. (2009). The phenology of phytoplankton blooms: ecosystem indicators from remote sensing. *Ecol. Model.* 220, 3057–3069. doi: 10.1016/j.ecolmodel.2008.11.022
- Racault, M.-F., Le Qué, C., Buitenhuis, E., Sathyendranath, S., and Platt, T. (2012). Phytoplankton phenology in the global ocean. *Ecol. Indic.* 14, 152–163. doi: 10.1016/j.ecolind.2011.07.010
- Raisots, D. E., Lavender, S. J., Maravelias, C. D., Haralabous, J., Richardson, A. J., and Reid, P. C. (2008). Identifying four phytoplankton functional types from space: an ecological approach. *Limnol. Oceanogr.* 53, 605–613. doi: 10.4319/lmo.2008.53.2.0605
- Reid, P., Colebrook, J., Matthews, J., Aiken, J., and Team, C. P. R. (2003). The continuous plankton recorder: concepts and history, from plankton indicator to undulating recorder. *Prog. Oceanogr.* 58, 117–173. doi: 10.1016/j.pocan.2003.08.002
- Reid, P., Lancelot, C., Gieskes, W., Hagmeier, E., and Weichert, G. (1990). Phytoplankton of the north sea and its dynamics: a review. *Netherlands J. Sea Res.* 26, 295–331. doi: 10.1016/0077-7579(90)90094-W
- Richardson, A., Walne, A., John, A., Jonas, T., Lindley, J., Sims, D., et al. (2006). Using continuous plankton recorder data. *Prog. Oceanogr.* 68, 27–74. doi: 10.1016/j.pocan.2005.09.011
- Robinson, G. (1970). Continuous plankton records: variation in the seasonal cycle of phytoplankton in the north atlantic. *Bull. Mar. Ecol.* 6, 333–345.
- Rousseau, V., Leynaert, A., Daoud, N., and Lancelot, C. (2002). Diatom succession, silicification and silicic acid availability in belgian coastal waters (Southern North Sea). *Mar. Ecol. Prog. Ser.* 236, 61–73. doi: 10.3354/meps236061
- Sadeghi, A., Dinter, T., Vouantas, M., Taylor, B., Altenburg-Soppa, M., Peeken, I., et al. (2012). Improvement to the PhytoDOAS method for identification of coccolithophores using hyper-spectral satellite data. *Ocean Sci.* 8:1055. doi: 10.5194/os-8-1055-2012
- SAHFOS (2015). *Dinoflagellate, Diatom, Chlorophytes, Coccolithophores and Trichodesmium Data for the North Atlantic*. Plymouth: SAHFOS.
- Sapp, M., Wicshels, A., Willshire, K. H., and Gerdis, G. (2007). Bacterial community dynamics during the winter-spring transition in the north sea. *FEMS Microbiol. Ecol.* 59, 622–637. doi: 10.1111/j.1574-6941.2006.00238.x
- Sathyendranath, S., Aiken, J., Alvain, S., Barlow, R., Bouman, H., Bracher, A., et al. (2014). "Phytoplankton functional types from space" in *Reports of the International Ocean-Colour Coordinating Group (IOCCG)*, 15, 1–156.
- Sathyendranath, S., Watters, L., Devred, E., Platt, T., Caverhill, C., and Maass, H. (2004). Discrimination of diatoms from other phytoplankton using ocean-colour data. *Mar. Ecol. Prog. Ser.* 272, 59–68. doi: 10.3354/meps272059
- Schröder, M., Sondermann, M., Sures, B., and Hering, D. (2015). Effects of salinity gradients on benthic invertebrate and diatom communities in a german lowland river. *Ecol. Indic.* 57, 236–248. doi: 10.1016/j.ecolind.2015.04.038
- Shimodaira, H. (2002). An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.* 51, 492–508. doi: 10.1080/10635150290069913
- Shimodaira, H., and Hasegawa, M. (2001). Consel: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* 17, 1246–1247. doi: 10.1093/bioinformatics/17.12.1246
- Siegel, D., Maritorena, S., Nelson, N., Behrenfeld, M., and McClain, C. (2005). Colored dissolved organic matter and its influence on the satellite-based characterization of the ocean biosphere. *Geophys. Res. Lett.* 32:L20605. doi: 10.1029/2005GL024310
- Smyth, T. J., Moore, G. F., Groom, S. B., Land, P. E., and Tyrrell, T. (2002). Optical modeling and measurements of a coccolithophore bloom. *Appl. Opt.* 41, 7679–7688. doi: 10.1364/AO.41.007679
- Stramski, D., and Kiefer, D. A. (1991). Light scattering by microorganisms in the open ocean. *Prog. Oceanogr.* 28, 343–383. doi: 10.1016/0079-6611(91)90032-H
- Stramski, D., Reynolds, R., Babin, M., Kaczmarek, S., Lewis, M., Röttgers, R., et al. (2008). Relationships between the surface concentration of particulate organic carbon and optical properties in the eastern south pacific and eastern atlantic oceans. *Biogeochemistry* 5, 171–201. doi: 10.5194/bg-5-171-2008
- Subramaniam, A., Brown, C. W., Hood, R. R., Carpenter, E. J., and Capone, D. G. (2002). Detecting *Trichodesmium* blooms in SeaWiFS imagery. *Deep Sea Res. II Top. Stud. Oceanogr.* 49, 107–121. doi: 10.1016/S0967-0645(01)00096-0
- Sverdrup, H. U. (1953). On conditions for the vertical blooming of phytoplankton. *J. Conseil* 18, 287–295. doi: 10.1093/icesjms/18.3.287
- Thackeray, S., Jones, I., and Maberly, S. (2008). Long-term change in the phenology of spring phytoplankton: species-specific responses to nutrient enrichment and climatic change. *J. Ecol.* 96, 523–535. doi: 10.1111/j.1365-2745.2008.01355.x
- Thyssen, M., Alvain, S., Lefebvre, A., Dessailly, D., Rijkeboer, M., Guiselin, N., et al. (2015). High-resolution analysis of a north sea phytoplankton community structure based on *in situ* flow cytometry observations and potential implication for remote sensing. *Biogeochemistry* 12, 4051–4066. doi: 10.5194/bg-12-4051-2015
- Uitz, J., Claustre, H., Morel, A., and Hooker, S. B. (2006). Vertical distribution of phytoplankton communities in open ocean: an assessment based on surface chlorophyll. *J. Geophys. Res. Oceans* 111:C08005. doi: 10.1029/2005JC003207
- Van Iperen, J., Van Bennekom, A., and Van Weering, T. (1993). "Diatoms in surface sediments of the indonesian archipelago and their relation to hydrography," in *Twelfth International Diatom Symposium* (Dordrecht: Springer), 113–128.
- Vantrepotte, V., Loisel, H., Dessailly, D., and Mériaux, X. (2012). Optical classification of contrasted coastal waters. *Remote Sens. Environ.* 123, 306–323. doi: 10.1016/j.rse.2012.03.004
- Volpe, G., Santoleri, R., Vellucci, V., d'Alcala, M. R., Marullo, S., and D'Ortenzio, F. (2007). The colour of the Mediterranean Sea: global versus regional bio-optical algorithms evaluation and implication for satellite chlorophyll estimates. *Remote Sens. Environ.* 107, 625–638. doi: 10.1016/j.rse.2006.10.017
- Warner, A., and Hays, G. (1994). Sampling by the continuous plankton recorder survey. *Prog. Oceanogr.* 34, 237–256. doi: 10.1016/0079-6611(94)90011-6
- Whitmire, A. L., Pegau, W. S., Karp-Boss, L., Boss, E., and Cowles, T. J. (2010). Spectral backscattering properties of marine phytoplankton cultures. *Opt. Exp.* 18, 15073–15093. doi: 10.1364/OE.18.015073
- Williams, R., and Lindley, J. (1980). Plankton of the fladen ground during flex 76 I. Spring development of the plankton community. *Mar. Biol.* 57, 73–78. doi: 10.1007/BF00387372
- Willshire, K. H., Kraberg, A., Bartsch, L., Boersma, M., Franke, H.-D., Freund, J., et al. (2010). Helgoland roads, North Sea: 45 years of change. *Estuar. Coasts* 33, 295–310. doi: 10.1007/s12237-009-9228-y
- Zhai, L., Platt, T., Tang, C., Sathyendranath, S., and Walls, R. H. (2011). Phytoplankton phenology on the scotian shelf. *ICES J. Mar. Sci.* 68, 781–791. doi: 10.1093/icesjms/fsq175

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Réve-Lamarche, Alvain, Racault, Dessailly, Guiselin, Janet, Vantrepotte and Beaugrand. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Annexe 13 : Le *Continuous Plankton Recorder*

---

Le *Continuous Plankton Recorder* est un programme de suivi du plancton dans la couche supérieure des mers et des océans. Il a été mis en place depuis 1946, par Sir Alister Hardy and Sir Cyril Lucas, puis repris sous la tutelle du SAHFOS (*The Sir Alister Hardy Foundation for Ocean Science*) car il a montré son intérêt majeur. Ce programme a collecté régulièrement des échantillons dans l'Atlantique Nord et la Mer du Nord grâce à une machine appelé le CPR qui est placé sur des bateaux de marchandises (Reid *et al.*, 2003). De nombreuses études ont montré que cette machine donne une image satisfaisante de la zone épipélagique, aussi bien en termes de zooplancton que de phytoplancton (Lindley et Williams, 1980; Williams et Lindley, 1980).

Dans l'instrument CPR (Figure 1, ANNEXE 13), le plancton est filtré sur une membrane de soie qui se déroule à l'intérieur de la machine. La taille des pores est de 270  $\mu\text{m}$ . La vitesse du déroulement du filtre est proportionnelle à la vitesse du bateau grâce à une boîte de vitesse externe qui entraîne les roulements du réservoir et ainsi l'enroulage des filtres (Batten *et al.*, 2003). Le filtre où sont déposés les cellules planctoniques est recouvert lors de son entrainement vers le réservoir de stockage par un second filtre, piégeant les cellules planctoniques entre les deux tissus (Batten *et al.*, 2003). Ce double filtre renfermant les cellules est conservé sous forme d'une bobine dans le réservoir de stockage qui est rempli de formaldéhyde pour la conservation, dans le meilleur état possible, des cellules s'y étant déposées.

Les cellules planctoniques filtrées sont ensuite analysées en laboratoire selon les techniques décrites dans Warner et Hays (1994); Batten *et al.* (2003) :

- 1 L'indice de couleur du phytoplancton est déterminé selon des standards précis en quatre catégories : pas de couleur verte, vert très pale, vert pale, et vert.
- 2 Le phytoplancton est identifié et compté (estimation par pallier d'abondance) sur 20 zones de 275 $\mu\text{m}$  de diamètre. Ces 20 zones sont dessinées sur les deux diagonales de l'échantillon et représentent environ 1/8000e de la surface de celui-ci.
- 3 Chaque organisme zooplanctonique d'une taille inférieure à 2mm (en longueur) est identifié et compté sur 20 zones de 2,06mm de diamètre, situées sur les deux diagonales de l'échantillon, ce qui représente 1/49e de la surface de celui-ci.
- 4 Le zooplankton de grande taille (>2mm) est identifié et compté sur l'ensemble de l'échantillon.



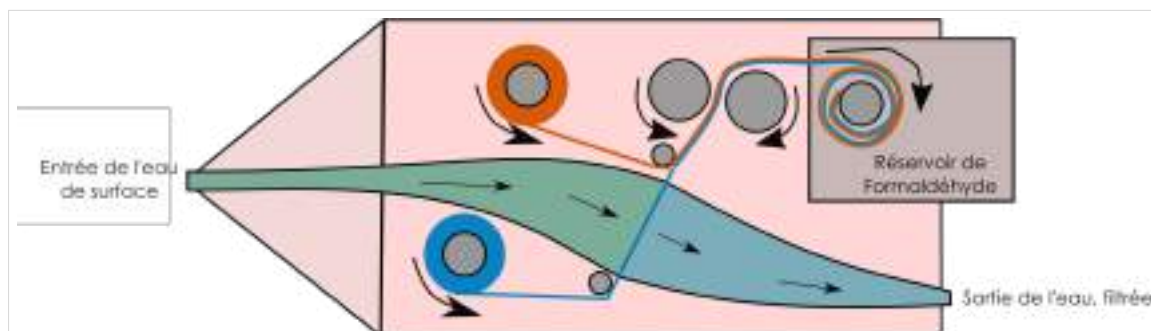


FIGURE 1, ANNEXE 13 – Schéma simplifié du fonction de la machine CPR. L'eau de surface entre par une petite ouverture. Elle est ensuite filtrée et ressort de la machine. Le filtre est lui entraîné vers un réservoir contenant du formaldéhyde. Il est recouvert au passage par un filtre protecteur pour éviter la perte de cellules planctoniques.

Plus de 400 espèces ou groupes taxonomiques ont été identifiés et/ou comptés chaque mois à l'échelle de l'Atlantique Nord et de ces mers adjacentes (Beaugrand, 2004).

En ce qui concerne les cellules phytoplanctoniques (point 1 ci-dessus), l'abondance est une estimation semi-quantitative (Warner et Hays, 1994; Batten *et al.*, 2003; Richardson *et al.*, 2006). Cependant, la proportion de cellules capturées par le CPR reflète les changements majeurs de l'abondance, de la distribution et de la composition du phytoplancton (Robinson, 1970; Colebrook, 1982; Batten *et al.*, 2003).

# Annexe 14 : Liste des espèces de diatomées des données CPR en Mer du Nord

---

Espèce de diatomée	Abréviation	Espèce de diatomée	Abréviation
<i>Actinoptychus</i> spp.1	ACTI1	<i>Meuniera membranacea</i>	MEUNMEM
<i>Asterionellopsis glacialis</i>	ASTEGLA	<i>Navicula</i> spp.1	NAVI1
<i>Bacillaria paxillifera</i>	BACIPAX	<i>Neocalyptrella robusta</i>	NEOCROB
<i>Bacteriastrum</i> spp.1	BACT1	<i>Nitzschia</i> spp.1	NITZ1
<i>Bellerochea horologicalis</i>	BELLHOR	<i>Nitzschia bicapitata</i>	NITZBIC
<i>Biddulphia alternans</i>	BIDDALT	<i>Nitzschia longissima</i>	NITZLON
<i>Bleakeleya notata</i>	BLEANOT	<i>Nitzschia sigma</i>	NITZSIG
<i>Centric diatoms</i>	CENT1	<i>Odontella aurita</i>	ODONAUT
<i>Cerataulina pelagica</i>	CERAPEL	<i>Odontella granulata</i>	ODONGRA
<i>Chaetoceros</i> spp.1	CHAE1	<i>Odontella mobiliensis</i>	ODONMOB
<i>Chaetoceros</i> spp.2	CHAE2	<i>Odontella regia</i>	ODONREG
<i>Corethron hystrix</i>	COREHYS	<i>Odontella rhombus</i>	ODONRHO
<i>Coscinodiscus</i> spp.1	COSCI1	<i>Odontella sinensis</i>	ODONSIN
<i>Coscinodiscus concinnus</i>	COSCCON	<i>Paralia sulcata</i>	PARASUL
<i>Coscinodiscus wailesii</i>	COSCWAI	<i>Podosira stelligera</i>	PODOSTE
<i>Cylindrotheca closterium</i>	CYLICLO	<i>Proboscia alata</i>	PROBALA
<i>Dactyliosolen antarcticus</i>	DACTANT	<i>Proboscia curvirostris</i>	PROBCUR
<i>Dactyliosolen fragilissimus</i>	DACTFRA	<i>Proboscia indica</i>	PROBIND
<i>Detonula confervacea</i>	DETOCON	<i>Proboscia inermis</i>	PROBINE
<i>Detonula pumila</i>	DETOPUM	<i>Pseudonitzschia delicatissima</i>	PSEUDEL
<i>Diploneis</i> spp. 1	DIPL1	<i>Pseudonitzschia seriata</i>	PSEUSER
<i>Ditylum brightwellii</i>	DITYBRI	<i>Rhaphoneis amphiceros</i>	RHAPAMP
<i>Ephemera planamembranacea</i>	EPHEPLA	<i>Rhizosolenia acuminata</i>	RHIZACU
<i>Eucampia</i> spp.	EUCA1	<i>Rhizosolenia bergonii</i>	RHIZBER
<i>Eucampia zodiacus</i>	EUCAZOD	<i>Rhizosolenia hebetata</i>	RHIZHEB
<i>Fragillaria</i> spp.1	FRAG1	<i>Rhizosolenia imbricata</i>	RHIZIMB
<i>Grammatophora</i> spp.1	GRAM1	<i>Rhizosolenia pungens</i>	RHIZPUN
<i>Guinardia cylindrus</i>	GUINCYL	<i>Rhizosolenia setigera</i>	RHIZSET
<i>Guinardia delicatula</i>	GUINDEL	<i>Rhizosolenia styliformis</i>	RHIZSTY
<i>Guinardia flaccida</i>	GUINFLA	<i>Skeletonema costatum</i>	SKELCOS
<i>Guinardia striata</i>	GUINSTR	<i>Stephanopyxis</i> spp.1	STEP1
<i>Gyrosigma</i> spp.1	GYRO1	<i>Thalassiosira</i> spp.1	THAL1
<i>Helicotheca tamesis</i>	HELITAM	<i>Thalassionema nitzschioides</i>	THALNIT
<i>Hemiaulus</i> spp.1	HEMI1	<i>Thalassiothrix longissima</i>	THAXLON
<i>Lauderia annulata</i>	LAUDANN		
<i>Leptocylindrus danicus</i>	LEPTDAN		
<i>Leptocylindrus mediterraneus</i>	LEPTMED		



# Annexe 15 : Méthode *Indicator Value* : résultats pour toutes les diatomées du CPR

---

Le tableau ci-dessous représente une combinaison des résultats de la classification des échantillons CPR et du calcul des valeurs indicatrices de la méthode IndVal pour l'ensemble des espèces (ou taxon) de diatomées du CPR. Chaque colonne représente un niveau de regroupement (chiffre arabe), pour lequel on délimite donc  $n$  assemblages (chiffre romain). Les valeurs indicatrices sont obtenues pour les niveaux successifs allant du premier niveau (deux assemblages) au sixième niveau (sept assemblages). Pour une espèce/taxon à un niveau de regroupement spécifique (i.e., pour une case particulière du tableau), l'IndVal est donné ainsi que le numéro de l'assemblage pour lequel cet IndVal a été maximum (indiqué par le chiffre romain en parenthèses). Les valeurs de Indval significatives selon Dufrière et Legendre (1997) pour chaque espèce ou taxon ont été mises en évidence en gras (p-value < 0,05, voir Section 5.1.1.2).

Espèce/taxon de diatomée	1 (II)	2 (III)	3 (IV)	4 (V)	5 (VI)	6 (VII)
<i>Actinoptychus spp.1</i>	0,21 (I)	0,15 (II)	0,18 (II)	0,13 (II)	0,26 (III)	0,22 (III)
<i>Asterionellopsis glacialis</i>	<b>9,8 (I)</b>	<b>8,81 (II)</b>	<b>7,05 (II)</b>	<b>6,17 (II)</b>	<b>7,06 (V)</b>	<b>5,22 (IV)</b>
<i>Bacillaria paxillifera</i>	2,64 (I)	<b>2,14 (II)</b>	<b>3,15 (IV)</b>	<b>2,82 (IV)</b>	<b>2,4 (V)</b>	<b>2,14 (VI)</b>
<i>Bacteriastrium spp.1</i>	2,47 (I)	1,92 (I)	1,36 (I)	1,56 (V)	<b>1,79 (II)</b>	<b>1,24 (II)</b>
<i>Bellerochea horologicalis</i>	1,2 (I)	0,9 (I)	<b>1,36 (IV)</b>	1,09 (IV)	0,92 (V)	0,8 (VI)
<i>Biddulphia alternans</i>	0,77 (I)	0,79 (II)	0,56 (II)	0,5 (II)	<b>1,02 (III)</b>	<b>0,87 (III)</b>
<i>Bleakeleya notata</i>	0,09 (I)	0,12 (II)	0,1 (IV)	0,1 (IV)	0,08 (V)	0,09 (II)
<i>Centric diatoms</i>	<b>1,25 (II)</b>	<b>1,08 (III)</b>	<b>0,78 (III)</b>	<b>0,69 (III)</b>	<b>0,68 (IV)</b>	<b>0,66 (V)</b>
<i>Cerataulina pelagica</i>	0,11 (I)	0,09 (I)	0,08 (IV)	0,18 (V)	0,14 (VI)	0,14 (VII)
<i>Chaetoceros spp.1</i>	40,25 (I)	33,9 (II)	<b>27,8 (III)</b>	<b>38,3 (II)</b>	<b>61,6 (II)</b>	<b>59,3 (II)</b>
<i>Chaetoceros spp.2</i>	<b>43,83 (I)</b>	31,4 (I)	<b>70,2 (I)</b>	<b>65,1 (I)</b>	<b>60,6 (I)</b>	<b>59,4 (I)</b>
<i>Corethron hystrix</i>	1,89 (I)	1,39 (I)	1,2 (I)	<b>2,34 (I)</b>	<b>2,15 (I)</b>	<b>1,73 (I)</b>
<i>Coscinodiscus spp.1</i>	<b>11,76 (II)</b>	<b>8,29 (III)</b>	<b>6,25 (III)</b>	<b>5,12 (III)</b>	<b>4,35 (IV)</b>	<b>3,91 (V)</b>
<i>Coscinodiscus concinnus</i>	0,35 (II)	0,27 (III)	0,2 (III)	0,18 (III)	0,15 (IV)	0,21 (IV)
<i>Coscinodiscus wailesii</i>	<b>6,17 (II)</b>	<b>4,82 (III)</b>	<b>3,85 (III)</b>	<b>3,31 (III)</b>	<b>2,83 (IV)</b>	<b>2,52 (V)</b>
<i>Cylindrotheca closterium</i>	<b>4,72 (I)</b>	3,02 (II)	2,57 (II)	2 (II)	1,75 (II)	<b>1,77 (IV)</b>
<i>Dactyliosolen antarcticus</i>	0,18 (I)	0,11 (II)	0,14 (II)	0,1 (II)	0,16 (II)	<b>0,36 (II)</b>
<i>Dactyliosolen fragilissimus</i>	1,63 (I)	<b>2,31 (I)</b>	<b>1,92 (I)</b>	<b>1,35 (I)</b>	<b>1,21 (I)</b>	<b>1,1 (I)</b>
<i>Detonula confervacea</i>	0,18 (I)	0,16 (II)	0,1 (II)	0,09 (II)	0,11 (III)	0,09 (III)
<i>Detonula pumila</i>	0,11 (I)	0,07 (II)	0,09 (II)	0,09 (V)	0,1 (II)	0,2 (II)
<i>Diploneis spp. 1</i>	0,16 (I)	0,14 (II)	0,21 (IV)	0,19 (IV)	0,15 (V)	0,13 (VI)
<i>Ditylum brightwellii</i>	<b>8,44 (I)</b>	<b>7,02 (II)</b>	<b>4,81 (IV)</b>	<b>4,14 (IV)</b>	<b>3,31 (V)</b>	2,47 (II)
<i>Ephemera planamembranacea</i>	0,02 (I)	0,02 (II)	0,03 (II)	0,03 (II)	0,04 (II)	0,12 (II)
<i>Eucampia spp.</i>	0,07 (I)	0,09 (I)	0,09 (I)	0,19 (I)	0,18 (I)	0,13 (I)
<i>Eucampia zodiacus</i>	1,76 (I)	1,24 (I)	<b>1,92 (IV)</b>	<b>1,56 (IV)</b>	1,29 (V)	<b>1,14 (VI)</b>

(suite Annexe 15)

Espèce/taxon de diatomée	1 (II)	2 (III)	3 (IV)	4 (V)	5 (VI)	6 (VII)
<i>Fragillaria spp.1</i>	<b>6,45 (I)</b>	<b>5,69 (II)</b>	<b>3,88 (II)</b>	<b>3,44 (II)</b>	<b>5,33 (III)</b>	<b>4,69 (III)</b>
<i>Grammatophora spp.1</i>	0,02 (I)	0,02 (II)	0,03 (II)	0,03 (II)	0,04 (II)	0,05 (IV)
<i>Guinardia cylindrus</i>	0,02 (I)	0,02 (II)	0,03 (II)	0,03 (II)	0,04 (II)	0,12 (II)
<i>Guinardia delicatula</i>	1,45 (I)	1,4 (I)	1,12 (I)	1,55 (V)	<b>1,38 (VI)</b>	1,21 (VII)
<i>Guinardia flaccida</i>	0,53 (I)	0,63 (I)	0,4 (I)	<b>0,7 (V)</b>	<b>0,63 (VI)</b>	<b>0,59 (VII)</b>
<i>Guinardia striata</i>	2,78 (I)	<b>4,58 (I)</b>	<b>3,52 (I)</b>	<b>3,53 (V)</b>	<b>3,32 (VI)</b>	<b>2,99 (VII)</b>
<i>Gyrosigma spp.1</i>	2,53 (I)	2,03 (I)	1,49 (I)	1,16 (I)	0,98 (I)	0,84 (I)
<i>Helicotheca tamesis</i>	0,21 (II)	0,16 (I)	0,16 (I)	0,21 (I)	0,21 (I)	0,21 (I)
<i>Hemiaulus spp.1</i>	0,02 (I)	0,02 (II)	0,18 (IV)	0,18 (IV)	0,18 (V)	0,18 (VI)
<i>Lauderia annulata</i>	2,69 (I)	1,75 (II)	1,22 (II)	1,1 (V)	1,12 (II)	<b>1,38 (IV)</b>
<i>Leptocylindrus danicus</i>	1,21 (I)	0,84 (II)	0,95 (IV)	0,78 (IV)	0,68 (V)	<b>0,98 (II)</b>
<i>Leptocylindrus mediterraneus</i>	1,84 (I)	1,26 (II)	<b>2,5 (IV)</b>	<b>2,14 (IV)</b>	<b>1,96 (II)</b>	<b>1,47 (VI)</b>
<i>Meuniera membranacea</i>	0,05 (I)	0,05 (I)	0,05 (I)	0,13 (V)	0,12 (VI)	0,1 (VII)
<i>Navicula spp.1</i>	2,56 (I)	2,02 (I)	1,5 (I)	1,33 (V)	<b>1,46 (III)</b>	<b>1,28 (III)</b>
<i>Neocalyptrella robusta</i>	0,07 (I)	0,13 (I)	0,13 (I)	<b>0,31 (V)</b>	<b>0,3 (VI)</b>	0,27 (VII)
<i>Nitzschia spp.1</i>	1,02 (I)	0,54 (II)	0,39 (II)	0,56 (V)	0,45 (VI)	0,4 (VII)
<i>Nitzschia bicapitata</i>	0,02 (I)	0,02 (II)	0,03 (II)	0,03 (II)	0,04 (II)	0,05 (IV)
<i>Nitzschia longissima</i>	0,18 (I)	0,24 (II)	<b>0,57 (IV)</b>	<b>0,57 (IV)</b>	<b>0,53 (V)</b>	<b>0,48 (VI)</b>
<i>Nitzschia sigma</i>	0,14 (I)	0,11 (II)	0,08 (IV)	0,07 (IV)	0,06 (II)	0,06 (IV)
<i>Odontella aurita</i>	0,02 (I)	0,02 (II)	0,18 (IV)	0,18 (IV)	0,18 (V)	0,18 (VI)
<i>Odontella granulata</i>	0,2 (I)	0,14 (II)	0,1 (II)	0,09 (I)	0,11 (II)	0,14 (IV)
<i>Odontella mobiliensis</i>	0,2 (I)	0,26 (II)	0,15 (IV)	0,15 (IV)	0,16 (III)	0,13 (III)

Espèce/taxon de diatomée	(suite Annexe 15)					
	1 (II)	2 (III)	3 (IV)	4 (V)	5 (VI)	6 (VII)
<i>Odontella regia</i>	1,31 (I)	1,06 (II)	0,73 (IV)	0,62 (IV)	<b>1,44 (III)</b>	<b>1,32 (III)</b>
<i>Odontella rhombus</i>	0,35 (II)	0,23 (I)	0,24 (IV)	0,24 (V)	0,21 (VI)	0,2 (VII)
<i>Odontella sinensis</i>	<b>32,84 (II)</b>	<b>27,51 (III)</b>	<b>21,81 (III)</b>	<b>19,5 (III)</b>	<b>16,07 (III)</b>	<b>14,91 (V)</b>
<i>Paralia sulcata</i>	11,54 (II)	<b>10,59 (I)</b>	<b>8,41 (I)</b>	<b>8,21 (V)</b>	<b>7,4 (VI)</b>	<b>6,32 (VII)</b>
<i>Podosira stelligera</i>	0,28 (I)	0,21 (I)	0,2 (I)	0,39 (V)	0,32 (VI)	0,27 (VII)
<i>Proboscia alata</i>	<b>33,69 (II)</b>	<b>27,52 (III)</b>	<b>23,47 (III)</b>	<b>19,96 (III)</b>	<b>18,74 (IV)</b>	<b>16,94 (V)</b>
<i>Proboscia curvirostris</i>	0,29 (II)	0,29 (III)	0,29 (III)	0,29 (III)	0,29 (IV)	0,29 (V)
<i>Proboscia indica</i>	0,52 (II)	0,42 (III)	0,31 (III)	0,28 (III)	0,24 (IV)	0,27 (II)
<i>Proboscia inermis</i>	1,69 (II)	1,19 (III)	0,9 (III)	0,73 (III)	0,63 (IV)	0,54 (V)
<i>Pseudonitzschia delicatissima</i>	<b>14,64 (II)</b>	<b>11,16 (III)</b>	<b>8,49 (IV)</b>	<b>7,29 (IV)</b>	<b>6,48 (V)</b>	<b>5,52 (VI)</b>
<i>Pseudonitzschia seriata</i>	27,83 (I)	23,17 (II)	87,19 (IV)	82,86 (IV)	<b>74,8 (IV)</b>	<b>71,6 (IV)</b>
<i>Rhaphoneis amphiceros</i>	1,57 (I)	1,06 (I)	0,81 (I)	0,73 (I)	<b>1,35 (III)</b>	<b>1,2 (III)</b>
<i>Rhizosolenia acuminata</i>	0,07 (I)	0,11 (I)	0,1 (I)	0,08 (V)	0,07 (VI)	0,07 (VII)
<i>Rhizosolenia bergonii</i>	0,02 (I)	0,02 (II)	0,03 (II)	0,03 (II)	0,04 (II)	0,05 (IV)
<i>Rhizosolenia hebetata</i>	<b>11,48 (I)</b>	<b>14,8 (I)</b>	<b>11,13 (I)</b>	<b>9,29 (V)</b>	<b>8,65 (VI)</b>	<b>7,63 (VII)</b>
<i>Rhizosolenia imbricata</i>	23,92 (I)	49,57 (I)	44,27 (I)	<b>86,1 (V)</b>	<b>82,9 (VI)</b>	<b>81,9 (VI)</b>
<i>Rhizosolenia pungens</i>	1,32 (I)	1,2 (I)	<b>1,52 (IV)</b>	<b>1,1 (IV)</b>	<b>1,04 (V)</b>	<b>0,89 (VI)</b>
<i>Rhizosolenia setigera</i>	3,16 (I)	2,09 (II)	<b>2,77 (IV)</b>	<b>2,29 (IV)</b>	<b>2,02 (V)</b>	<b>1,63 (VI)</b>
<i>Rhizosolenia styliformis</i>	7,31 (I)	<b>8,66 (I)</b>	<b>6,93 (I)</b>	<b>5,39 (I)</b>	<b>4,85 (I)</b>	<b>4,18 (I)</b>
<i>Skeletonema costatum</i>	<b>11,61 (I)</b>	<b>12,64 (II)</b>	<b>12,02 (II)</b>	<b>11,15 (II)</b>	<b>15,6 (V)</b>	<b>12,18 (IV)</b>
<i>Stephanopyxis spp.1</i>	0,57 (II)	0,43 (I)	0,32 (I)	0,39 (V)	0,33 (VI)	0,28 (VII)
<i>Thalassiosira spp.1</i>	<b>92,4 (II)</b>	<b>83,9 (III)</b>	<b>76,7 (III)</b>	<b>72,5 (IV)</b>	<b>67,9 (V)</b>	<b>67 (V)</b>
<i>Thalassionema nitzschioides</i>	31,93 (I)	30 (II)	<b>83,6 (II)</b>	<b>81,1 (III)</b>	<b>77,8 (III)</b>	<b>77,1 (III)</b>
<i>Thalassiothrix longissima</i>	1,9 (I)	1,5 (II)	1,11 (II)	0,9 (II)	0,87 (II)	0,75 (IV)

