



HAL
open science

Combinaison multi-capteurs de données de couleur de l'eau : application en océanographie opérationnelle

Claire Pottier

► **To cite this version:**

Claire Pottier. Combinaison multi-capteurs de données de couleur de l'eau : application en océanographie opérationnelle. Autre. Université Paul Sabatier - Toulouse III, 2006. Français. NNT: . tel-00179729

HAL Id: tel-00179729

<https://theses.hal.science/tel-00179729>

Submitted on 16 Oct 2007

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.

UNIVERSITE TOULOUSE III – PAUL SABATIER

ECOLE DOCTORALE SCIENCES DE L'UNIVERS, DE L'ENVIRONNEMENT ET DE L'ESPACE
Formation Doctorale Océan Atmosphère Environnement

THESE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE TOULOUSE III

Discipline : Traitement d'Image en Océanographie Spatiale

présentée et soutenue par

Claire POTTIER

le 22 décembre 2006.

**Combinaison multi-capteurs
de données de couleur de l'eau :
Application en océanographie opérationnelle**

Directrice de thèse : Véronique Garçon, Directeur de Recherches CNRS, *LEGOS*

JURY

Serge Chauzy, Professeur, *Université Toulouse III*,
Nicolas Hoepffner, Senior Scientist, *JRC Ispra (Italie)*,
Hussein Yahia, Chargé de Recherches, *INRIA Rocquencourt*,
Gilles Larnicol, Ingénieur d'études, *CLS Toulouse*,
Pierre-Yves Le Traon, Cadre de Recherche IIIA, *Ifremer Brest*,
Eric Thouvenot, Ingénieur, *CNES Toulouse*,
Antonio Turiel, Chercheur, *ICM Barcelone (Espagne)*,

Président
Rapporteur
Rapporteur
Examineur
Examineur
Examineur
Examineur

Laboratoire d'Etudes en Géophysique et Océanographie Spatiales (LEGOS)
UMR 5566 – CNRS/IRD/UPS/CNES – Observatoire Midi-Pyrénées
14 avenue Edouard Belin – 31400 Toulouse

CLS / ARGOS

8-10 rue Hermès – Parc Technologique du Canal – 31520 Ramonville Saint-Agne

UNIVERSITE TOULOUSE III – PAUL SABATIER

ECOLE DOCTORALE SCIENCES DE L'UNIVERS, DE L'ENVIRONNEMENT ET DE L'ESPACE
Formation Doctorale Océan Atmosphère Environnement

THESE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE TOULOUSE III

Discipline : Traitement d'Image en Océanographie Spatiale

présentée et soutenue par

Claire POTTIER

le 22 décembre 2006.

Combinaison multi-capteurs de données de couleur de l'eau : Application en océanographie opérationnelle

Directrice de thèse : Véronique Garçon, Directeur de Recherches CNRS, *LEGOS*

JURY

Serge Chauzy, Professeur, <i>Université Toulouse III</i> ,	Président
Nicolas Hoepffner, Senior Scientist, <i>JRC Ispra (Italie)</i> ,	Rapporteur
Hussein Yahia, Chargé de Recherches, <i>INRIA Rocquencourt</i> ,	Rapporteur
Gilles Larnicol, Ingénieur d'études, <i>CLS Toulouse</i> ,	Examineur
Pierre-Yves Le Traon, Cadre de Recherche IIIA, <i>Ifremer Brest</i> ,	Examineur
Eric Thouvenot, Ingénieur, <i>CNES Toulouse</i> ,	Examineur
Antonio Turiel, Chercheur, <i>ICM Barcelone (Espagne)</i> ,	Examineur

Laboratoire d'Etudes en Géophysique et Océanographie Spatiales (LEGOS)

UMR 5566 – CNRS/IRD/UPS/CNES – Observatoire Midi-Pyrénées
14 avenue Edouard Belin – 31400 Toulouse

CLS / ARGOS

8-10 rue Hermès – Parc Technologique du Canal – 31520 Ramonville Saint-Agne

REMERCIEMENTS

Je remercie chaleureusement les membres de mon jury, Serge Chauzy, Nicolas Hoepffner, Hussein Yahia, Gilles Larnicol, Pierre-Yves Le Traon, Eric Thouvenot, Antonio Turiel et Véronique Garçon, pour avoir lu, commenté et corrigé mon travail. Merci aussi d'avoir été présent le jour de ma soutenance de thèse, particulièrement les plus éloignés : il est souvent plus sportif que l'on croit d'aller d'un endroit à l'autre... même (et surtout !) en avion !

Je tiens à remercier tout particulièrement Véronique Garçon. Tout d'abord, car cette thèse n'aurait pas vu le jour sans elle. Toujours présente, elle m'a admirablement bien encadrée durant ces trois années de thèse, m'a fait partager sa grande expérience dans le domaine de l'océanographie, m'a guidée tout au long de ce travail. Merci pour la patience dont elle a fait preuve (là, je pense surtout aux divers écrits et oraux...). Je la remercie aussi pour la confiance qu'elle m'a accordée et ses encouragements tout au long de ma thèse. Elle m'a transmis sa rigueur scientifique, son amour du travail bien fait, sa volonté d'aller toujours plus loin... Ces trois années furent très riches à beaucoup de niveaux, et je l'en remercie sincèrement.

Je remercie aussi Gilles Larnicol pour m'avoir accueillie au sein de son équipe à CLS et Pierre-Yves Le Traon qui était aussi présent à l'origine de cette thèse.

J'exprime toute ma gratitude à Antonio Turiel, pour m'avoir si bien initiée au monde des ondelettes et pour m'avoir si bien encadrée et conseillée dans le domaine du traitement d'images. Ça a été un réel plaisir intellectuel de travailler avec lui.

Je voudrais vivement remercier Joël Sudre pour m'avoir fait partager ses compétences en informatique et traitement d'images, pour sa gentillesse, sa générosité, sa disponibilité, son aide précieuse à des moments critiques et pour m'avoir soutenue et encouragée durant ces trois années.

Je souhaiterais également remercier Nicolas Hoepffner et Watson Gregg pour m'avoir permis de participer à la rédaction du rapport IOCCG sur la combinaison de données couleur

de l'eau. Ce fut extrêmement enrichissant de dialoguer avec des experts mondiaux du domaine.

Je remercie Frédéric Mélin, Philippe Schaeffer et Simon Ruiz, Pierre Mazzega, Jean-Pierre Céron, Marie Drévillon et Sophie Belamari pour leurs avis éclairés et leur aide dans les domaines de la couleur de l'eau, l'analyse objective, les statistiques, les mathématiques, le traitement du signal et l'Océan Austral respectivement.

Il est quelque peu difficile de récupérer sans difficulté des mesures *in situ*. C'est pourquoi je remercie avec ferveur Watson Gregg, Diana Ruiz-Pino, Yves Dandonneau, Gwenaëlle Moncoiffé et Christophe Maes d'avoir répondu présents si rapidement.

Merci également à Isabelle Dadou, Guillaume Charria, Jérôme Llido, Mathilde Faillot, Erika Sternberg, Alix Lombard, Guillaume Ramillien, Mathilde Cauhope et Frédéric Frappart pour leur sympathie, leur soutien et leurs encouragements.

Merci aux collègues de bureau de CLS, Hervé Johannes, Isabelle Pujol, Annabelle Olivier et Charles Desportes, ainsi que Frédéric Montazaud pour leur bonne humeur et leur sympathie.

Merci aux Philippes de CLS (Sicard et Poilbarbe) pour toute l'aide informatique qu'ils m'ont apportée ; à Bruno Buisson, Christian Naskas et Philippe Téchiné pour régler nos petits problèmes informatiques au LEGOS ; à Nadine Lacroux, Catherine Donati et Martine Mena pour les longues discussions et les rires.

Merci à mes parents et mon frère, à Anthony Ouvrard, Justine Dussarrat, Sandrine et Noé Serres, Ludovic Pouyadou, Gaël Rivière, Hervé Dehapiot et tous les autres, et surtout toi, Nicolas, pour avoir été présents, m'avoir encouragée et soutenue (et supportée) durant ces trois années, même (et surtout) dans les moments les plus difficiles.

Merci à tous, du fond du coeur.

TABLE DES MATIERES

INTRODUCTION GENERALE	15
------------------------------------	-----------

CHAPITRE 1 – LA « COULEUR DE L’EAU » : INTERET ET PRINCIPE DE LA MESURE	21
--	-----------

1 – LA « COULEUR DE L’EAU »	25
-----------------------------------	----

1.1– Le phytoplancton	26
-----------------------------	----

1.1.1 – Qu’est-ce que le phytoplancton ?.....	26
---	----

1.1.2 – Pourquoi le phytoplancton est-il important ?	27
--	----

1.1.3 – Conclusion.....	29
-------------------------	----

1.2 – Le carbone absorbé par les océans.....	30
--	----

1.2.1 – Interaction océan / atmosphère.....	30
---	----

1.2.2 – Le cycle du carbone.....	32
----------------------------------	----

1.3 – Conclusion	36
------------------------	----

2 – MESURE DE LA CONCENTRATION EN CHLOROPHYLLE	37
--	----

2.1 – Introduction	37
--------------------------	----

2.2 – Les données satellitaires	37
---------------------------------------	----

2.2.1 – Introduction.....	37
---------------------------	----

2.2.2 – Qu’est-ce que la « couleur de l’eau » ?.....	38
--	----

2.2.3 – Comment obtient-on une image « true color » ?	41
---	----

2.2.4 – SeaWiFS (Sea- viewing Wide Field-of-view Sensor).....	47
--	----

2.2.5 – MODIS (MODerate resolution Imaging Spectroradiometer)	49
---	----

2.2.6 – Produits disponibles et produits utilisés	52
---	----

2.2.7 – Conclusion.....	54
-------------------------	----

2.3 – Les données <i>in situ</i>	55
--	----

2.3.1 – Introduction.....	55
---------------------------	----

2.3.2 – Principe de la mesure	55
-------------------------------------	----

2.3.3 – AMT (Atlantic Meridional Transect, UK)	56
--	----

2.3.4 – SeaBASS (SeaWiFS Bio-optical And Storage System, USA)	57
---	----

2.3.5 – NODC (National Oceanographic Data Center, USA)	60
--	----

2.3.6 – CIEM (Conseil International pour l’Exploitation de la Mer)	60
--	----

2.3.7 – NIO (National Institute of Oceanography, Inde)	60
--	----

2.3.8 – Les campagnes OISO (Océan Indien Service d’Observation).....	61
--	----

2.3.9 – Le programme MINERVE	61
------------------------------------	----

2.3.10 – Le programme ARGAU.....	62
----------------------------------	----

2.3.11 – Le programme DIAPAZON	62
--------------------------------------	----

2.3.12 – Le programme DYFAMED	63
-------------------------------------	----

2.3.13 – Campagnes ponctuelles.....	63
-------------------------------------	----

2.3.14 – Conclusion	63
---------------------------	----

3 – CONCLUSION DU CHAPITRE	64
----------------------------------	----



CHAPITRE 2 – ETUDE DES DONNEES	65
1 – PROVINCES BIOGEOCHIMIQUES	69
2 – UNE REPARTITION LOGNORMALE	71
3 – COMPARAISON DES DONNEES SATELLITAIRES AVEC LES MESURES <i>IN SITU</i> ...	73
3.1 – Introduction.....	73
3.2 – Définition des formules de comparaison par l'exemple	73
3.2.1 – Calcul sur les données « brutes »	74
3.2.2 – Calcul sur le logarithme base 10 des données.....	76
3.2.3 – Conclusion.....	79
3.3 – Résultat des « matchups »	79
3.3.1 – Méthodologie de création des « matchups »	79
3.3.2 – Comparaison des données satellitaires et des observations <i>in situ</i>	81
3.3.3 – Comparaison des comportements des capteurs	81
3.3.4 – Conclusion.....	81
3.4 - Construction des cartes d'erreur et de biais.....	84
3.5 – Conclusion	88
4 – INTERCOMPARAISON DES DONNEES SATELLITAIRES	88
4.1 – Introduction.....	88
4.2 – Méthodologie.....	88
4.3 – Résultats et discussion.....	89
4.4 – Conclusion	90
5 – CONCLUSION DU CHAPITRE	90
CHAPITRE 3 – COMBINAISON DES DONNEES DE CHLOROPHYLLE : MOYENNE PONDEREE ET ANALYSE OBJECTIVE.....	93
1 – MOYENNE PONDEREE	98
1.1 – Introduction.....	98
1.2 – Méthodologie.....	98
1.2.1 – Introduction.....	98
1.2.2 – Approche MP1	99
1.2.3 – Approche MP2.....	100
1.2.4 – Implémentation de l'algorithme	100
1.2.5 – Conclusion.....	102
1.3 – Application aux données de chlorophylle.....	104
1.3.1 – Introduction.....	104
1.3.2 – Les paramètres d'entrée	104
1.3.3 – Les résultats	104
1.3.4 – Comparaison des 4 cas : choix de l'option à retenir	106
1.4 – Conclusion	107

2 – ANALYSE OBJECTIVE.....	107
2.1 – Introduction.....	107
2.2 – Méthodologie.....	108
2.2.1 – Introduction.....	108
2.2.2 – Formulation.....	108
2.2.3 – Implémentation de l'algorithme	109
2.2.4 – Conclusion.....	110
2.3 – Calculs préliminaires	112
2.3.1 – Introduction.....	112
2.3.2 – Climatologie mensuelle.....	112
2.3.3 – Variance spatio-temporelle	112
2.3.4 – Fonction de corrélation	116
2.3.5 – Rayons de corrélations spatiaux.....	122
2.3.6 – Conclusion.....	123
2.4 – Tests de sensibilité.....	123
2.4.1 – Introduction.....	123
2.4.2 – Rayon de corrélation temporel.....	123
2.4.3 – Centrage des données	125
2.4.4 – Nombre minimum d'observations dans la bulle d'influence.....	126
2.4.5 – Nombre maximum d'observations dans la bulle d'influence	126
2.4.6 – Sous-échantillonnage des positions d'étude	127
2.4.7 – Rayons de corrélations spatiaux.....	127
2.4.8 – Erreur de mesure.....	128
2.4.9 – Biais.....	128
2.4.10 – Variance	129
2.4.11 – Coefficient de la fonction de corrélation	129
2.4.12 – Conclusion.....	130
2.5 – Application aux données de chlorophylle.....	130
2.6 – Conclusion	132
3 – COMPARAISON ENTRE LES DEUX METHODES.....	133
3.1 – Introduction.....	133
3.2 – Comparaison avec les données <i>in situ</i>	133
3.3 – Comparaison des tendances temporelles et latitudinales	134
3.3.1 – Méthodologie	134
3.3.2 – Résultats	135
3.4 – Conclusion	138
4 – CONCLUSION DU CHAPITRE	139
CHAPITRE 4 – COMBINAISON DES DONNEES DE CHLOROPHYLLE : APPROCHE PAR ONDELETTES.....	141
1 – GENERALITES SUR LA THEORIE DES ONDELETTES.....	145
1.1 – Introduction par la musique	145
1.2 – Transformée en ondelettes.....	146
1.2.1 – Introduction.....	146



1.2.2 – Famille d'ondelettes.....	146
1.2.3 – Transformée en ondelettes continue.....	147
1.2.4 – Transformée en ondelettes discrète.....	147
1.2.5 – Conclusion.....	148
1.3 – Multi-résolution : cas des bases orthonormales	149
1.3.1 – Introduction.....	149
1.3.2 – Définition de l'analyse multi-résolution	149
1.3.3 – Approximation et fonction échelle	150
1.3.4 – Filtres miroirs conjugués.....	151
1.3.5 – Espace des détails et ondelettes	152
1.3.6 – Conclusion.....	153
1.4 – Ondelettes et bancs de filtres	154
1.4.1 – Introduction.....	154
1.4.2 – Algorithme pyramidal ou transformée en ondelettes orthogonale rapide.....	154
1.4.3 – Cas des signaux finis ou problèmes de bords.....	155
1.4.4 – Formulation matricielle.....	156
1.4.5 – Cas de la dimension 2	158
1.4.6 – Conclusion.....	161
1.5 – Exemple de bases d'ondelettes orthonormales	161
1.5.1 – Introduction.....	161
1.5.2 – Choix de l'ondelette	161
1.5.3 – L'ondelette de Haar	162
1.5.4 – Les ondelettes à support compact d'Ingrid Daubechies	163
1.5.5 – Les Symmlets.....	166
1.5.6 – Les Coiflets.....	166
1.5.7 – L'ondelette de Shannon.....	167
1.5.8 – Les ondelettes splines de Battle-Lemarié	168
1.5.9 – Conclusion.....	170
1.6 – Conclusion	170
2 – METHODOLOGIE.....	171
2.1 – Introduction.....	171
2.2 – Estimation des données manquantes.....	171
2.2.1 – Introduction.....	171
2.2.2 – Statistiques à l'origine de la méthode.....	172
2.2.3 – Principe de l'estimateur par ondelette ou calcul de l'inférence fils → père	178
2.2.4 – Implémentation de l'algorithme	180
2.2.5 – Tests.....	189
2.2.6 – Conclusion.....	191
2.3 – Fusion d'images par ondelettes.....	191
2.4 – Application à la combinaison des données de chlorophylle	193
3 – CONCLUSION DU CHAPITRE	195

CHAPITRE 5 – APPLICATION OCEANOGRAPHIQUE : LES ONDES DOMINANTES DE L'OCEAN AUSTRAL.....	197
1 – DESCRIPTIF DE LA ZONE D'ETUDE ET PROBLEMATIQUE.....	201
2 – METHODES D'ANALYSE.....	205
3 – APPLICATION AUX DONNEES ALTIMETRIQUES.....	209
3.1 – Introduction.....	209
3.2 – Les données utilisées.....	210
3.3 – Résultats.....	210
3.4 – Discussion et conclusion.....	216
4 – APPLICATION AUX DONNEES DE CHLOROPHYLLE.....	219
4.1 – Introduction.....	219
4.2 – Application aux données mensuelles issues de SeaWiFS.....	219
4.2.1 – Les données utilisées.....	219
4.2.2 – Résultats.....	220
4.3 – Application aux données hebdomadaires issues de la combinaison de SeaWiFS et MODIS/Aqua.....	225
4.3.1 – Les données utilisées.....	225
4.3.2 – Résultats.....	225
4.4 – Discussion et conclusion.....	231
5 – CONCLUSION DU CHAPITRE.....	237
 CONCLUSION GENERALE.....	 239
 BIBLIOGRAPHIE.....	 247
 ANNEXE A.....	 259
1 – MATCHUPS PAR PROVINCE.....	261
1.1 – SeaWiFS.....	261
1.2 – MODIS/Aqua.....	263
2 – ANALYSE OBJECTIVE.....	265
2.1 – Les cartes de climatologie mensuelle.....	265
2.2 – Tests de sensibilité.....	267
2.2.1 – Cartes pour le 25 mars 2005.....	267
2.2.2 – Rayon de corrélation temporel.....	267
2.2.3 – Centrage des données.....	268
2.2.4 – Nombre minimum d'observations dans la bulle d'influence.....	269
2.2.5 – Nombre maximum d'observations dans la bulle d'influence.....	269
2.2.6 – Sous-échantillonnage des positions d'étude.....	270



2.2.7 – Rayons de corrélations spatiaux.....	271
2.2.8 – Erreur de mesure.....	272
2.2.9 – Biais.....	273
2.2.10 – Variance	274
2.2.11 – Coefficient de la fonction de corrélation	275
3 – TRANSFORMEE EN ONDELETTES : TESTS.....	276
4 – MISSIONS SATELLITAIRES ALTIMETRIQUES	278
4.1 – Intérêt de la mesure.....	278
4.2 – Principe de la mesure.....	278
4.3 – La mission TOPEX-POSEIDON	281
4.3.1 – Introduction.....	281
4.3.2 – Le satellite	281
4.3.3 – L'orbite.....	283
4.3.4 – Le traitement des données	284
4.3.5 – Les produits TOPEX-POSEIDON	285
4.4 – Les missions ERS-1/2	285
4.5 – La mission Jason-1.....	287
4.5.1 – Introduction.....	287
4.5.2 – Le satellite	287
4.5.3 – Le segment sol	289
4.6 – La mission Envisat.....	289
5 – RESULTATS COMPLEMENTAIRES DES ANALYSES DE CERON ET D'HAYASHI... 272	
5.1 – Analyses entre septembre 1997 et mai 2006.....	290
5.1.1 – Les données utilisées	290
5.1.2 – Maxima des DSP obtenues par l'analyse d'Hayashi pour l'altimétrie	290
5.1.3 – Résultats de l'analyse d'Hayashi pour le couple 2/0	291
5.1.4 – Résultats de l'analyse de Céron appliquée aux données mensuelles SeaWiFS de concentration en chlorophylle pour le couple 2/0.....	292
5.1.5 – Résultats de l'analyse de Céron appliquée aux données altimétriques pour le couple 2/0	293
ANNEXE B – PUBLICATIONS	295
1 - C. POTTIER, J.-P. CÉRON, J. SUDRE, I. DADOU, S. BELAMARI ET V. GARÇON, 2004: DOMINANT PROPAGATING SIGNALS IN SEA LEVEL ANOMALIES IN THE SOUTHERN OCEAN, <i>GEOPHYSICAL RESEARCH LETTERS</i> , VOL. 31, 2004.	297
2 - C. POTTIER, V. GARÇON, G. LARNICOL, J. SUDRE, P. SCHAEFFER, ET P.Y. LE TRAON, 2006: MERGING SEAWIFS AND MODIS/AQUA OCEAN COLOR DATA IN NORTH AND EQUATORIAL ATLANTIC USING WEIGHTED AVERAGING AND OBJECTIVE ANALYSIS, <i>IEEE TRANSACTION ON GEOSCIENCES AND REMOTE SENSING</i> , VOLUME 44, ISSUE 10, PART 1.....	305

Introduction

Générale

INTRODUCTION GENERALE

L'Océan... Cet univers fascinant occupe environ les 3/4 de la surface de notre planète, soit plus de 99% de l'espace habitable sur Terre. De nombreux organismes y ont élu domicile, dont le plus gros jamais recensé, la baleine bleue, pouvant mesurer plus de 35 mètres et peser plus de 200 tonnes. Mais l'aspect de la vie dans l'océan qui nous intéresse ici est d'une toute autre taille. En effet, l'objet de notre étude est le phytoplancton, ces végétaux microscopiques en suspension qui dérivent avec les courants à travers les océans et qui jouent un rôle absolument primordial pour toute vie sur notre planète. Sans phytoplancton, la vie sur Terre telle que nous la connaissons ne serait jamais apparue. Sans lui, elle cesserait d'exister. En effet, non seulement le plancton océanique est la base de la chaîne alimentaire dans l'océan, mais il est un acteur de premier plan dans le cycle du carbone sur Terre, de par l'absorption du dioxyde de carbone au cours de la photosynthèse permettant la séquestration d'une partie de ce gaz dans l'océan profond en tant que carbone sédimentaire durant plusieurs siècles [Behrenfeld *et al*, 2002].

L'analyse du phytoplancton permet, en particulier, de déterminer et d'étudier les zones de l'océan dites « puits de CO₂ », connaissance d'autant plus importante que les émissions de dioxyde de carbone dues aux activités humaines ne cessent d'augmenter. Comme le phytoplancton participe à cette diminution de CO₂ atmosphérique (ce phénomène est appelé « pompe biologique »), de longues séries temporelles complètes de données biologiques sont indispensables pour effectuer la surveillance et la prévision de l'environnement marin. C'est dans cette optique que les programmes européens comme GMES (Global Monitoring for Environment and Security) ou MERSEA (Marine Environment and Security for the European Area) ont ajouté une composante biogéochimique aux systèmes opérationnels de prévision océanique tels que celui élaboré par le programme MERCATOR Océan en France. De plus, les modèles utilisés dans ce cadre ne sont pas encore assez précis et sont donc continuellement validés par comparaison avec les observations.

Les données de chlorophylle-a, pigment principal du phytoplancton océanique, sont également cruciales pour répondre aux demandes socio-économiques. Par exemple, elles peuvent être utiles à la gestion de la pêche car elles sont nécessaires à la modélisation des écosystèmes qui permettent l'estimation des stocks. Elles peuvent aussi servir d'outil d'aide



à la pêche dans le cadre de recommandation de zones de pêche, puisque les bancs de poissons sont souvent associés aux zones riches en phytoplancton.

Si les campagnes en mer offrent la possibilité d'acquérir des données à haute fréquence (quelques heures voire quelques minutes entre 2 mesures consécutives) et à petite échelle spatio-temporelle (quelques kilomètres voire quelques mètres entre 2 mesures consécutives), l'observation spatiale, grâce à sa couverture spatio-temporelle, permet de suivre la variabilité de la couleur de l'océan pour des échelles allant de la journée à la décennie, et de la méso-échelle à l'océan global. En effet, les instruments embarqués sur satellites mesurent la réponse spectrale de la surface océanique, en exploitant un intervalle de fréquences allant du visible au proche infrarouge [Mobley, 1994]. Après calibrations, corrections atmosphériques et autres traitements, les radiances correspondant au signal océanique sont obtenues : c'est ce qu'on appelle la « couleur de l'eau ». Elles sont ensuite converties en concentration en chlorophylle-*a* par des algorithmes empiriques [O'Reilly *et al*, 1998].

Neuf satellites embarquant des capteurs couleur de l'eau ont déjà été en orbite, dont le plus ancien, CZCS (Coastal Zone Color Scanner, sur Nimbus-7 (NASA, USA)) depuis le 24 octobre 1978 et jusqu'au 22 juin 1986. Puis il n'y eut aucun capteur pendant plus de dix ans. Actuellement, neuf satellites embarquant des capteurs couleur de l'eau sont en orbite, pour des applications à l'échelle globale ou régionale, dont les deux auxquels nous nous sommes intéressés : SeaWiFS (Sea-viewing Wide Field-of-view Sensor, sur OrbView-2 (NASA, USA), depuis le 1^{er} août 1997) et MODIS/Aqua (MODerate resolution Imaging Spectroradiometer, sur Aqua (NASA, USA), depuis le 4 mai 2002). Sept autres missions sont aussi déjà planifiées, les prochaines missions, régionale (GOCI sur COMS-1 (KARI/KORDI, Corée)) et globale (VIIRS sur NPP (NASA, USA)) sont prévues pour 2008.

Chaque mission satellitaire qui mesure la couleur de l'eau est limitée en couverture océanique par la distance entre deux traces du satellite, sa répétitivité et par des phénomènes atmosphériques et de surface faussant la vision de l'océan, en particulier les nuages et l'éblouissement du capteur par la réflexion du soleil sur la surface océanique (« effet miroir »). Par exemple, considérant des conditions moyennes de présence de nuages et d'éblouissement, SeaWiFS peut couvrir environ 15% de la couverture océanique en un jour [Gregg *et al*, 1998]. Cette couverture journalière peut augmenter jusqu'à 25% en combinant les jeux de données issues de trois satellites. Ainsi, non seulement la confiance statistique dans les données est augmentée, mais l'étude des phénomènes physiques et

biologiques (comme la production primaire et les floraisons d'algues, par exemple) est aussi rendue plus facile avec des cartes complètes.

Les capteurs satellitaires ont des caractéristiques différentes, tant au niveau des bandes spectrales que des calibrations et des algorithmes de traitement, rendant difficile la combinaison des données couleur de l'eau. En particulier, de nombreuses études ont contribué à réconcilier SeaWiFS et MODIS/Aqua. Par exemple, SeaWiFS est équipé d'un instrument optique de polarisation, induisant une sensibilité à la polarisation négligeable pour lui. Comme ce n'est pas le cas pour MODIS/Aqua, il est obligatoire d'appliquer une correction de polarisation à ses données pour qu'il y ait concordance entre les mesures des deux capteurs [Meister *et al*, 2005]. Un travail impressionnant d'intercomparaison et d'intercalibration entre SeaWiFS et MODIS/Aqua a aussi été réalisé dans le cadre du Projet SIMBIOS de la NASA [McClain *et al*, 2002; Kwiatkowska, 2003; Franz, 2003].

L'objectif du travail présenté ici est de concevoir, développer et tester des méthodes de combinaison de données couleur de l'eau haute résolution (quelques kilomètres), provenant des capteurs SeaWiFS et MODIS/Aqua, afin d'élaborer des produits de couleur de l'eau pour des applications en temps réel relevant de l'océanographie opérationnelle.

La combinaison des données peut se faire au niveau des radiances ou au niveau des produits dérivés comme la chlorophylle-a. Maritorena et Siegel [2005] utilisent un modèle semi-analytique (GSM01) pour combiner les radiances de SeaWiFS et MODIS/Aqua et produire globalement trois composantes biogéochimiques, dont la chlorophylle-a. Gregg et Conkright [2001] ont utilisé une « blended analysis » pour combiner les mesures *in situ* (ou mesures terrain) et les données de chlorophylle issues du capteur CZCS pour construire une représentation saisonnière moyenne de la distribution chlorophyllienne globale. Dans le cadre du projet SIMBIOS, plusieurs méthodes ont été testées, du simple « binning » (la valeur en un pixel ou « bin » de la grille des données combinées est la moyenne de toutes les données journalières de chlorophylle issues de plusieurs sources à l'intérieur de ce pixel) à des méthodes plus ou moins sophistiquées [Kwiatkowska and Fargion, 2001, 2002a, 2002b].

Dans le cadre de cette étude, nous avons décidé de nous intéresser directement aux données de concentration en chlorophylle, puisqu'il s'agit de la variable que nous souhaitons cartographier.



Trois approches ont été retenues :

- ☞ la moyenne pondérée par l'erreur de mesure de chaque capteur : c'est une méthode simple à mettre en œuvre, ne considérant que les données existantes, robuste, mais qui présente certaines limitations,
- ☞ l'interpolation optimale : c'est une méthode statistique considérant le voisinage des pixels qui a déjà fait ses preuves en océanographie pour des variables telles que la hauteur de mer [Le Traon *et al*, 1998], la température (Reynolds et Smith [1994] ; Ruiz et Larnicol [2004] dans le cadre du projet Medspiration) et la salinité (dans le cadre de la mission ESA SMOS ; <http://earth.esa.int/symposia/smos04/presentations/SynergySMOSAqua.pdf>),
- ☞ l'analyse en ondelettes : c'est une méthode nouvelle et plus exploratoire.

Ce travail s'est déroulé dans le cadre du projet européen intégré MERSEA (<http://www.mersea.eu.org/>). MERSEA a pour objectif de développer, d'ici à 2008, un système opérationnel pour la surveillance (temps réel) et la prévision (de quelques jours à quelques mois) de la physique océanique, de la biogéochimie et des écosystèmes, à des échelles globales et régionales. Au sein du Workpackage 2 dédié aux activités de télédétection, la tâche 2.3 est directement liée à la couleur de l'eau. Une partie du travail a justement consisté à développer et tester différentes méthodes de combinaison de données couleur de l'eau à partir de celles issues de plusieurs capteurs, afin d'obtenir des données combinées haute résolution et le moins entachées d'erreur possible.

Ce manuscrit comporte 5 chapitres. Le 1^{er} chapitre introduit la notion de concentration en chlorophylle et présente les deux types de données utilisées, *in situ* et satellitaires. Le 2^e chapitre est une étude quantitative et qualitative des données utilisées. Au chapitre 3, la moyenne pondérée par l'erreur et l'analyse objective, qui sont les 2 premières méthodes utilisées pour la combinaison des données couleur de l'eau, sont décrites, testées et comparées de façon globale pour l'année 2003. Le chapitre 4 montre les travaux concernant la 3^e méthode, basée sur la transformée en ondelettes. Le chapitre 5 montre l'intérêt de la combinaison multi-capteurs de données couleur de l'eau au travers de l'étude des ondes dominantes dans la zone de l'Océan Austral. Nous terminons avec une conclusion générale ainsi que les perspectives envisagées.

Chapitre 1

La couleur de l'eau :

intérêt et principe de la mesure

CHAPITRE 1

LA « COULEUR DE L'EAU » :

INTERET ET PRINCIPE DE LA MESURE

Nous dépendons des ressources océaniques pour de nombreuses choses : nourriture, eau, transport, loisirs, minéraux et énergie. Un changement majeur de la circulation océanique globale s'accompagne généralement d'un changement dans le climat, causant ainsi des événements extrêmes tels que sécheresses, inondations, ouragans, etc. L'étude des océans est primordiale pour la compréhension de l'évolution de notre planète.

Ce chapitre est destiné à nous familiariser avec la donnée océanique sur laquelle porte cette étude : la couleur de l'eau. Nous commençons par présenter son contenu biologique, ainsi que le rôle joué par sa composante la plus intéressante : le phytoplancton. Après avoir introduit les différentes méthodes de mesure, nous montrons les données utilisées.

Sommaire

1 – LA « COULEUR DE L'EAU »	25
1.1 – <i>Le phytoplancton</i>	26
1.2 – <i>Le carbone absorbé par les océans</i>	30
1.3 – <i>Conclusion</i>	36
2 – MESURE DE LA CONCENTRATION EN CHLOROPHYLLE.....	37
2.1 – <i>Introduction</i>	37
2.2 – <i>Les données satellitaires</i>	37
2.3 – <i>Les données in situ</i>	55
3 – CONCLUSION DU CHAPITRE.....	64

1 - La « couleur de l'eau »

La « couleur » de l'océan est déterminée par les interactions de la lumière avec l'eau. Nous voyons de la couleur quand la lumière est réfléchiée par les objets autour de nous. La lumière blanche est composée d'une combinaison (ou « spectre ») de couleurs, que les gouttes d'eau séparent en un arc en ciel. Quand la lumière atteint la surface d'un objet, les différentes couleurs peuvent être absorbées, transmises, dispersées ou réfléchies selon différentes intensités. La couleur que nous voyons dépend des couleurs réfléchies. Par exemple, un livre qui nous apparaît rouge absorbe la plupart du vert et du bleu de la lumière blanche qui l'éclaire, et réfléchit le rouge. La lumière qui est dispersée ou transmise par la plupart des objets n'est souvent pas visible à nos yeux.

Les substances dans l'eau de mer qui affectent la couleur réfléchiée sont le phytoplancton, les particules inorganiques, la matière organique dissoute et l'eau elle-même [Feldmann, 1994]. Le phytoplancton contient de la chlorophylle, qui absorbe le rouge (620-700 nm) et le bleu (446-500 nm) et réfléchit le vert (500-578 nm). Les particules peuvent réfléchir et absorber de la lumière, ce qui réduit la clarté (transmission de la lumière) de l'eau. La matière organique dissoute absorbe fortement la lumière bleue, et sa présence peut interférer, de ce fait, avec les mesures de chlorophylle.



Figure 1.1 – Courant des Malouines (à l'est de l'Argentine), le 27 novembre 2001. Les eaux turquoise (probablement un bloom phytoplanctonique de coccolithophoridés) contrastent clairement avec les eaux froides bleu foncé plus loin de la côte. Les îles en bas de l'image sont les Îles Falkland.



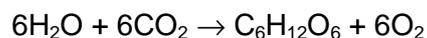
Figure 1.2 – Eaux turbides. A gauche, James Bay (Canada) le 26 octobre 1998. A droite, Mer de Azov (entre l'Ukraine et la Russie) le 9 octobre 1998.

Quand nous regardons l'océan ou que nous l'observons de l'espace, nous voyons que l'océan est bleu car l'eau absorbe le rouge et réfléchit le bleu. En utilisant des instruments qui sont plus sensibles que l'œil humain, nous pouvons mesurer un large panel de dégradés de bleu-verts, qui révèlent la présence de différentes quantités de phytoplancton (Figure 1.1), sédiments et matière organique dissoute (Figure 1.2).

1.1 - Le phytoplancton

1.1.1 - Qu'est-ce que le phytoplancton ?

Le phytoplancton utilise la lumière du soleil et le dioxyde de carbone pour produire du carbone organique. Ce procédé est appelé « photosynthèse » et sa formule est la suivante :



(i.e. 6 molécules d'eau plus 6 molécules de dioxyde de carbone produisent 1 molécule de sucre plus 6 molécules de dioxygène). La photosynthèse est rendue possible car le phytoplancton contient des amas de pigments, ou unités photosynthétiques, qui piègent l'énergie de la lumière du soleil. Le pigment principal, la chlorophylle, donne la couleur verte à la plupart des plantes marines et terrestres.

Mesurer la couleur d'une zone permet d'estimer la quantité de phytoplancton dans cette zone. De l'eau très pure apparaît bleue foncée, presque noire. De l'eau très productive avec beaucoup de phytoplancton apparaît bleu-vert. Les Figures 1.1, 1.2 et 1.5 montrent un tel changement de couleur.

De l'espace, les variations de la couleur de l'eau peuvent être mesurées avec des instruments adéquats, par exemple des radiomètres ou des spectromètres. Les espèces phytoplanctoniques contiennent différents types de pigments en plus de la chlorophylle (par exemple : caroténoïdes, xanthophylles). L'association de ces pigments à l'intérieur de la cellule représente une vraie signature optique de certains groupes phytoplanctoniques qui peut être reconnu par des instruments optiques de terrain ou satellitaires (cf. partie 2.2.2).

1.1.2 - Pourquoi le phytoplancton est-il important ?

Le phytoplancton se situe au début de la chaîne alimentaire océanique. Alors qu'il croît et se multiplie, de petits poissons et autres animaux se nourrissent de lui, qui sont à leur tour mangés par de plus grands et ainsi de suite. La concentration en phytoplancton permet d'estimer la productivité des océans. Les pêcheurs trouvent des zones propices à la pêche en regardant les données de couleur de l'eau pour localiser les endroits riches en phytoplancton : les cartes de concentration en chlorophylle sont l'un des paramètres océaniques distribués par CLS via le logiciel CatSat aux flotilles de pêche (Figure 1.3).

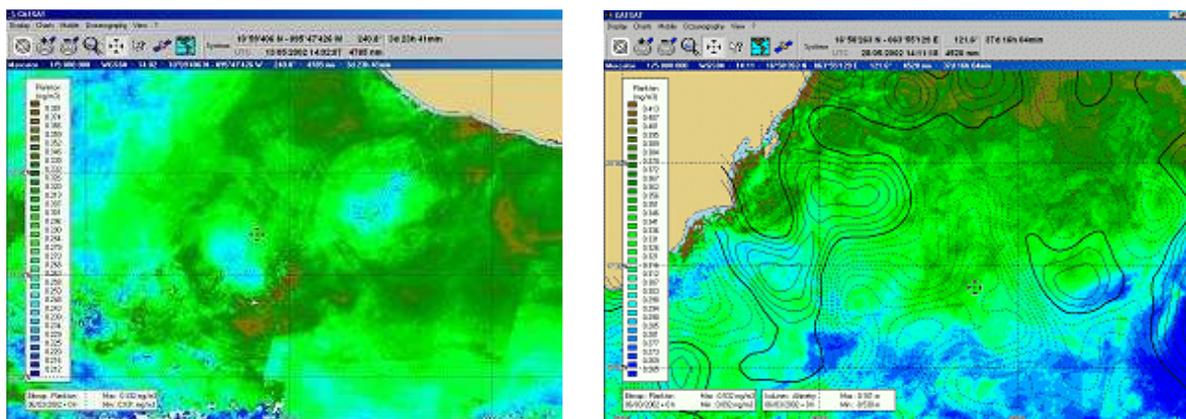


Figure 1.3 – CatSat, logiciel d'aide aux pêcheurs de CLS. A gauche : visualisation des données de concentration en chlorophylle. A droite : superposition d'une carte d'altimétrie avec une carte de concentration en chlorophylle, permettant de voir, par exemple, si un tourbillon de chlorophylle est au même endroit que sa signature altimétrique.

Regarder la couleur de l'océan nous renseigne aussi sur la chimie de l'océan. En effet, en plus de la lumière et du dioxyde de carbone, la croissance du phytoplancton est limitée par la disponibilité des nutriments (azote, phosphore, silice, fer). Ainsi, la cartographie spatio-temporelle des structures de concentration en chlorophylle océanique à l'échelle du globe a permis de montrer dans un premier temps que la production océanique est très faible au niveau des grands gyres subtropicaux dans les océans Pacifique, Atlantique et Indien (« 1 » sur la Figure 1.4). Ces régions océaniques présentent ces caractéristiques désertiques car

elles sont oligotrophes en raison de la courantologie, c'est-à-dire qu'elles sont pauvres en nutriments. D'autres zones du globe sont au contraire très productives à l'instar des régions côtières du fait de la proximité de zones de résurgence des substances nutritives et des conditions de circulation favorables à l'accumulation de nourriture (par exemple, les zones d'upwelling¹ ou les marges continentales). Le gyre subpolaire Atlantique est aussi caractérisé par une forte floraison, liée à un apport important de nutriments en hiver, lorsque la couche de mélange est très profonde, et à la restratification de la colonne d'eau au printemps. Il existe aussi des régions dites HNLC (High Nutrient Low Chlorophyll), comme l'océan Pacifique subarctique et équatorial, et l'océan Austral. Ces régions présentent une très faible concentration en chlorophylle alors que les nutriments (azote, phosphore, silice) ne manquent pas. Ce phénomène semble être cause par un déséquilibre stœchiométrique (par ex., insuffisance relative en fer) et/ou un fort taux de broutage du phytoplancton par le micro-zooplancton.

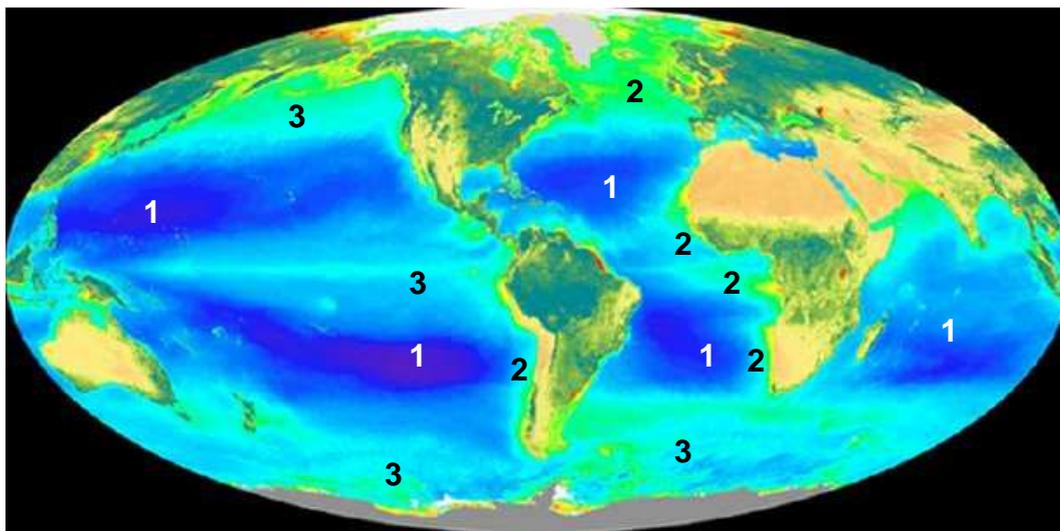


Figure 1.4 – 1 : Déserts océaniques (gyres subtropicaux). 2 : Zones productives (upwellings, plateaux, etc.). 3 : Zones HNLC.

Les données couleur de l'eau constituent également un bon moyen pour connaître la santé de l'océan, en particulier pour déterminer les zones pouvant être atteintes par des pollutions d'origine humaine ou naturelle. Par exemple, il est possible de cartographier d'autres algues, comme les cyanobactéries (algues « bleues ») dont certaines peuvent être toxiques (Figure 1.5). La surveillance dans le temps et l'espace de la distribution de ces algues à la surface de l'eau permet de déterminer les zones à risque et de mettre en place

¹ Remontée des eaux profondes.

des mesures de précaution visant l'exploitation des ressources potentiellement affectées. Autre exemple : mises en relation avec d'autres données océaniques, comme la température de surface, les données de concentration en chlorophylle ont été utilisées lors d'un programme des Nations Unies pour l'Environnement (PNUE) pour la détection de zones côtières libanaises soumises à des apports polluifs terrigènes et anthropiques (<http://www.grid.unep.ch/activities/capacitybuilding/lebanon/index.fr.php>). Dans cette région, si une forte concentration en chlorophylle est liée à une basse température de surface de l'océan, alors on peut penser que les nutriments nécessaires à la croissance du phytoplancton ont une origine naturelle, provenant d'eaux froides issues d'upwelling. Au contraire, aucun lien température / chlorophylle n'est constatable si l'origine des nutriments est polluive. D'autres conditions peuvent aussi affecter la croissance du phytoplancton, comme les changements saisonniers. Puisque le phytoplancton a des besoins spécifiques pour croître, il est souvent le premier à être affecté par les changements, mêmes subtiles, de son environnement (température, salinité, mélange océanique, etc.).

Puisque le phytoplancton dérive avec l'eau et peut donc être considéré comme un traceur passif, la dynamique du phytoplancton peut aussi être utilisée pour l'étude des courants océaniques.



Figure 1.5 – Bloom toxique de cyanobactéries, dans la Mer Baltique. De tels blooms épuisent les concentrations de nutriments dans la Mer Baltique. De plus, les stocks de poissons s'effondrent dans cette région, car ces algues consomment l'oxygène dont les poissons ont besoin pour survivre.

1.1.3 - Conclusion

Nous avons vu dans un premier temps que le phytoplancton est la composante principale de la couleur de l'eau. Ces algues microscopiques contiennent des pigments verts, la chlorophylle. En utilisant des instruments adéquats, mesurant les couleurs, il est possible de suivre les évolutions de la concentration en chlorophylle, de caractériser sa variabilité spatio-temporelle, etc.

Nous avons aussi constaté que la couleur de l'eau est un outil de recherche précieux pour l'étude de la biologie, la chimie et la physique océaniques. En effet, le phytoplancton est non seulement à la base de la chaîne alimentaire mais sa dynamique spatiale et temporelle donne de nombreuses informations : secteurs propices pour la pêche, régions polluées, zones de changement environnemental.

Nous allons maintenant nous intéresser à l'absorption du dioxyde de carbone atmosphérique par l'océan et au rôle de première importance joué par le phytoplancton.

1.2 - Le carbone absorbé par les océans

1.2.1 - Interaction océan / atmosphère

La surface de la mer est le lieu de transmission d'énergie et de matière de l'océan vers l'atmosphère, et du milieu extérieur (atmosphère, espace) vers l'océan.

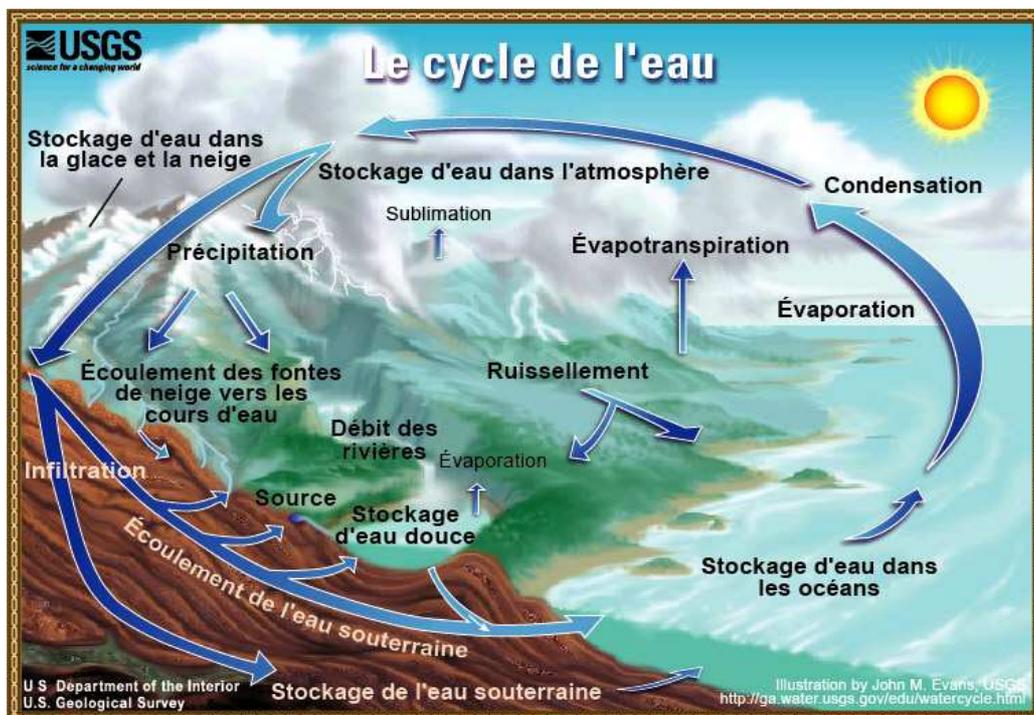


Figure 1.6 – Le cycle de l'eau.

La surface océanique est la source essentielle d'eau atmosphérique. La vapeur d'eau est le principal gaz à effet de serre, et les nuages, contribuant au cycle de l'eau (Figure 1.6), arrêtent partiellement le rayonnement solaire. Le rôle de l'océan n'est pas seulement d'être une source d'eau, mais aussi de chaleur : en effet, l'inertie thermique de l'eau est bien plus élevée que celle de l'air ou du sol, ce qui ralentit considérablement le refroidissement saisonnier. De plus, les courants marins tels que le Gulf Stream ou le Kuro Shio ralentissent les refroidissements hivernaux de température et adoucissent les contrastes avec la latitude.

La source essentielle d'énergie pour la Terre est le Soleil. Celui-ci émet un rayonnement centré dans le domaine du visible, qui traverse partiellement l'atmosphère avant de toucher la surface. Celle-ci réfléchit une partie de ce rayonnement (selon son albédo¹) et absorbe le reste. La pénétration de la lumière peut atteindre plusieurs dizaines de mètres dans l'océan, permettant un réchauffement des couches supérieures de la mer, alors que seule la surface du sol est chauffée par le Soleil, cette chaleur pénétrant ensuite par conduction.

L'atmosphère n'est pas totalement transparente au rayonnement solaire : les molécules de gaz composant l'atmosphère (azote : 78% ; oxygène : 21% ; autres dont argon, dioxyde de carbone, néon, hélium : 1%), les gouttelettes et cristaux nuageux, et les poussières (aérosols) ont des pouvoirs significatifs d'absorption dans l'infrarouge (gaz et nuages), de diffusion et réflexion. Ainsi, l'atmosphère est elle-même chauffée par le Soleil et elle émet à son tour un rayonnement dans l'infrarouge thermique (vers le bas et vers le haut) ; les nuages réfléchissent vers le haut (espace) une partie du rayonnement, absorbent, réémettent et diffusent le reste dans toutes les directions. Le bilan radiatif de la basse atmosphère tient compte de tous ces rayonnements. Par exemple, lors d'une nuit claire et sèche, l'atmosphère contribue très peu à réchauffer la surface de la Terre qui, de ce fait, se refroidit radiativement. A l'inverse, la présence de vapeur d'eau et de gaz carbonique (gaz à effet de serre²), qui absorbent une partie importante de rayonnement et le réémettent, contribuent au réchauffement de la basse atmosphère et de la surface.

Mais deux autres flux importants à la surface s'ajoutent à ces flux radiatifs : les flux de chaleur sensible³ et de chaleur latente¹. Lorsque la température de la surface est plus élevée

¹ L'albédo du système Terre-Atmosphère est la fraction de l'énergie solaire qui est réfléchiée vers l'espace. Sa valeur est comprise entre 0 et 1. Plus une surface est réfléchissante, plus son albédo est élevé. Les éléments qui contribuent le plus à l'albédo de la Terre sont les nuages, les surfaces de neige et de glace, les aérosols.

² Phénomène thermique bien connu sur Terre et sur Vénus, où l'atmosphère laisse passer une partie du rayonnement ultraviolet du Soleil qui vient frapper le sol. Réchauffé, celui-ci émet un rayonnement infrarouge qui est en partie ou totalement piégé par l'atmosphère rendue « imperméable » par la présence, entre autres, d'une forte proportion de CO₂. Il y a alors réchauffement global de la planète.

³ Chaleur émise ou absorbée par un milieu dont la température croît ou décroît en conséquence, mais sans que ce milieu subisse un changement de phase.

que celle de l'atmosphère au-dessus, l'excès de chaleur est transféré vers le haut par conduction et, surtout, par la turbulence ; ce transfert est d'autant plus fort que l'écart de la température est élevé et que le vent souffle fort. De même, l'évaporation à la surface croît avec l'écart d'humidité entre la surface et la première couche de l'atmosphère, et augmente avec le vent. L'évaporation se traduit localement par un refroidissement de la surface, mais la vapeur d'eau, en se condensant plus haut dans l'atmosphère, restitue la chaleur latente à l'atmosphère.

Le bilan d'énergie de surface de l'océan a les caractéristiques suivantes :

- ☞ Energie perdue par l'océan : les rayonnements infrarouge (dépend de sa température de surface ; affecte une pellicule de l'ordre du micromètre), les chaleurs sensible et latente. Le refroidissement de la surface par évaporation de l'eau est la principale perte d'énergie de l'océan.
- ☞ Sources d'énergie pour l'océan : la chaleur sensible, le Soleil (surtout la partie visible du spectre ; mais l'albédo de l'océan est faible – 0.08 – donc une fraction importante du rayonnement solaire pénètre dans la mer), le rayonnement de l'atmosphère (infrarouge thermique), la rotation de la Terre et la gravité terrestre (induisent des courants et des mouvements verticaux contribuant à répartir la chaleur reçue en surface majoritairement autour de l'équateur vers les hautes latitudes et à faire plonger dans les profondeurs l'eau plus froide et plus dense dans certaines régions des latitudes tempérées et polaires) et le vent (provoque vagues, houle et courant de surface par frottement à la surface de l'océan ; intervient également dans le refroidissement de la surface par un mécanisme de mélange vertical).
- ☞ L'océan, en retour, fournit à l'atmosphère : de l'eau (par évaporation), du CO₂ et autres gaz (en fonction du vent et de l'écart de pression partielle entre les deux niveaux), des particules (sel, etc.).

Ces échanges de différentes natures ne sont pas totalement indépendants, car température de surface et vent jouent un rôle important dans les variations de la plupart des flux.

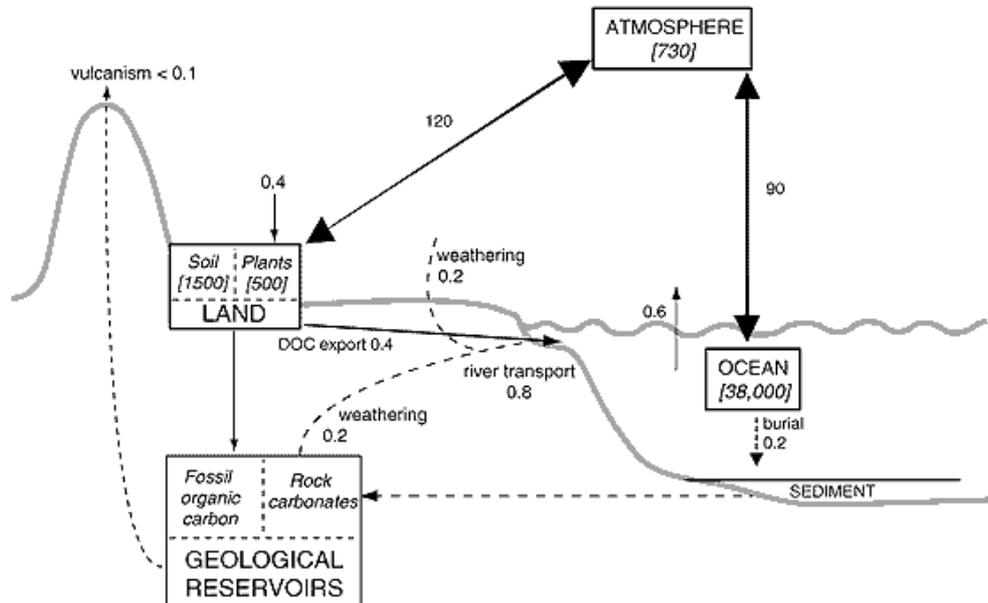
1.2.2 - Le cycle du carbone

Outre le fait qu'il s'agisse du premier maillon de la chaîne alimentaire, le phytoplancton tient une place importante dans la chimie de l'océan : il est l'un des éléments clés du « cycle du carbone » océanique et de la « pompe biologique » (Figures 1.7 et 1.8). Durant la photosynthèse, le phytoplancton absorbe le CO₂ inorganique présent dans l'océan (i.e. le CO₂ atmosphérique qui s'est dissous à la surface de l'océan). L'oxygène alors produit est

¹ Chaleur émise ou absorbée par un milieu qui donne lieu, non à une variation de température mais à un changement d'état physique.

rejeté dans l'atmosphère. S'il existait moins de phytoplancton, la quantité de dioxyde de carbone dans l'atmosphère s'accroîtrait.

a) Main components of the natural carbon cycle



b) The human perturbation

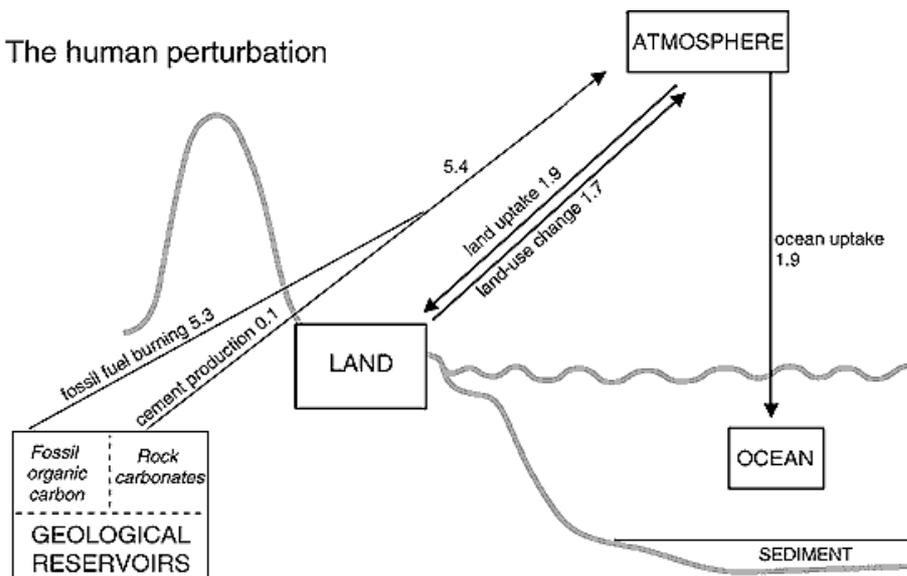


Figure 1.7 – Le cycle du carbone global : stocks (en PgC) et flux (en PgC/an) estimés pour les années 1980s. (a) Composantes principales du cycle naturel. Les flèches en gras indiquent les flux les plus importants : production primaire et respiration par la biosphère terrestre, et échange physique océan-atmosphère. Ces flux sont à peu près en équilibre chaque année. Les flèches fines montrent les flux naturels additionnels, importants sur des échelles temporelles plus longues. (b) Perturbation d'origine anthropique. La combustion des fuels fossiles et des forêts est le processus principal à l'origine de libération de CO₂ dans l'atmosphère. Environ la moitié de CO₂ reste dans l'atmosphère ; le reste est absorbé par la biosphère continentale (plantes et sols) et par l'océan. Cet apport de CO₂ atmosphérique supplémentaire met en péril la balance naturelle entre l'atmosphère et l'océan. [Prentice *et al.*, 2001]

Le CO₂ inorganique est donc transformé en carbone organique puis pris par la chaîne alimentaire : le zooplancton qui se nourrit de phytoplancton, les organismes supérieurs qui se nourrissent de zooplancton, et ainsi de suite. Les cadavres des organismes planctoniques et les pelotes fécales, composés de substances qui contiennent du carbone, vont sédimenter dans la colonne d'eau puis être pris par la circulation océanique (Figure 1.8). Certains éléments carbonés peuvent atteindre le fond de l'eau et rapidement être recouverts par d'autres matériels sédimentaires. Ils peuvent aussi être pris dans la boucle microbienne et reminéralisés dans la colonne d'eau. De cette manière, les océans agissent en tant que puits, un endroit où disposer du carbone global qui, sinon, s'accumulerait dans l'atmosphère en tant que dioxyde de carbone. Environ 10% du carbone absorbé par le phytoplancton est piégé au fond avec un temps de résidence important (500 à 1000 ans).

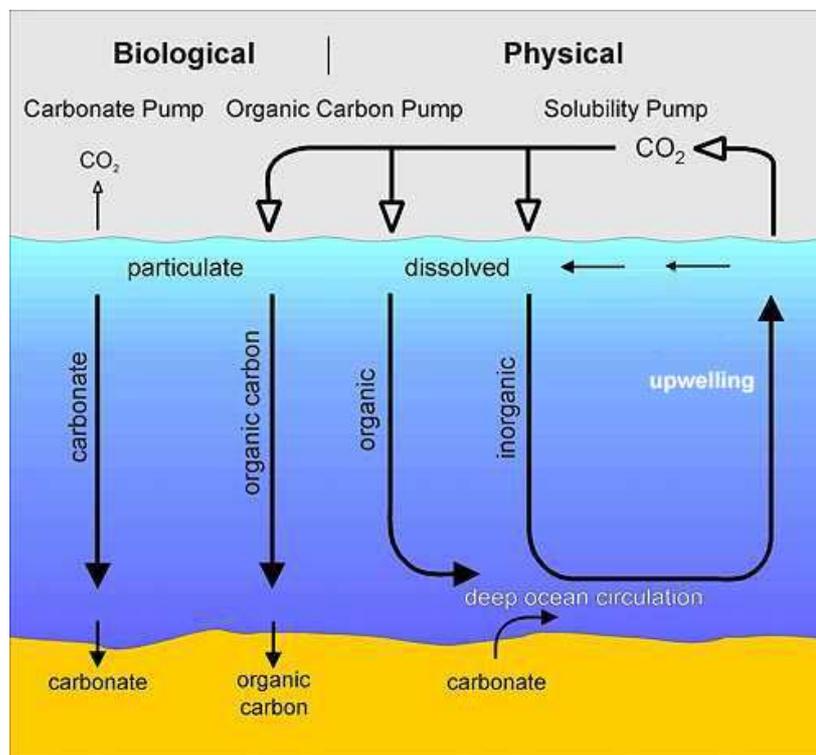


Figure 1.8 – Les pompes physique et biologique.

La biosphère terrestre agit aussi comme un puits de carbone. Néanmoins, le carbone qui y est piégé retourne fréquemment dans l'atmosphère, alors que les végétaux terrestres se décomposent ou brûlent. La déforestation contribue à l'accumulation de dioxyde de carbone dans l'atmosphère car elle réduit ainsi la végétation qui assimile le dioxyde de carbone. Les différentes sources du dioxyde de carbone dans l'atmosphère terrestre sont entre autres : les activités humaines (brûler les combustibles fossiles et les forêts, etc.), la

décomposition de la matière organique (les arbres, par exemple), le dioxyde de carbone exhalé par les animaux et les êtres humains, et l'activité volcanique. Les dernières estimations concernant le bilan du CO₂ anthropique sont résumées dans le Tableau 1.1.

Tableau 1.1 – Budget du CO₂ anthropique pour les périodes 1800-1994 et 1980-1999 [Sabine *et al.*, 2004].

CO ₂ sources and sinks	1800 to 1994 (Pg C)*	1980 to 1999 (Pg C)¶
<i>Constrained sources and sinks</i>		
(1) Emissions from fossil fuel and cement production	244† ± 20	117 ± 5
(2) Storage in the atmosphere	-165‡ ± 4	-65 ± 1
(3) Uptake and storage in the ocean	-118§ ± 19	-37 ± 8
<i>Inferred net terrestrial balance</i>		
(4) Net terrestrial balance = [-(1) - (2) - (3)]	39 ± 28	-15 ± 9
<i>Terrestrial balance</i>		
(5) Emissions from land-use change	100 to 180	24 ± 12
(6) Terrestrial biosphere sink = [-(1) - (2) - (3)] - (5)	-61 to -141	-39 ± 18

*Errors as estimated by respective sources; errors of sums and differences are calculated by quadratic error propagation. †From (79), with an error estimate of ±8%. ‡Calculated from the change in atmospheric PCO₂ (1800: 281 ± 2 ppm; 1994: 359 ± 0.4 ppm). §This study includes anthropogenic CO₂ storage in marginal seas and the Arctic Ocean. ||Based on (2); see text for details. ¶From (23), integrated for the period 1980 to 1999.

Ce tableau montre clairement l'influence humaine sur les flux de carbone, déséquilibre datant de la Révolution Industrielle au XIX^e siècle. En effet, la dépendance de l'humanité par rapport aux combustibles fossiles a provoqué une augmentation dramatique de la quantité de dioxyde de carbone atmosphérique (Figure 1.9).

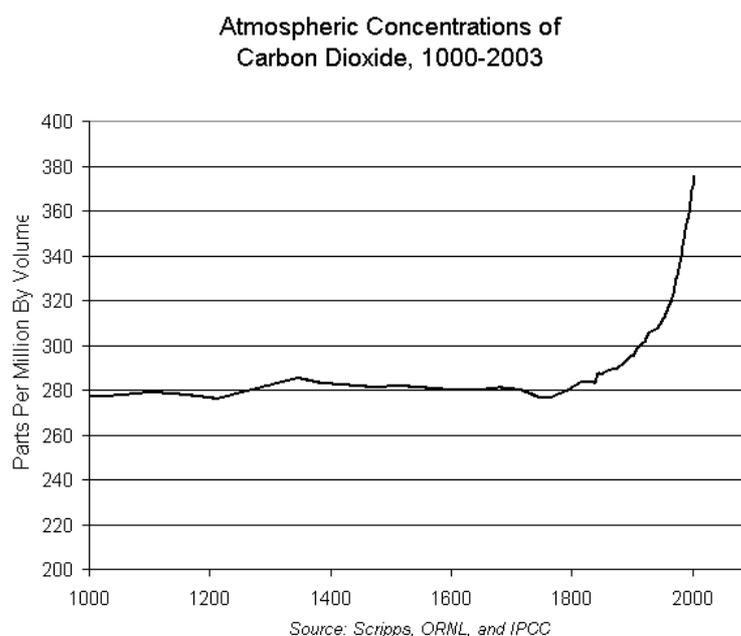


Figure 1.9 – Evolution du dioxyde de carbone sur le dernier millénaire.



Cette augmentation signifie que moins d'énergie grande longueur d'onde émise sur Terre peut s'échapper vers l'espace. De nombreux scientifiques pensent que cela peut mener à un réchauffement graduel de la Terre, tandis que d'autres sont d'avis que différents facteurs vont contrer cet effet de réchauffement. Par exemple, la couverture nuageuse réfléchit la lumière du soleil avant qu'elle n'atteigne la Terre, réduisant ainsi la quantité de lumière solaire qui atteint la surface de la Terre. L'étude de ces processus est difficile, car ils sont complexes et incluent des boucles de rétroaction positives et / ou négatives.

1.3 - Conclusion

Cette partie nous a permis de toucher du doigt la notion de « couleur de l'océan ». Tout d'abord, nous avons vu que les dégradés du bleu de l'océan montraient la présence de différentes espèces en différentes quantités. Puis, nous nous sommes intéressés au phytoplancton, composé d'algues microscopiques. C'est lui le principal responsable de la couleur de l'océan. Nous avons pu mesurer l'importance de la connaissance de la variabilité spatio-temporelle du phytoplancton, tant sur le plan environnemental (rôle de premier ordre dans le système Terre, indicateur de pollution) que sur le plan économique (gestion des zones de pêches). Dans un troisième temps, nous avons constaté que la surface de la mer est le lieu de transmission d'énergie et de matière de l'océan vers l'atmosphère, et vice-versa. Nous nous sommes alors intéressés au rôle de première importance du phytoplancton, à savoir l'assimilation du dioxyde de carbone atmosphérique par l'océan, durant la photosynthèse.

Personne ne sait encore quelle est la quantité de carbone que peuvent absorber les océans et les terres. Personne ne sait, non plus, comment l'environnement de la Terre s'ajustera à l'augmentation de la quantité de dioxyde de carbone (essentiellement d'origine anthropique) dans l'atmosphère. L'étude de la répartition et des changements du phytoplancton à l'aide de la couleur de l'eau et d'autres données peut aider à trouver des éléments de réponse à cette question.

La partie suivante introduit les deux types de mesure permettant d'accéder à la concentration en chlorophylle, pigments verts dont est composé le phytoplancton.

2 - Mesure de la concentration en chlorophylle

2.1 - Introduction

Des siècles durant, les navires, les côtes, et les îles ont été les seuls endroits à partir desquels nous pouvions observer, mesurer et étudier l'océan. Ces mesures sont qualifiées d'*in situ* (« sur place », en latin). Mais, de cette manière, nous pouvons seulement regarder une toute petite partie des océans du monde. Nous avons besoin d'un moyen plus performant pour étudier les océans. La mesure satellitaire permet une vision globale et régulière dans le temps.

Dans une première partie, nous présentons les données satellitaires utilisées au cours de ce travail. Après avoir introduit le principe de la mesure, nous dressons l'inventaire des capteurs exploités. Nous décrivons le principe de la mesure satellitaire de concentration en chlorophylle-*a*, qui est une mesure indirecte, obtenue par une relation empirique. Les données sont donc entachées d'une incertitude par rapport à la « vérité vraie ». Pour estimer leur qualité, ainsi que celle du produit combiné, il est donc nécessaire de les comparer à des données *in situ*, considérées comme référence car leur incertitude de mesure est très faible. Ces données seront présentées dans une seconde partie.

2.2 - Les données satellitaires

2.2.1 - Introduction

L'utilisation de l'imagerie satellitaire dans l'étude des processus océaniques est devenue essentielle pour la recherche océanographique et la surveillance des océans. Les données globales jouent un rôle fondamental, puisque les plates-formes conventionnelles ne peuvent pas couvrir convenablement la vaste étendue des océans dont la variabilité spatiale et temporelle est rapide, de surcroît. Depuis que la première mission couleur de l'eau – CZCS (Coastal Zone Color Scanner) – a démontré la possibilité d'observer à partir de l'espace la distribution du phytoplancton océanique à l'échelle globale, les données ont été utilisées dans de nombreuses études, dans le but de mieux comprendre le rôle de l'océan dans les cycles biogéochimiques, particulièrement la pompe biologique (cf. partie 1.2).

Neuf satellites, ayant un capteur couleur de l'eau à bord, sont actuellement en orbite, pour des applications à l'échelle globale ou régionale. Ils sont présentés dans le Tableau 1.2.

Tableau 1.2 – Informations générales sur les capteurs couleur de l'eau actuellement en orbite.

Capteur	Agence	Satellite	Date de lancement	Application
MERIS	ESA (Europe)	ENVISAT (Europe)	01/03/2002	Globale
MMRS	CONAE (Argentine)	SAC-C (Argentine)	21/11/2000	Côtier Fluvial
MODIS/Aqua	NASA (USA)	Aqua (USA)	04/05/2002	Globale
MODIS/Terra	NASA (USA)	Terra (USA)	18/12/1999	Globale
OCI	NEC (Japon)	ROCSAT-1 (Taïwan)	27/01/1999	Globale
OCM	ISRO (Inde)	IRS-P4 (Inde)	26/05/1999	Globale
OSMI	KARI (Corée)	KOMPSAT (Corée)	20/12/1999	Globale
PARASOL	CNES (France)	Série Myriade (France)	18/12/2004	Globale
SeaWiFS	NASA (USA)	Orbview-2 (USA)	01/08/1997	Globale

Après avoir expliqué le principe de la mesure et décrit les différentes étapes d'obtention d'une image couleur de l'eau, nous nous pencherons sur les produits disponibles. Puis nous nous intéresserons aux deux capteurs utilisés dans cette étude : les capteurs américains SeaWiFS (Sea-viewing Wide Field-of-view Sensor) et MODIS (MOderate Resolution Imaging Spectroradiometer) sur Aqua. Nous aurions souhaité utiliser aussi le capteur européen MERIS (Medium Resolution Imaging Spectrometer) ; malheureusement, malgré une mise en place depuis 4 ans déjà, des données journalières globales sont impossibles à récupérer.

2.2.2 - Qu'est-ce que la « couleur de l'eau » ?

La couleur de l'océan est le résultat de la réflexion de l'éclairement solaire sur la surface de l'océan. Océanographiquement parlant, le paramètre « couleur de l'eau » est défini comme les variations spectrales (i.e. variations selon la longueur d'onde considérée λ) de la réflectance diffuse des couches océaniques supérieures $R_{rs}(\lambda)$. Cette réflectance est elle-même le rapport entre l'éclairement ascendant issu de la surface de l'océan L_w et l'éclairement descendant y pénétrant E_d [Jerlov, 1976] (Figure 1.10) :

$$R_{rs}(\lambda) = \frac{\pi L_w}{E_d}$$

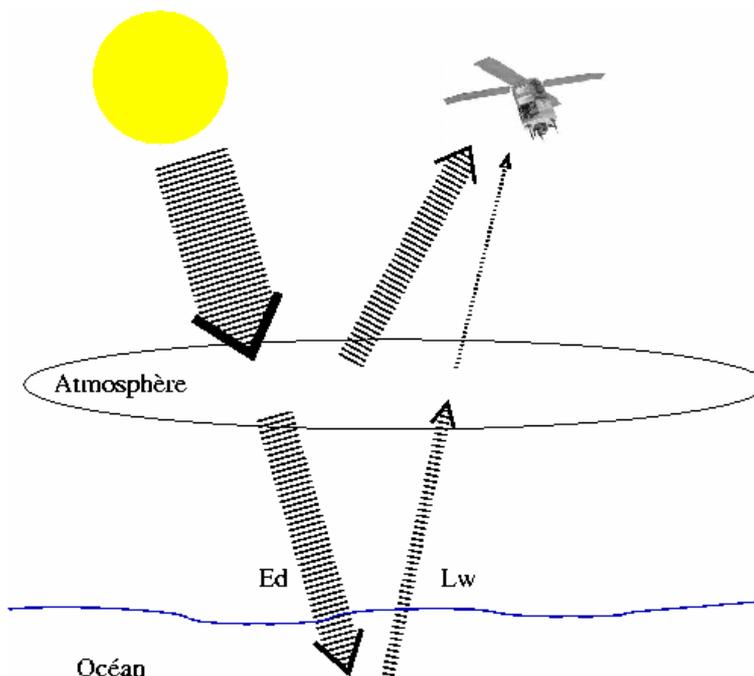


Figure 1.10 – Schéma du signal mesuré par le capteur satellitaire. L'épaisseur relative des flux montre l'importance de la part atmosphérique dans le signal mesuré (jusqu'à 90%).

Les variations de $R_{rs}(\lambda)$ sont dues aux diverses substances dissoutes ou particules en suspension dans l'eau, et également à l'eau de mer. C'est pourquoi, par des méthodes d'inversion, il est possible de récupérer des informations sur les propriétés ou les concentrations de ces « additifs » à partir des mesures de R_{rs} pour diverses longueurs d'ondes du visible. En optique marine, les masses d'eau sont souvent classées en deux catégories. Les eaux dites du « Cas 1 » sont les eaux du large, non affectées par les influences de la côte ou du fond. Dans ces eaux du Cas 1, qui représentent environ 95% de l'océan mondial, les propriétés optiques sont déterminées d'une part par les propriétés optiques de l'eau de mer elle-même, et d'autre part par celles du phytoplancton et de l'ensemble de ses produits associés (détritus organiques sous forme de particules ou de substances dissoutes et organiques hétérotrophes¹). En opposition, dans les eaux dites du « Cas 2 », les propriétés optiques dépendent non seulement des mêmes composantes que dans les eaux du Cas 1, mais aussi d'autres substances actives d'un point de vue optique, comme des sédiments en suspension ou des « substances jaunes » provenant des apports fluviaux. Les multiples combinaisons possibles entre phytoplancton et particules et substances dissoutes exogènes² génèrent une multitude de possibilités pour le spectre de la

¹ Organisme qui dépend des substances organiques pour son alimentation et sa croissance.

² Non produites par l'activité phytoplanctonique.

réflectance. L'inversion de ce signal pour estimer les concentrations en diverses composantes de ces eaux devient alors très difficile.

La réflectance peut être exprimée en fonction du coefficient de rétrodiffusion b_b , et du coefficient d'absorption a [Morel et Prieur, 1977] :

$$R_{rs}(\lambda) = F_{L,\beta} \frac{b_b(\lambda)}{a(\lambda)}$$

où $F_{L,\beta}$ est un coefficient de proportionnalité qui exprime le fait que R_{rs} dépend de la distribution des luminances¹ sous-marines L (F varie donc avec la hauteur du soleil dans la mesure où le champ radiatif sous la surface océanique change au-dessus de la surface), et de la forme de l'indicatrice totale de diffusion β [Antoine, 1998]. Cette forme change avec le rapport η entre diffusion moléculaire et diffusion due aux particules [Morel et Gentili, 1991].

Les propriétés optiques inhérentes a et b_b peuvent également être décomposées en la somme des coefficients correspondant aux différentes composantes, les propriétés optiques inhérentes étant additives. Pour les eaux du Cas 1, ces N composantes sont l'eau de mer pure (indice w), l'ensemble des particules (indice p), dont le phytoplancton (indice ϕ), et les autres particules, non algales (indice nap , en général obtenu par différence entre p et ϕ), et enfin les substances dissoutes colorées (indice ds).

$$R_{rs}(\lambda) = F_{L,\beta} \frac{b_{b,w}(\lambda) + b_{b,p}(\lambda)}{a_w(\lambda) + a_p(\lambda) + a_{ds}(\lambda)} \text{ avec } a_p(\lambda) = a_\phi(\lambda) + a_{nap}(\lambda).$$

Enfin, cette équation peut-être exprimée en utilisant les coefficients spécifiques correspondants, qui sont les coefficients normalisés par unité de concentration de la substance concernée (notés par une étoile *) :

$$R_{rs}(\lambda) = F_{L,\beta} \sum_{i=1}^N \frac{b_{b,i}^*(\lambda)[i]}{a_i^*(\lambda)[i]}$$

où $[i]$ est la concentration de l' i -ème composante et $a_i^*(\lambda)$ et $b_{b,i}^*(\lambda)$ ses coefficients spécifiques d'absorption et de rétrodiffusion à la longueur d'onde λ .

Pour plus de détails, le lecteur pourra se référer aux travaux de Mélin [2003], extrêmement bien approfondis sur le sujet.

¹ Intensité d'une source de rayonnement électromagnétique dans une direction donnée, divisée par l'aire apparente de cette source dans cette même direction.

2.2.3 - Comment obtient-on une image « True Color » ?

La Figure 1.11 montre les différentes étapes permettant d'obtenir la concentration en pigments chlorophylliens à partir du signal reçu par le capteur. Ces étapes vont être détaillées dans cette partie.

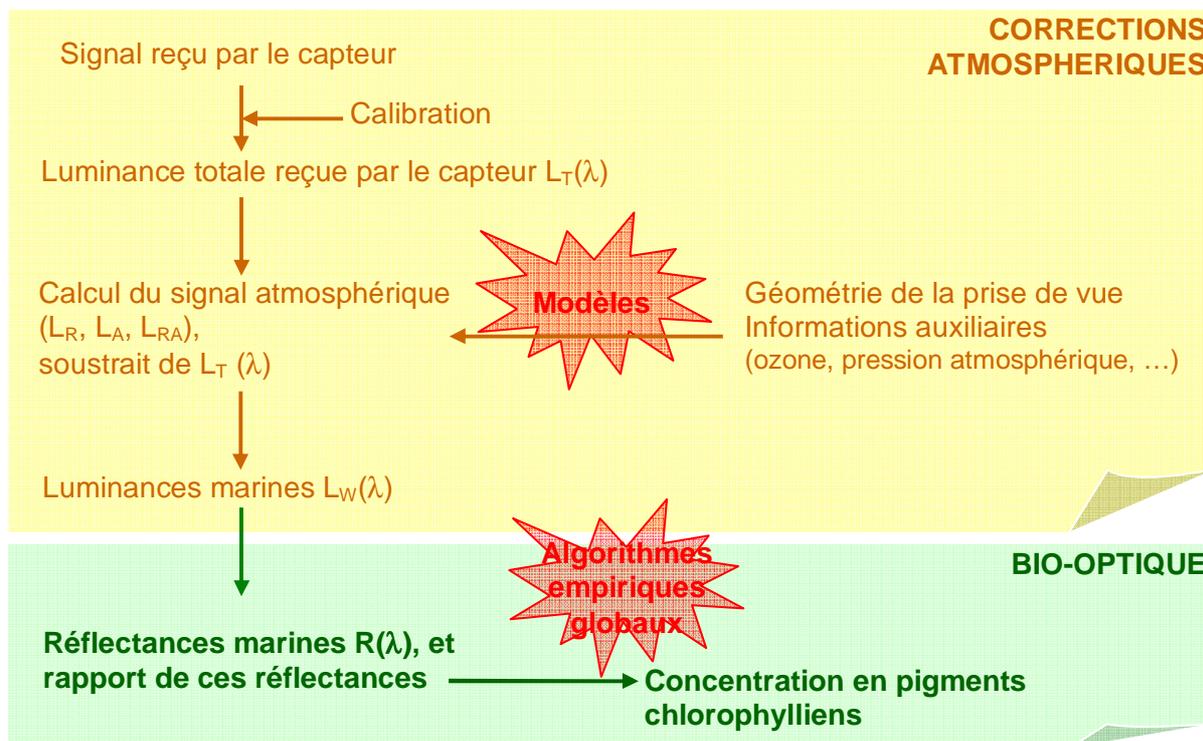


Figure 1.11 – Les différentes étapes d'obtention de la concentration en pigments chlorophylliens à partir du signal reçu par le capteur.

Principe de la mesure de couleur de l'eau

La mesure de la couleur de l'eau est une télédétection passive, dans le sens où le capteur mesure le rayonnement solaire réfléchi par l'océan, à l'instar de la photographie. Les données sont collectées selon différentes bandes spectrales du domaine du visible¹ et du Proche InfraRouge (PIR)². Pour chaque bande, le capteur mesure l'intensité de lumière qui l'atteint. Par exemple, une forte intensité de vert (longueur d'onde : 555 nm) traduit la présence de chlorophylle-a (principal pigment du phytoplancton), qui absorbe essentiellement le bleu violet et le rouge, et réfléchit le vert. Visuellement parlant (Figure 1.12 pour le capteur SeaWiFS), ces données consistent en une série d'images en dégradé de gris, s'apparentant à des photographies noir et blanc semblant avoir été prises avec

¹ Le domaine du visible correspondant à la bande de longueurs d'onde 400 nm (violet) à 700 nm (rouge).

² On considère comme proche infrarouge la bande 700 nm à 1400 nm.

différents filtres de couleur posés devant la lentille. Sur la Figure 1.12, on peut constater que les formes ont des intensités qui diffèrent selon les bandes. Par exemple, les nuages et l'eau apparaissent en brillant dans les bandes bleue et violette, alors que la terre est foncée. Par contre, dans les bandes rouge et infrarouge, c'est la terre qui apparaît brillante alors que l'eau est foncée.

Nous allons maintenant nous intéresser aux étapes permettant de passer de ces images noir et blanc à des images dites « True Color » (« couleur vraie »).

“Dark correction”

Quand le capteur pointe une zone totalement noire, ce qu'il détecte n'est pas nul. Ceci est la conséquence de la combinaison de deux facteurs : d'une part, il y a toujours des traces de lumière environnementale et d'autre part, les variations dans l'électronique du détecteur crée des faibles valeurs positives.

Pour corriger cet effet indésirable, le capteur détecte périodiquement une zone sombre connue. La valeur ainsi obtenue est alors soustraite à l'ensemble des autres données : les données ainsi produites sont qualifiées de « dark corrected ».

La mesure de luminance hors de l'atmosphère

Un instrument dédié à la télédétection de la couleur de l'océan doit enregistrer une certaine quantité d'énergie radiative en la convertissant en un courant électrique dont l'intensité est fonction de l'énergie reçue. Le signal ainsi créé peut ensuite être digitalisé, traité (corrections radiométriques élémentaires : première calibration, application des gains, etc.), stocké pour les traitements suivants et envoyé vers les stations de réceptions situées au sol. La quantité d'énergie enregistrée dépend de la portion de spectre électromagnétique concernée et également du champ de vue instantané (I_{FOV} : Instantaneous Field Of View), qui correspond à l'élément de base observable (pixel). De l'ouverture angulaire de ce champ dépend la taille du pixel de la surface océanique. L'une des difficultés rencontrées lors de la construction de ces instruments réside d'ailleurs dans l'établissement de rapports signal / bruit suffisamment élevés malgré la faiblesse du signal observé. Le rayonnement électromagnétique étant enregistré dans de petits angles solides pointés sur des directions données, il s'agit d'une mesure de luminance (la « luminance totale »), dans la mesure où son intensité dans une direction donnée est déterminée aussi bien par les propriétés optiques de l'océan que par celles de l'interface air-mer (la lumière solaire ne pénètre que sur une certaine portion de la surface de l'océan : la « couche optique ») et surtout celles de l'atmosphère (y compris les nuages, masquant la surface de l'océan au capteur s'ils sont présents). Cette luminance contient donc des informations sur la totalité du système océan-

atmosphère. Puisque l'on souhaite récupérer les caractéristiques océaniques, il faut corriger le signal mesuré des effets de l'atmosphère.

Les corrections atmosphériques

Afin d'analyser la lumière rétrodiffusée par l'océan, il faut extraire des luminances totales L_T la partie qui provient de l'océan, les « luminances marines » L_W . L'équation de la correction atmosphérique est la suivante : $L_T = L_R + L_A + L_{RA} + L_{WC} + L_G + tL_W$ où :

- ☞ t est la transmittance¹ diffuse de l'atmosphère.
- ☞ L_R désigne la contribution de la dispersion moléculaire (Rayleigh), qui peut être modélisée avec précision. Ce paramètre dépend de la polarisation, des vents et de la pression atmosphérique. MODIS, en particulier, nécessite aussi une quantification précise du changement de la réflectivité du miroir en fonction de l'angle d'incidence.
- ☞ $L_A + L_{RA}$ est la contribution des aérosols. La concentration en aérosols ainsi que leurs propriétés optiques ont besoin d'être déterminées pour résoudre ce terme. La plupart des techniques actuellement utilisées reposent sur l'observation du système océan + atmosphère dans au moins deux canaux du proche infrarouge, pour lesquels le signal océanique est nul (phénomène visible dans certaines zones du Cas 1). Ceci est dû à la très forte absorption de l'eau et aux faibles capacités de diffusion des différents matériels présents. A partir de l'intensité de ce signal et de sa dépendance spectrale entre les deux longueurs d'onde considérées, on obtient suffisamment d'information sur l'aérosol présent pour pouvoir en extrapoler la contribution vers les longueurs d'onde du domaine visible à l'aide de modèles d'aérosols montrés dans le Tableau 1.3 (adaptation de Shettle et Fenn, 1979).
- ☞ L_{WC} désigne la contribution du « panneau de recouvrement blanc », estimée à partir d'une relation statistique avec la vitesse du vent.
- ☞ L_G est le reflet du soleil, c'est-à-dire la somme des réflectances directe et diffuse de l'éclairement solaire sur la surface océanique. L'effet sur SeaWiFS est minimal, car le capteur a la capacité de s'incliner (« tilting », cf. partie 2.2.4) pour éviter ce phénomène. Ce n'est pas le cas de MODIS, et le reflet du soleil doit donc être supprimé, en fonction des vecteurs de vent et de la polarisation.

¹ Rapport de l'énergie qui subsiste après la traversée d'un milieu matériel à l'énergie incidente.

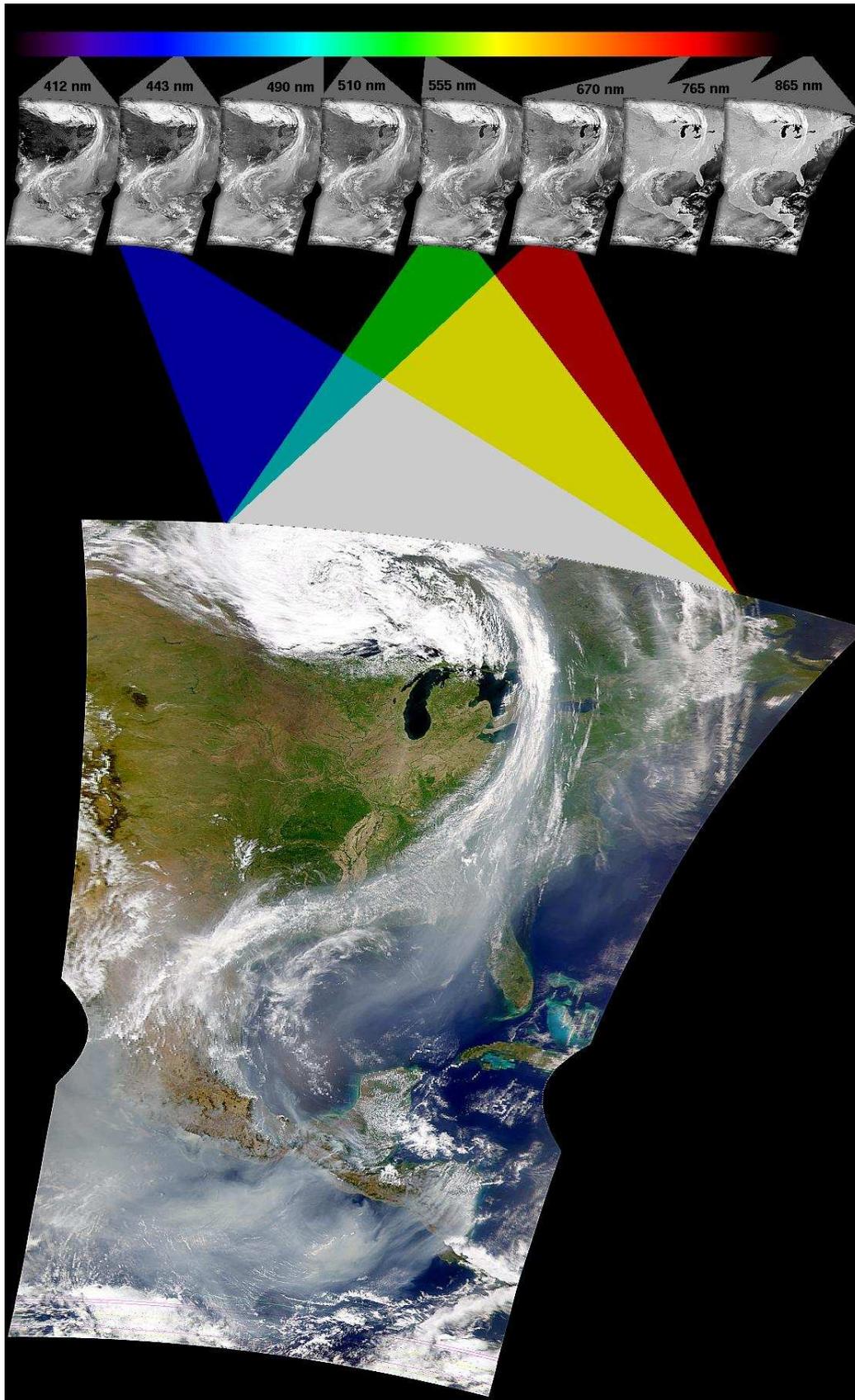


Figure 1.12 – Obtention d'une image « True Color » SeaWiFS [SeaWiFS Website].

Tableau 1.3 – Modèles d'aérosols utilisés dans les procédures standards. Il y a 3 modèles primaires avec des quantités variables d'humidité relative. Le modèle côtier est un mélange entre les modèles terrestres et maritimes de Shettle et Fenn [1979], tandis que le modèle océanique est un cas océanique extrême. Il est important d'avoir des modèles pour différents niveaux d'humidité car la taille des particules d'aérosols augmente avec elle. Ceci modifie les effets de la dispersion de la lumière sur les particules d'aérosols [Shettle et Fenn, 1979] et nécessite l'utilisation de modèles séparés pour représenter les conditions à un taux d'humidité relative donné.

O99	Océanique, humidité 99%
M50	Maritime, humidité 50%
M70	Maritime, humidité 70%
M90	Maritime, humidité 90%
M99	Maritime, humidité 99%
C50	Côtier, humidité 50%
C70	Côtier, humidité 70%
C90	Côtier, humidité 90%
C99	Côtier, humidité 99%
T50	Troposphérique, humidité 50%
T90	Troposphérique, humidité 90%
T99	Troposphérique, humidité 99%

La géolocalisation

Les données sont collectées suivant une série de points selon les lignes de scrutation, créant ainsi un tableau rectangle de « pixels ». Néanmoins, la courbure de la surface de la Terre ajoutée à la géométrie de l'instrument de balayage du satellite fait que les données ne sont pas toujours représentées sur une zone parfaitement rectangulaire. Pour corriger cet effet, les longitude et latitude terrestres exactes de chaque pixel sont calculées et les données sont ensuite projetées sur une grille longitude / latitude. Les diverses bandes de l'image de la Figure 1.12 ne sont pas rectangulaires car, justement, elles ont été géolocalisées.

La co-registation

Pour de nombreux satellites, les données provenant des différentes bandes sont mesurées à des instants légèrement différents (de l'ordre de la microseconde). Puisque le satellite est en mouvement, les différentes bandes ne correspondent donc pas exactement à la même région de la Terre. Cela se traduit par des bords normalement nets, comme le rivage des lacs ou la limite des nuages, qui apparaissent non alignés dans l'image « true color » finale. Pour corriger cet effet, les bandes doivent être co-registrées, de façon à ce que chaque pixel de chaque bande coïncide exactement avec le même pixel des autres bandes. Ceci est très facile à réaliser quand les données sont géolocalisées ; toutes les bandes sont simplement alignées en fonction de leurs coordonnées longitude / latitude. Cette étape n'est pas nécessaire pour les données SeaWiFS car le capteur a été conçu de façon à collecter les 8 bandes exactement au même instant.



Représentation en image « RVB »

Au final, les 3 bandes qui représentent au mieux le rouge, le vert et le bleu (RVB) dans le spectre visible sont combinées. Chaque bande est visualisée selon une échelle monochromatique correspondant à la couleur appropriée. Ces images sont alors mixées de manière à produire l'intervalle entier des couleurs du visible, créant ainsi une image très proche de ce que peut percevoir l'œil humain.

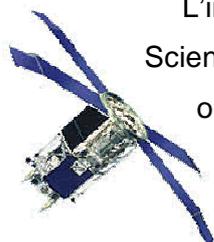
Conclusion

Une des difficultés majeures de la télédétection spatiale de la couleur de l'océan provient de la faiblesse du signal que l'on cherche à identifier par rapport au signal que l'on cherche à éliminer. Le premier, la luminance marine, contient les informations sur la composition des eaux océaniques et ne représente que 10% du signal total mesuré au niveau du capteur. Le second est exclusivement dû à l'atmosphère et à l'interface océan-air. La prise en compte des aérosols représente la principale difficulté du traitement de l'image. D'importants efforts sont actuellement entrepris pour mieux connaître les aérosols et leur climatologie. Le signal mesuré doit aussi être parfaitement calibré, sans quoi les données peuvent être déficientes [Evans et Gordon, 1994].

Toutes les espèces phytoplanctoniques contiennent de la chlorophylle-*a*, mais aussi un large panel d'autres pigments. Cette chlorophylle, qui absorbe préférentiellement dans le bleu, provoque le changement de couleur qui permet de la quantifier, et elle est utilisée comme indice des propriétés optiques des eaux du Cas 1. Néanmoins, un flou important existe sur la relation entre les propriétés optiques et la concentration en chlorophylle-*a*, ayant pour origine la variabilité des assemblages des pigments, de particules et de substances dissoutes qui accompagnent le phytoplancton. Les propriétés optiques, et donc L_w , varient en conséquence d'un facteur 2 ou 3 environ pour une concentration en chlorophylle donnée. De nombreux efforts sont actuellement entrepris pour essayer de mieux décrire, comprendre et quantifier cette variabilité naturelle qui limite les capacités de la télédétection de la couleur de l'océan. Le problème des eaux du Cas 2 est encore plus complexe, en raison de la présence de sédiments et de substances dissoutes exogènes en proportions variables géographiquement.

Intéressons-nous maintenant aux différents capteurs utilisés.

2.2.4 - SeaWiFS (Sea-viewing Wide Field-of-view Sensor)



L'instrument SeaWiFS a été lancé sur le satellite SeaStar par l'OSC (Orbital Sciences Corporation) en août 1997. Il observe maintenant la Terre depuis une orbite descendante, héliosynchrone, quasi-polaire et circulaire, à une altitude de 705 km, et passant à l'Equateur à midi et 20 minutes, heure locale (Tableau 1.4). La période de revisite est de 2 jours.

SeaWiFS est équipé d'instruments de balayage qui conduisent un télescope plié en dehors de son axe et un miroir demi-angle rotatif qui est synchronisé en phase avec le télescope et qui tourne à la moitié de la vitesse de celui-ci (Figure 1.13) [<http://www.oceancolor.gsfc.nasa.gov/SeaWiFS/SEASTAR/SPACECRAFT.html>]. En plus de l'équipement électronique prévu pour la polarisation, cet arrangement permet d'atteindre un niveau minimum de polarisation¹. Bien que l'angle de balayage maximal soit $\pm 58.3^\circ$, la trace des données GAC (Global Area Coverage ; cf. partie 2.2.6) est limitée à 1502 km (Tableau 1.4). Durant chaque orbite, SeaWiFS s'incline de façon à réduire les effets de l'éblouissement dû à la réflexion du soleil sur la surface de l'eau. En conséquence, la trace de SeaWiFS montre une bande horizontale sans donnée aux alentours de l'équateur (la latitude de cette bande varie en fonction de la saison, comme montré Figure 1.14).

SeaWiFS possède 8 bandes spectrales, toutes utilisées pour la mesure de la couleur de l'eau (Tableau 1.5). La résolution, de 1.1 km au nadir, est identique pour chaque bande. Les données GAC sont sous-échantillonnées tous les 4 pixels et toutes les 4 lignes, la résolution des données SeaWiFS est donc de 4.5 km environ.

Tableau 1.4 – Caractéristiques du capteur SeaWiFS.

Date de lancement	1 ^{er} août 1997
Données disponibles depuis le...	4 septembre 1997
Heure de passage à l'Equateur	12h20
Orbite	Héliosynchrone à 705km, descendante
Période de revisite	2 jours
Inclinaison (« tilt »)	-20°, 0°, +20°
Largeur de trace	2801km (LAC) 1502km (GAC)
Résolution spatiale au nadir²	1130m (LAC) 4500m (GAC)

¹ Phénomène lumineux lié à l'orientation des vibrations lumineuses autour de leur direction de propagation.

² Point sur la Terre situé juste à la verticale du satellite.

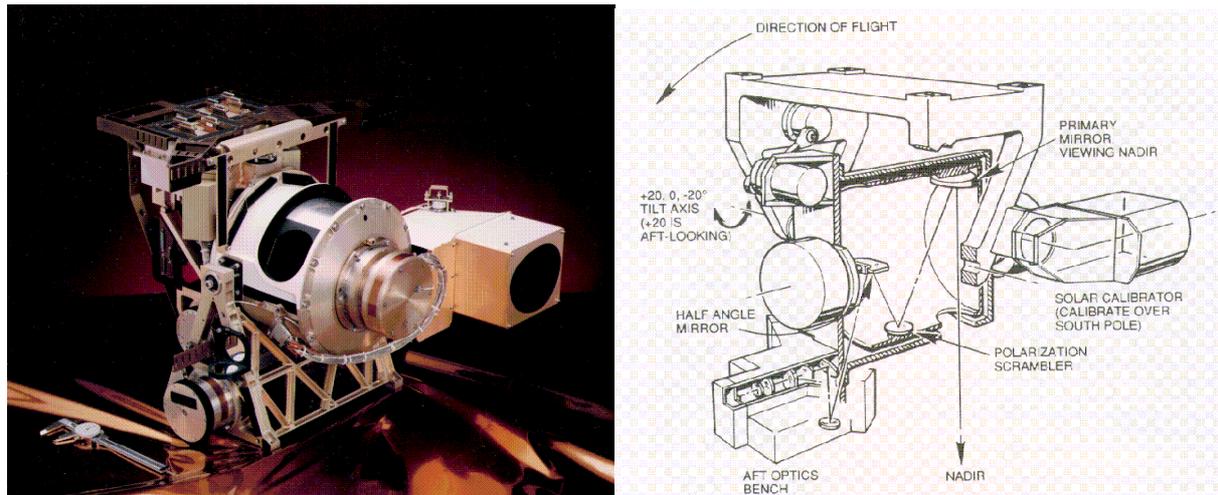


Figure 1.13 – A gauche, photographie du capteur SeaWiFS. A droite, schéma explicatif [<http://oceancolor.gsfc.nasa.gov/SeaWiFS/SEASTAR/SPACECRAFT.html>].

Tableau 1.5 – Bandes spectrales de SeaWiFS.

Bande #	Longueur d'onde centrale	Largeur de bande	Couleur	Utilisation principale
1	412 nm	20 nm	Violet	Matière organique dissoute
2	443 nm	20 nm	Bleu	Chlorophylle
3	490 nm	20 nm	Bleu / vert	Chlorophylle
4	510 nm	20 nm	Vert	Chlorophylle
5	555 nm	20 nm	Vert / Jaune	Chlorophylle
6	670 nm	20 nm	Rouge	Aérosols atmosphériques
7	765 nm	40 nm	Infrarouge	Aérosols atmosphériques
8	865 nm	40 nm	Infrarouge	Aérosols atmosphériques

L'algorithme bio-optique empirique utilisé pour obtenir la concentration en chlorophylle-a à partir des radiances est OC4v4 :

$$Chl_a = 10^{0.366 - 3.067R + 1.93R^2 + 0.649R^3 - 1.532R^4}$$

où $R = \log_{10} \left(\frac{\max(R_{rs\ 443}, R_{rs\ 490}, R_{rs\ 510})}{R_{rs\ 555}} \right)$ est le **rapport de bande maximal** et $R_{rs\ X}$ désigne la réflectance relative à la longueur d'onde X .

La Figure 1.14 montre une image globale de concentration en chlorophylle-a mesurée par le capteur SeaWiFS les 26 mars et 13 août 2003, pour le reprocessing #5.1 (il s'agit du dernier en date, à l'instant où le manuscrit est rédigé ; disponible depuis le 5 juillet 2005, le lecteur peut obtenir plus de renseignements à son sujet sur le site <http://oceancolor.gsfc.nasa.gov/REPROCESSING/SeaWiFS/R5.1/>).

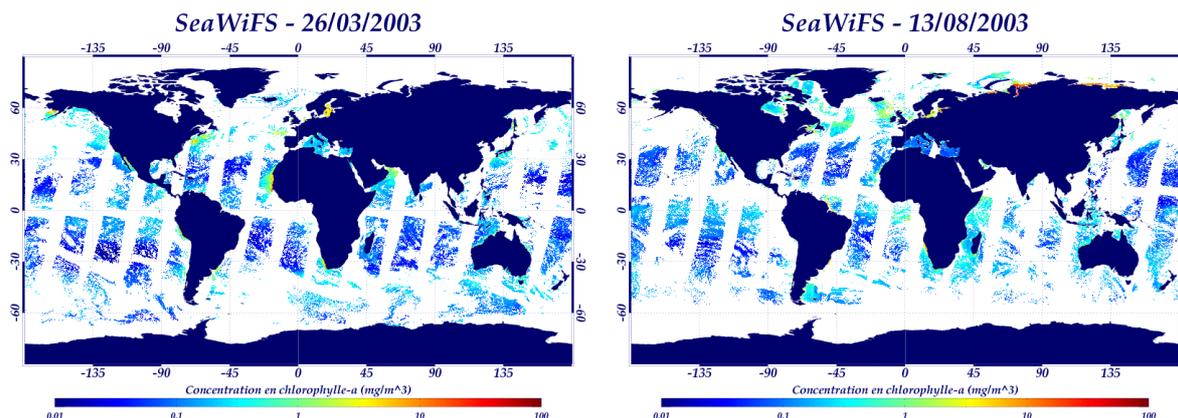


Figure 1.14 – Carte SeaWiFS de concentration en chlorophylle-*a* pour les 26 mars (à gauche) et 13 août 2003 (à droite).

2.2.5 - MODIS (MODERate resolution Imaging Spectroradiometer)



L'instrument MODIS a été lancé sur le satellite Aqua (ou EOS-PM) en mai 2002. Pour le différencier du capteur MODIS (le même) lancé sur le satellite Terra (ou EOS-AM) en décembre 1999, il est noté MODIS/Aqua. Ce capteur observe la Terre depuis une orbite ascendante, héliosynchrone, quasi-polaire et circulaire, à une altitude de 705 km, et passant à l'Equateur à 13h30, heure locale (Tableau 1.6). La période de revisite est de 2 jours.

Tableau 1.6 – Caractéristiques du capteur MODIS sur le satellite Aqua.

Date de lancement	4 mai 2002
Données disponibles depuis le...	29 novembre 2002
Heure de passage à l'Equateur	13h30
Orbite	Héliosynchrone à 705km, ascendante
Période de revisite	2 jours
Inclinaison (« tilt »)	Non
Largeur de trace	2330km
Résolution spatiale au nadir	250m (bandes #1-2) 500m (bandes #3-7) 1000m (bandes #8-36)

Pour observer la Terre sur $\pm 55^\circ$ en latitude, le capteur MODIS utilise un miroir à balayage double-face qui tourne de façon continue [Xiong *et al.*, 2005] : sa largeur de vision est de 10 km le long de la trace au nadir et 2330 km en largeur de trace à $\pm 55^\circ$ (Tableau 1.6). MODIS/Aqua ne bénéficie pas du même arrangement que SeaWiFS quant à la polarisation (cf. partie précédente) : la sensibilité de polarisation nominale de SeaWiFS est donc 20 fois moins importante que le maximum de sensibilité à la polarisation de MODIS/Aqua et une correction de polarisation est donc nécessaire pour améliorer les données de ce capteur [Meister *et al.*, 2005]. Contrairement à SeaWiFS, MODIS ne s'incline pas pour éviter la réflexion du soleil sur la surface de l'océan.

Tableau 1.7 – Bandes spectrales de MODIS/Aqua.

Bande #	Longueur d'onde centrale	Largeur de bande	Couleur	Utilisation principale
1	645 nm	25 nm	Rouge	Limites des terres, des nuages et des aérosols
2	856 nm	15 nm	Infrarouge	
3	469 nm	20 nm	Bleu	Propriétés des terres, des nuages et des aérosols
4	555 nm	20 nm	Vert / jaune	
5	1240 nm	20 nm	Infrarouge	
6	1640 nm	24 nm	Infrarouge	
7	2130 nm	50 nm	Infrarouge	
8	412 nm	15 nm	Violet	Couleur de l'eau / Phytoplancton / Biogéochimie
9	443 nm	10 nm	Bleu	
10	493 nm	10 nm	Bleu / vert	
11	531 nm	10 nm	Vert	
12	551 nm	10 nm	Vert / jaune	
13	667 nm	10 nm	Rouge	
14	678 nm	10 nm	Rouge	
15	748 nm	10 nm	Infrarouge	
16	869 nm	15 nm	Infrarouge	
17	905 nm	30 nm	Infrarouge	
18	936 nm	10 nm	Infrarouge	
19	940 nm	50 nm	Infrarouge	Température de surface et des nuages
20	3.750 µm	0.180 µm	Infrarouge	
21	3.959 µm	0.030 µm	Infrarouge	
22	3.959 µm	0.030 µm	Infrarouge	
23	4.050 µm	0.060 µm	Infrarouge	Température atmosphérique
24	4.465 µm	0.065 µm	Infrarouge	
25	4.515 µm	0.067 µm	Infrarouge	Vapeur d'eau des nuages
26	1.375 µm	0.030 µm	Infrarouge	
27	6.715 µm	0.360 µm	Infrarouge	
28	7.325 µm	0.300 µm	Infrarouge	Propriétés des nuages
29	8.550 µm	0.300 µm	Infrarouge	
30	9.730 µm	0.300 µm	Infrarouge	Ozone
31	11.030 µm	0.500 µm	Infrarouge	Température de surface et des nuages
32	12.020 µm	0.500 µm	Infrarouge	
33	13.335 µm	0.300 µm	Infrarouge	Altitude du sommet des nuages
34	13.635 µm	0.300 µm	Infrarouge	
35	13.935 µm	0.300 µm	Infrarouge	
36	14.235 µm	0.300 µm	Infrarouge	

MODIS possède 36 bandes spectrales (Tableau 1.7), dont 9 utilisées pour la couleur de l'océan (bandes #8 à #16, les autres étant nécessaires à la mesure des propriétés des terres, des nuages et des aérosols, à la vapeur d'eau contenue dans l'atmosphère ou les nuages, à la température de surface, des nuages et atmosphérique, à l'altitude du sommet des nuages, et à l'ozone ; <http://modis.gsfc.nasa.gov/about/specifications.php>). Les bandes spectrales #1 et #2 de MODIS ont une résolution de 250m au nadir, les bandes #3-7 de 500m et les bandes #8-36 de 1km [Xiong et al., 2005].

Le centre des bandes spectrales de MODIS dédiées à la couleur de l'eau ne sont pas exactement les mêmes que celles de SeaWiFS. L'algorithme empirique pour calculer la

concentration en chlorophylle-a à partir des radiances n'est donc pas exactement le même, même s'il s'en rapproche le plus possible. Il s'agit d'OC3M :

$$Chl_a = 10^{0.283 - 2.753R + 1.457R^2 + 0.659R^3 - 1.403R^4}$$

où $R = \log_{10}\left(\frac{\max(R_{rs443}, R_{rs488})}{R_{rs555}}\right)$ est le rapport de bande maximal et R_{rsX} désigne la réflectance relative à la longueur d'onde X .

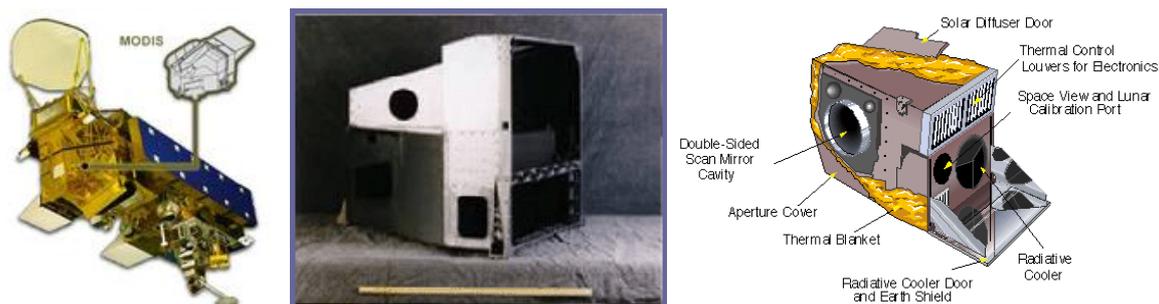


Figure 1.15 – A gauche, emplacement du capteur MODIS sur le satellite Aqua. Au milieu, photographie du capteur MODIS. A droite, schéma explicatif. [<http://modis.gsfc.nasa.org>]

La Figure 1.16 montre une image globale de concentration en chlorophylle-a mesurée par le capteur MODIS/Aqua les 26 mars et 13 août 2003 pour le reprocessing #1 .1 (il s'agit, là encore, du dernier en date, à l'instant où le manuscrit est rédigé ; disponible depuis le 4 août 2005, le lecteur peut obtenir plus de renseignements à son sujet sur le site <http://oceancolor.gsfc.nasa.gov/REPROCESSING/Aqua/R1.1/>).

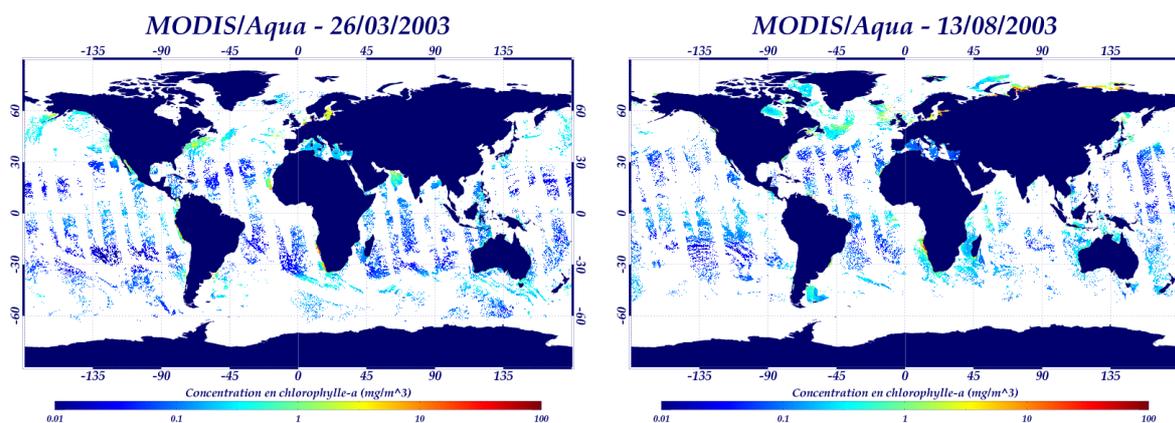


Figure 1.16 – Carte MODIS/Aqua de concentration en chlorophylle-a pour les 26 mars (à gauche) et 13 août 2003 (à droite).

2.2.6 - Produits disponibles et produits utilisés

Les produits opérationnels et les « catalogues d'images » (« browse products ») sont disponibles à différents niveaux de traitement :

Niveau 1

Les produits de niveau 1 sont les images « true color ». Ils contiennent les valeurs numériques correspondant aux luminances brutes mesurées par le capteur dans chaque canal, avec les données de télémétrie, calibration et navigation associées. Les données de niveau 1 sont nécessaires à la géolocalisation, à la calibration et au traitement des données.

En ce qui concerne SeaWiFS, ces données sont fournies en mode LAC (Local-Area Coverage, pleine résolution, 1 km) ou en mode GAC (résolution réduite). Les données GAC sont obtenues par sous-échantillonnage des données LAC à raison de 1 pixel sur 4 sauvegardé pour chaque ligne scrutée et 1 ligne sur 4 sauvegardée pour chaque trace.

Niveau 2

Les données de niveau 2 contiennent les paramètres géophysiques (luminances marines, luminances des aérosols, concentration en chlorophylle-a, etc.) pour chaque pixel, dérivés des mesures de radiances du niveau 1 après application des facteurs de calibration, des corrections atmosphériques et des algorithmes bio-optiques (cf. partie 2.2.4 pour SeaWiFS et 2.2.5 pour MODIS/Aqua). La couverture géographique de chaque produit de niveau 2 correspond exactement à celle de son homologue de niveau 1.

Niveau 3

Les données de niveau 3 peuvent être journalières, à 8 jours, mensuelles et annuelles. Elles consistent en l'accumulation de toutes les données de niveau 2 correspondant à la période souhaitée. Il existe 2 types de niveau 3 : les produits dits « binnés » et les produits SMI (Standard Mapped Image). Les premiers sont sur une grille à aire égale, c'est-à-dire une grille régulière en kilomètres ; chaque pixel représente une valeur de chlorophylle. Les produits SMI, quant à eux, sont la représentation imagée des produits binnés, c'est-à-dire un tableau à 2 dimensions d'octets représentant le globe projeté sur une grille régulière en degrés.

Format des données utilisées dans l'article Pottier et al. [2006]

Les données utilisées au cours de l'étude sur l'Atlantique Nord et Equatorial dont les résultats ont été publiés dans Pottier *et al.* [2006] sont les données de niveau 3 binnées issues des reprocessings #4 pour SeaWiFS et #1 pour MODIS/Aqua et reprojctées sur une grille régulière en degrés. Les résolutions des nouvelles grilles de projection sont : 0.1°x 0.1°

pour SeaWiFS (i.e. environ 11.1 km x 11.1 km à l'équateur) et 0.05° x 0.05° pour MODIS/Aqua (i.e. environ 5.6 km x 5.6 km). La Figure 1.17 montre les positions respectives du pixel MODIS/Aqua correspondant à un pixel SeaWiFS situé à la longitude i et la latitude j . Ces résolutions avaient été choisies car elles présentaient un bon compromis : en effet, elles conservaient le rapport entre les résolutions originales (1/12° pour SeaWiFS et 1/24° pour MODIS/Aqua) et l'arrondi était plus pratique pour le calcul tout en étant relativement proches des résolutions originales.

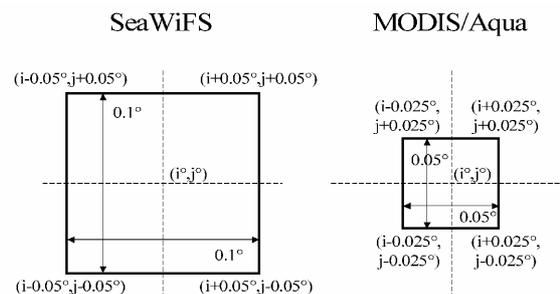


Figure 1.17 – Exemple de l'emplacement d'un pixel de chlorophylle- a SeaWiFS et de son homologue MODIS/Aqua à la longitude i et la latitude j .

Dans un but d'océanographie opérationnelle, le calcul sur des matrices permet une gestion rapide de l'accès aux données. D'autre part, une grille régulière est nécessaire pour deux des méthodes de combinaison utilisées : la moyenne pondérée et l'analyse en ondelettes. Enfin, beaucoup d'études océanographiques sont réalisées sur des grilles régulières (simplicité, moyens informatiques, etc.).

Format des données utilisées pour notre étude en global (ce manuscrit)

Notre étude en global présentée dans ce manuscrit porte sur les nouveaux reprocessings de SeaWiFS et MODIS/Aqua (#5.1 et #1.1 respectivement). A cette occasion, l'algorithme de reprojexion des données binnées sur une grille régulière a été complètement modifié car l'ancien, utilisant une interpolation basée sur le plus proche voisin, présentait des artefacts au niveau des latitudes extrêmes, comme montré Figure 1.18 (cercles verts) en exemple de ce problème. Du coup, nous avons choisi de garder la résolution de SeaWiFS et MODIS/Aqua à l'équateur. La résolution des nouvelles grilles est donc : 1/12° x 1/12° pour SeaWiFS (i.e. environ 9.2 km x 9.2 km à l'équateur) et 1/24° x 1/24° pour MODIS/Aqua (i.e. environ 4.6 km x 4.6 km). La Figure 1.19 montre les positions respectives des 4 pixels MODIS/Aqua correspondant à un pixel SeaWiFS situé à la longitude i et la latitude j .

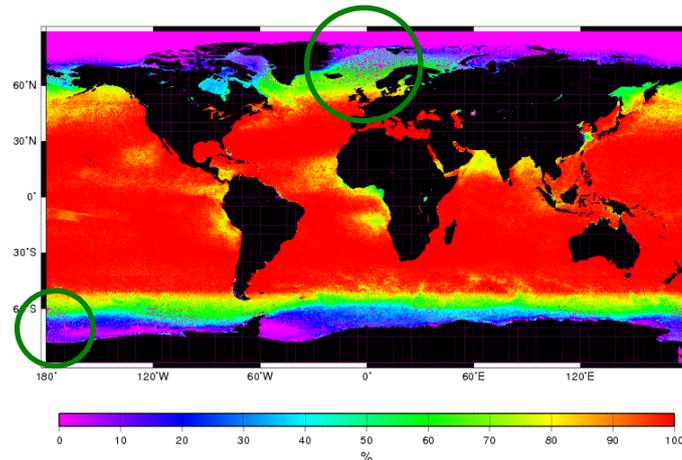


Figure 1.18 – Pourcentage de présence des pixels SeaWiFS pour une étude des données mensuelles comprises entre septembre 1997 et août 2003.

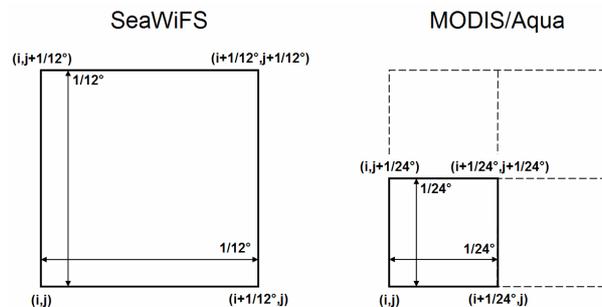


Figure 1.19 – Exemple de l'emplacement d'un pixel de chlorophylle-*a* SeaWiFS et de ses 4 homologues MODIS/Aqua à la longitude *i* et la latitude *j*.

2.2.7 - Conclusion

Nous avons vu, dans cette partie, le principe de la mesure de la couleur de l'eau, qui est une télédétection passive. Les différentes étapes jusqu'à l'obtention de la carte des concentrations en chlorophylle journalière utilisée ont été explicitées. Les données utilisées sont les cartes journalières de concentration en chlorophylle-*a* répartie sur une grille régulière en degrés. Dans un dernier temps, les caractéristiques des deux capteurs mis en jeu, SeaWiFS et MODIS sur Aqua, ont été précisées. Au cours de ce travail de thèse, l'année étudiée pour le développement et la comparaison des algorithmes de combinaison des données *in situ* est la première année que les capteurs étudiés ont en commun, à savoir l'année 2003.

2.3 - Les données *in situ*

2.3.1 - Introduction

Les mesures *in situ* de concentration en chlorophylle-a ont été utilisées pour estimer la qualité des données satellitaires et des données combinées. Il n'existe malheureusement pas de base de données mondiale, regroupant l'ensemble des mesures. L'exercice de récupération du maximum de données *in situ* est donc extrêmement délicat car il faut contacter chaque organisation océanographique nationale, chaque scientifique susceptible d'avoir fait des campagnes en mer. Cet exercice requiert patience et persévérance. Notons également que nous ne nous intéressons qu'aux mesures comprises entre 0 et 5 m, ce qui restreint encore le nombre de données. En effet, c'est la profondeur moyenne de l'océan vue par un capteur satellitaire. Cette profondeur dépend, évidemment, de la clarté de l'eau : on voit beaucoup plus profondément dans une eau claire (une dizaine de mètres) que dans une eau opaque (quelques mètres).

Dans cette partie, nous commençons par introduire les différentes techniques pour mesurer la concentration en chlorophylle-a dans un volume d'eau. Puis, nous nous intéressons aux données utilisées pour l'année 2003, dont les caractéristiques sont indiquées dans le Tableau 1.8. Les bases de données AMT (Atlantic Meridional Transect, UK), SeaBASS (SeaWiFS Bio-optical And Storage System, USA), celle du NODC (National Oceanographic Data Center, USA), celle du CIEM (Conseil National pour l'Exploitation de la Mer) et celle du NIO (National Institute of Oceanography, Inde) sont tout d'abord présentées. Puis, sont cités les programmes scientifiques ayant réalisé des mesures de concentration en chlorophylle-a durant l'année 2003, OISO, MINERVE, ARGAU, DIAPALIS et DYFAMED. Enfin, seront abordées les mesures effectuées au cours de campagnes en mer ponctuelles.

2.3.2 - Principe de la mesure

Il existe différentes techniques pour mesurer la concentration en chlorophylle dans un volume d'eau, dont la spectrophotométrie, la chromatographie liquide haute performance (HPLC : High Performance Liquid Chromatography) et la fluorométrie.

La spectrophotométrie est la méthode classique pour déterminer la quantité de chlorophylle à la surface de l'eau. Les différentes étapes de mesure sont les suivantes [JGOFS, 1994] :

- 1) Récupérer des échantillons d'eau de mer de quel ques litres.
- 2) Filtrer un volume fini (quelques centaines de millilitres à quelques litres) sur des filtres GF/F de 25 mm de diamètre sous une pression n'excédant pas 200 mbars.

3) Si attente pour analyse, transférer les filtres immédiatement pour centrifugation puis les sceller et les conserver à -80°C.

4) Extraire la chlorophylle- a avec 6 mL d'acétone à 90% pendant 24h à 5°C dans le noir.

5) Le surnageant, obtenu après centrifugation de 5 min, est mesuré pour son contenu en chlorophylle-a avec un fluorimètre (la calibration de l'instrument doit respecter le protocole JGOFS).

Connaissant les propriétés optiques de la chlorophylle, le composé ainsi extrait est ensuite analysé. La spectrophotométrie mesure la quantité de lumière absorbée par le composé à des longueurs d'onde spécifiques et doit être sensible aux longueurs d'onde relatives au rouge et à l'infrarouge pour détecter correctement la chlorophylle-a. La fluorométrie mesure la quantité de lumière émise à une longueur d'onde particulière, le composé étant exposé à une lumière de longueur d'onde différente. La mesure HPLC sépare les composants de l'extrait dissous préalablement dans un solvant par injections régulières d'échantillons d'un mélange de solvants.

Pour plus de renseignements, le lecteur pourra se référer à l'ouvrage d'Arnold Greenberg [1995].

2.3.3 - AMT (Atlantic Meridional Transect, UK)



Deux fois l'année, le navire de recherche *James Clark Ross* (ou *Discovery*) effectue une traversée méridionale de l'Océan Atlantique entre le Royaume-Uni et les îles Falkland (ou, parfois, Cape Town) : le trajet nord-sud se déroule aux alentours de septembre et le retour 6 mois plus tard, vers le mois d'avril de l'année suivante. Le programme AMT [Robins et Aiken, 1996 ; Aiken *et al.*, 2000] exploite ce circuit avec l'objectif principal d'étudier les processus physiques et biologiques ainsi que mesurer les propriétés bio-optiques dans l'Océan Atlantique, de la méso-échelle à l'échelle du bassin. Ces études visent principalement à calibrer et valider les données satellitaires de couleur de l'eau : d'abord en reliant les mesures *in situ* aux mesures satellitaires (cette opération est appelée « matchups » ; voir le Chapitre 2), puis, dans un second temps, en mesurant les constituants de l'eau qui ont de l'influence sur ses propriétés optiques. Entre 1995 et 2005, 17 trajets ont été réalisés, auxquels ont contribué près de 180 scientifiques de 11 pays. Une nouvelle phase (8 trajets entre 2007 et 2012) est en train d'être préparée.

Les données sont traitées et distribuées par le centre de données océanographiques anglais (BODC : British Oceanographic Data Centre). On peut les récupérer sur Internet : <http://www.bodc.ac.uk/projects/uk/amt/>. Les données utilisées sont qualifiées d' « underway » (une petite quantité d'eau de surface est pompée et des mesures

sont effectuées de façon continue à raison d'une mesure par minute) par opposition aux profils « bouteilles » (mesures journalières effectuées à midi, heure locale, à partir d'échantillons d'eau prélevés régulièrement jusqu'à 200 m de profondeur). Dans notre étude, nous utilisons des données de concentration en chlorophylle obtenue par fluorométrie (voir plus haut). Pour obtenir des données en unité traditionnelle (mg/m^3), le fluoromètre est calibré à l'aide des échantillons bouteilles mesurés dans les 10 premiers mètres.

Les deux jeux de données utilisés sont ceux des missions AMT12 et AMT13.

AMT12 est parti de Port Stanley le 12 mai 2003 et est arrivé au Royaume-Uni le 17 juin. Cette mission a donné lieu à 47431 mesures *in situ* de concentration en chlorophylle réparties selon la Figure 1.20 (ligne violet clair). L'équation de calibration utilisée est la suivante :

$$\text{Underway chlorophyll - a calibrated} = \frac{F - 0.113057}{0.855530}$$

où F est la mesure de chlorophylle par le fluoromètre.

AMT13 est parti du Royaume-Uni le 11 septembre 2003 et est arrivé Port-Stanley le 13 octobre. Cette mission a donné lieu à 46743 mesures *in situ* de concentration en chlorophylle réparties selon la Figure 1.20 (ligne violet foncé). L'équation de calibration utilisée est la suivante (F étant la mesure de chlorophylle par le fluoromètre) :

Section 1 (jusqu'au 19/09/2003/00 : 59) : Underway chlorophyll - a calibrated = $1.14 * F - 0.11$
Section 2 (à partir du 19/09/2003/01 : 00) : Underway chlorophyll - a calibrated = $2.76 * F - 0.27$

2.3.4 - SeaBASS (SeaWiFS Bio-optical And Storage System, USA)



Le second jeu de mesures *in situ* provient de la base de données SeaBASS de la NASA [Werdell et Bailey, 2002]. Les données sont collectées, dans le cadre de missions diverses, à l'aide de différents instruments (profileurs, bouteilles, ...) sur des plates-formes variées (dont les navires et les bouées). Les données récentes ne sont normalement pas disponibles sur le site Internet http://seabass.gsfc.nasa.gov/cgi-bin/pigment_search.cgi, mais une autorisation spéciale nous a été aimablement accordée par Watson GREGG (NASA) dans le cadre de notre participation au groupe de travail sur la combinaison de données couleur de l'eau de l'IOCCG (International Ocean-Colour Coordinating Group : <http://www.ioccg.org>).

Tableau 1.8 – Listing et caractéristiques des données *in situ* (programmes, bases de données et campagnes) de l'année 2003 utilisées. Les couleurs associées (colonne de droite) permettent de visualiser les dites données sur la Figure 1.20 de la page suivante.

Programmes et bases de données	Nom de la campagne	Dates			Zone				Nombre de points			Couleur associée
		Du	Au	Nombre de jours	Nord	Sud	Ouest	Est	Total	Grille 1/12°	Grille 1/24°	
AMT	AMT12	13/05/2003	15/06/2003	34	49,43	-50,32	-53,68	-4,82	47431	2349	4577	
	AMT13	11/09/2003	13/10/2003	33	51,11	-51,69	-57,83	1,54	46743	2292	4407	
ARGAU	ARGAU3	07/02/2003	19/05/2003	61	-38,50	-77,71	-70,08	-20,02	314	304	307	
DIAPAZON	DIAPAZON7	03/02/2003	11/02/2003	9	-20,60	-22,58	166,05	167,46	10	9	9	
	DIAPAZON8	09/06/2003	13/06/2003	7	-21,84	-23,06	166,21	167,50	45	22	27	
	DIAPAZON9	08/10/2003	14/10/2003	7	-20,00	-22,58	166,10	167,60	8	8	8	
DYFAMED				9	48,40	43,40	7,00	7,16	98	14	14	
ICES				233	62,00	51,53	-13,63	15,98	1150	1097	1127	
MINERVE	MINERVE 0203-R2	01/01/2003	27/01/2003	8	-44,85	-66,45	140,11	147,01	32	32	32	
	MINERVE 0203-R4	18/02/2003	04/03/2003	11	-43,72	-66,56	140,02	147,37	49	49	49	
	MINERVE 0304-R0	21/10/2003	11/11/2003	16	-43,98	-66,63	143,37	147,28	61	61	61	
	MINERVE 0304-R1	06/12/2003	31/12/2003	11	-43,57	-66,39	140,47	147,40	51	51	51	
NIO				35	23,07	9,00	70,22	88,00	92	55	61	
NODC				277	71,30	-22,66	-175,20	167,46	2970	2413	2581	
OISO	OISO10	29/01/2003	15/02/2003	16	-43,43	-66,42	138,45	146,43	43	43	43	
SeaBASS				171					1272	836	935	
Yves Dandonneau	COLIBRI	05/04/2003	09/05/2003	27	50,65	6,87	-51,65	9,85	96	96	96	
	GUYAPLAC	20/05/2003	02/06/2005	14	9,02	5,03	-52,43	-47,45	52	51	51	
	SKOGA03	27/03/2003	04/05/2003	6	59,11	42,49	-70,26	-32,95	34	34	34	
TOTAL				345					100551	9746	14399	

Donnees in situ disponibles pour 2003

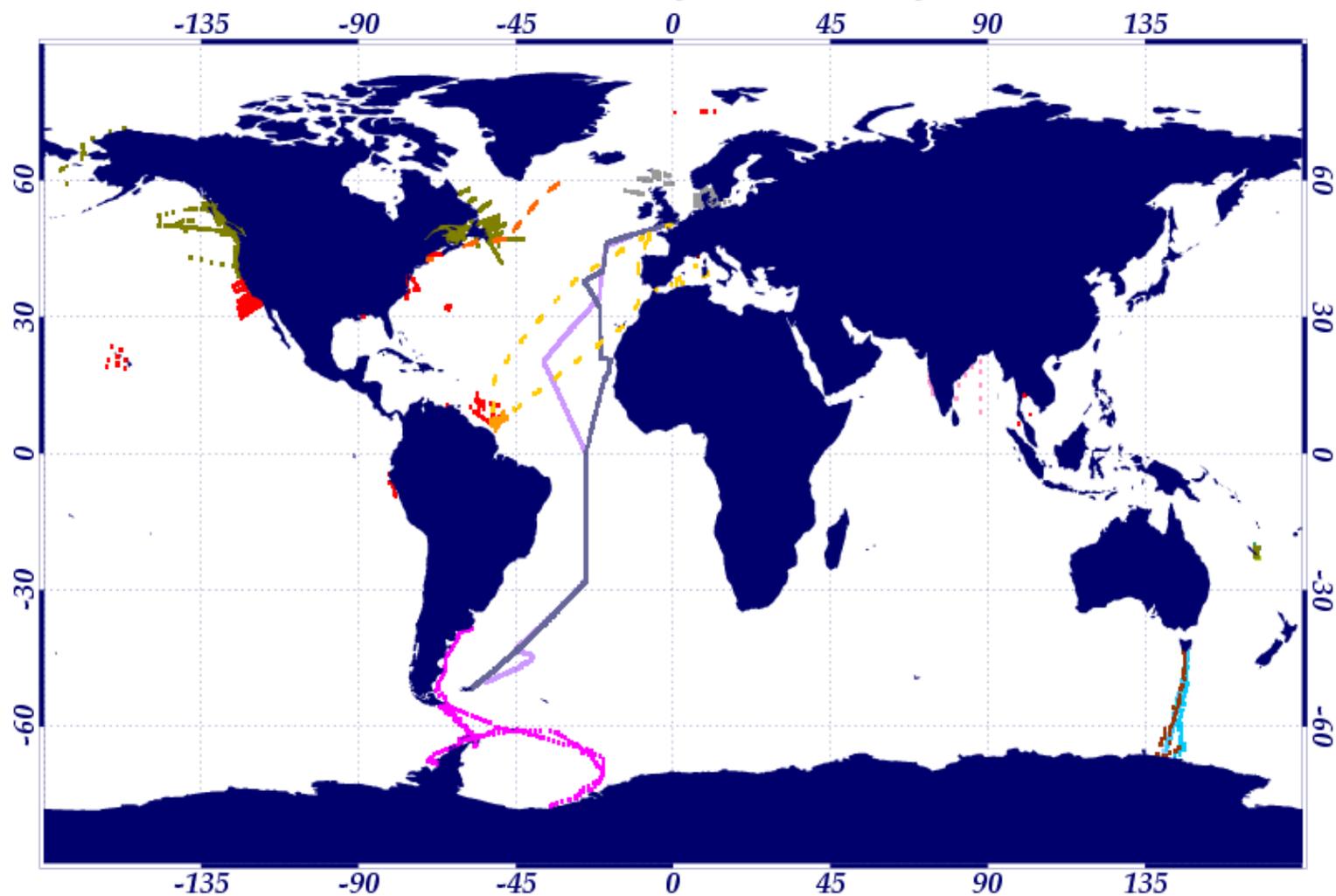


Figure 1.20 – Cartes indiquant l'emplacement des mesures *in situ* de l'année 2003 utilisées. Les couleurs sont celles définies dans le Tableau 1.8 de la page précédente.

Pour l'année 2003, il y a 1272 mesures *in situ* sur tout le globe. Il s'agit de valeurs de concentrations en chlorophylle mesurées à des profondeurs allant jusqu'à 5 m et obtenues par fluorométrie ou spectrophotométrie. Comme on peut le voir sur la Figure 1.20 (points rouges), ces données sont essentiellement concentrées au niveau de l'archipel d'Hawaï, des côtes américaines et au large de la Guyane.

2.3.5 - NODC (National Oceanographic Data Center, USA)



Le troisième jeu de données *in situ* provient de la base de données américaine NODC, dont le site Internet est <http://www.ndc.noaa.gov>. Comme dans le cas de SeaBASS, les données sont d'origines diverses.

Pour l'année 2003, il y a 2970 mesures *in situ* sur tout le globe. Là encore, il s'agit de valeurs de concentrations en chlorophylle mesurées à des profondeurs allant jusqu'à 5 m et obtenues par fluorométrie ou spectrophotométrie. Comme on peut le voir sur la Figure 1.20 (points vert kaki), ces données sont essentiellement concentrées au niveau des côtes américaines.

2.3.6 - CIEM (Conseil International pour l'Exploitation de la Mer)



Le CIEM est l'organisation qui coordonne et promeut la recherche marine dans l'Atlantique Nord, incluant aussi les mers adjacentes telles que la Mer Baltique et la Mer du Nord. Cette communauté regroupe plus de 1600 chercheurs provenant de 19 pays autour de l'Océan Atlantique Nord.

Pour l'année 2003, il y a 1150 mesures *in situ*, situées principalement autour du Royaume-Uni et du Danemark (cf. les points gris sur la Figure 1.20). Ces données sont disponibles sur Internet à l'adresse suivante : <http://www.ices.dk>.

2.3.7 - NIO (National Institute of Oceanography, Inde)



Le cinquième jeu d'observations *in situ* provient du centre de données de l'Institut National Océanographique d'Inde (Data Center of National Institute of Oceanography, Dona Paula, Goa, India).

Pour l'année 2003, avec les considérations de profondeur et de mesure énumérées plus haut, 92 mesures ont été effectuées [J.S. Sarupia, pers. com.] : 9, au large, par les navires indiens et 83 en côtier, réparties comme le montre la Figure 1.20 (points rose clair).

2.3.8 - Les campagnes OISO (Océan Indien Service d'Observation)



Le programme OISO (<http://www.ipsl.jussieu.fr/services/Observations/fr/OISO.htm>), labellisé Service d'Observation INSU en juillet 1997, est soutenu par 3 instituts : l'INSU (Institut National des Sciences de l'Univers du CNRS), l'IPEV (Institut Polaire français Paul-Emile Victor) et l'IPSL (Institut Pierre Simon Laplace). Ce programme met en place un réseau couplé d'observations océaniques et atmosphériques à long terme afin de mieux identifier et quantifier les variations des sources et puits de CO₂ océaniques, comprendre comment les échanges air-mer de CO₂ varient d'une saison à l'autre, d'une année à l'autre, estimer l'évolution de ces échanges en réponse à des changements climatiques et identifier le carbone anthropique dans l'océan et son évolution. Pour ses mesures, le programme utilise les trajets du Marion-Dufresne dans l'Océan Indien, navire affrété par l'IPEV. Il y a environ 2 campagnes par an, une en été austral, l'autre en hiver austral. Depuis le démarrage de l'opération en 1998, 12 campagnes ont été effectuées.

Les données utilisées sont les 43 mesures discrètes de surface de chlorophylle-a de la campagne OISO 10, du 23 janvier au 17 février 2003 [N. Metzl, pers.com.], réparties comme montré Figure 1.20 (points marron clair).

2.3.9 - Le programme MINERVE

Le programme MINERVE (Mesures à l'Interface Eau-air de la Variabilité des Echanges de CO₂) s'appuie sur des campagnes de valorisation de transit. Il a pour objectif d'observer et de comprendre les variabilités saisonnières de la pression partielle de CO₂ (pCO₂) et du carbone inorganique total dans les eaux de surface, en association avec les mesures hydrologiques et biogéochimiques *in situ* et avec l'aide de données satellitaires (température et couleur de l'eau). Les parcours du navire « Astrolabe » permettent d'accéder à des zones très peu étudiées et de les documenter afin de mieux comprendre les processus qui expliquent les variations spatio-temporelles de pCO₂ à moyenne échelle dans les régions océaniques australes. Les données sont utilisées pour faire des estimations de la variabilité interannuelle de flux net de CO₂ à l'interface océan-atmosphère.

Faisant suite au programme INDIVAT (1985-1987, à bord du Marion-Dufresne), le programme MINERVE a débuté en 1990. 40 campagnes ont été effectuées entre 1990 et 1997, la plupart dans la partie ouest de l'Océan Indien et le secteur austral indien, régions où très peu d'observations ont été faites avant 1990. Depuis 2002, l'appareillage a été modifié, installé sur l'Astrolabe et adapté aux conditions extrêmes de mesure rencontrées sur le trajet Hobart – Terre-Adélie. De 2002 à 2005, 3 campagnes ont été effectuées chaque année.



Les données utilisées sont les 193 mesures discrètes de surface de chlorophylle-*a* des campagnes MINERVE 0302-R2 (du 28 décembre 2002 au 27 janvier 2003), MINERVE 0203-R4 (du 18 février au 4 mars 2003), MINERVE 0304-R0 (du 21 octobre au 11 novembre 2003) et MINERVE 0302-R1 (du 6 décembre 2003 au 1^{er} janvier 2004) [B. Griffith, pers. com.], réparties comme montré Figure 1.20 (points bleus).

2.3.10 - Le programme ARGAU

Le programme de recherche ARGAU (« french projet in cooperation with ARGentina for the study of the AUstral atlantic ocean ») est basé sur une coopération entre la France et l'Argentine et a pour but principal d'étudier la variabilité à long terme de l'Océan Atlantique Sud occidental et austral. En effet, cette région de l'océan serait celle qui réagirait le plus rapidement et le plus intensément aux changements climatiques prédits comme conséquence de l'augmentation croissante de CO₂ atmosphérique [Sarmiento *et al.*, 1998].

Les campagnes ARGAU s'effectuent à bord du brise-glace argentin Almirante Irizar, appartenant à l'Institut Antarctique Argentin, dans le cadre soit de rotation d'été austral (d'une durée de 3 mois) permettant aussi le ravitaillement de bases militaires ; soit de campagnes d'hiver (de seulement 3 semaines à 1 mois), organisées exclusivement pour les études scientifiques.

Les données utilisées sont les 314 mesures discrètes de surface de chlorophylle-*a* de la campagne ARGAU3 (du 7 février au 19 mai 2003) fournies par le Dr D. Ruiz Pino (LOCEAN, Paris), réparties comme montré Figure 1.20 (points rose vif).

2.3.11 - Le programme DIAPAZON

Le programme DIAPAZON est l'un des projets de l'unité de recherche IRD CYANO. L'objectif est d'améliorer la connaissance de la diazotrophie¹ dans l'océan et de son rôle dans le cycle global du carbone.

Les 63 mesures utilisées sont issues des campagnes DIAPALIS 7 (3-11 février 2003 : 10 mesures), DIAPALIS 8 (10-14 juin 2003 : 45 mesures) et DIAPALIS 9 (7-16 octobre 2003 : 8 mesures). Il s'agit d'allers-retours entre Nouméa et les Iles Loyautés (Nouvelle-Calédonie), comme montré Figure 1.20 (points verts).

¹ Fixation biologique du diazote. Il s'agit d'un processus métabolique exclusivement réalisé par les organismes procaryotes (organismes unicellulaires dont le noyau est dépourvu de membrane et réduit à un seul chromosome). Il est responsable de la conversion du diazote (très abondant dans la nature mais relativement inerte) en substrats biologiquement utilisables. C'est le mécanisme principal permettant l'introduction d'azote dans la biosphère.

2.3.12 - Le programme DYFAMED



L'opération DYFAMED (DYnamique des Flux Atmosphériques en MEDiterranée) est une étude à long terme de l'environnement méditerranéen basée sur l'observation et la prédiction des cycles biogéochimiques du carbone et des substances qui lui sont associées. L'objectif est d'étudier les variations temporelles à diverses échelles, haute fréquence saisonnière et interannuelle, sur la dimension verticale, depuis l'atmosphère jusqu'au sédiment. En effet, la particularité du site (cf. le point marron foncé au large de Nice, Figure 1.20) est d'être situé dans une zone à faible advection protégée des apports continentaux directs autres qu'atmosphériques.

Les 98 mesures utilisées ont été récupérées sur le site Internet <http://www.obs-vlfr.fr/sodyf/home.htm>.

2.3.13 - Campagnes ponctuelles

Le Dr Y. Dandonneau (LOCEAN, Paris) m'a généreusement fourni les mesures qu'il a obtenues en 2003 durant 3 trajets :

-  Le navire COLIBRI a fait l'aller-retour France-Guyane entre le 5 avril et le 9 mai 2003. Il y a 96 mesures *in situ*, réparties comme indiquées sur la Figure 1.20 (points jaunes).
-  La campagne GUYAPLAC a consisté en une section au large de la Guyane, entre les 20 mai et 2 juin 2003. Il en résulte 52 mesures *in situ* (points orange pâle sur la Figure 1.20).
-  La campagne SKOGA03 est partie d'Islande le 27 avril 2003 à destination de Terre-Neuve où elle est arrivée le 4 mai 2003. Il y a 34 mesures *in situ* de concentration en chlorophylle-a. Leur répartition est indiquée sur la Figure 1.20 (points orange foncé).

2.3.14 - Conclusion

Dans l'objectif d'apprécier quantitativement les données satellitaires ou les données combinées, il est nécessaire de les comparer aux observations *in situ* pour déterminer l'erreur dont elles sont entachées. Comme il n'existe malheureusement pas de base de données internationale des mesures *in situ*, il est souvent difficile de se les procurer. De plus, les campagnes telles AMT, ayant une couverture spatiale très intéressante, sont malheureusement trop peu nombreuses.

Néanmoins, toutes les campagnes décrites précédemment ont fourni 100551 mesures de concentration en chlorophylle-a au cours de l'année 2003. Sur la Figure 1.20, on peut constater que ces mesures sont essentiellement présentes dans l'Océan Atlantique, le nord de l'Océan Pacifique et le sud de l'Australie.



3 - Conclusion du chapitre

Dans ce chapitre, nous avons tout d'abord abordé la notion de couleur de l'eau. Nous avons vu l'importance de la connaissance de la variabilité spatio-temporelle du phytoplancton, tant sur le plan environnemental (rôle de premier ordre dans le système Terre, indicateur de pollution) que sur le plan économique (gestion des zones de pêches). La composante principale du phytoplancton, la chlorophylle-*a*, peut se mesurer *in situ* (à partir de navires, plates-formes scientifiques, bouées...) – les données sont alors peu entachées d'erreur mais peu nombreuses – et de l'espace (couverture du globe en seulement deux jours).

Dans le chapitre suivant, nous allons nous intéresser à l'étude quantitative de ces données.

Chapitre 2

Etude des données

CHAPITRE 2

ÉTUDE DES DONNÉES

Après avoir présenté la biologie océanique et plus particulièrement le phytoplancton, nous allons nous intéresser, dans ce chapitre, aux facettes quantitatives des données de concentration en chlorophylle-a. Dans un premier temps, nous nous attachons à décrire la répartition spatiale du phytoplancton, qui nous amène à considérer la biologie de l'océan selon différentes régions. Puis, nous étudions la distribution particulière des données, qui a une conséquence directe sur tout calcul statistique qui leur est appliqué. Nous voyons ensuite que les données satellitaires de concentrations en chlorophylle sont entachées d'une erreur de mesure et d'un biais qu'il est possible d'évaluer en les comparant aux mesures *in situ*. Enfin, nous étudions le biais existant entre les données des différents capteurs pris en compte dans notre étude.

Sommaire

1 – PROVINCES BIOGEOCHIMIQUES.....	69
2 – UNE REPARTITION LOGNORMALE.....	71
3 – COMPARAISON DES DONNEES SATELLITAIRES AVEC LES MESURES <i>IN SITU</i>	73
3.1 – <i>Introduction</i>	73
3.2 – <i>Définition des formules de comparaison par l'exemple</i>	73
3.3 – <i>Résultat des « match-ups »</i>	78
3.4 – <i>Construction des cartes d'erreur et de biais</i>	84
3.5 – <i>Conclusion</i>	88
4 – INTERCOMPARAISON DES DONNEES SATELLITAIRES.....	88
4.1 – <i>Introduction</i>	88
4.2 – <i>Méthodologie</i>	88
4.3 – <i>Résultats et discussion</i>	89
4.4 – <i>Conclusion</i>	90
5 – CONCLUSION DU CHAPITRE.....	90

1 - Provinces biogéochimiques

Il est possible de diviser l'océan en régions ou « provinces » ayant chacune un comportement biologique et physique qui lui est propre : ce sont les **provinces biogéochimiques**, concept introduit par Longhurst il y a 11 ans [Longhurst, 1995]. Les provinces biogéochimiques globales sont celles montrées sur la Figure 2.1. Cette distribution en provinces avec des frontières fixes est cependant peu conforme à la fluidité de la réalité. Mélin [2003] propose une partition du bassin qui est dynamique, évoluant saisonnièrement. Néanmoins, pour nos études, nous garderons une approche simplifiée et utiliserons des frontières fixes.

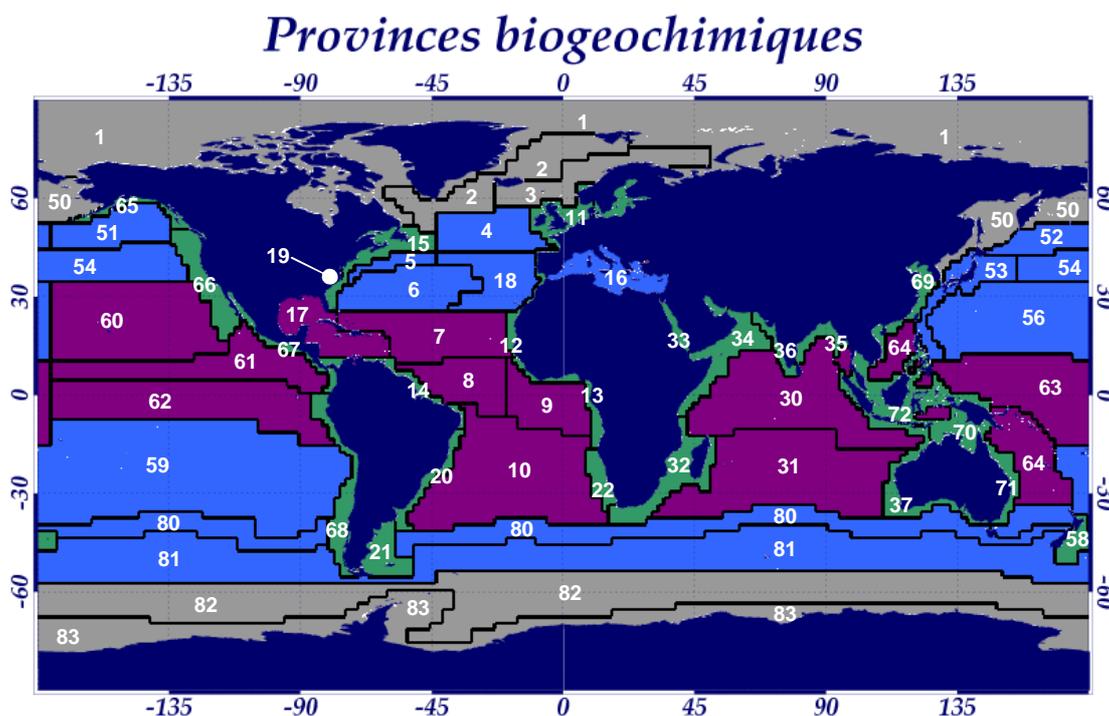


Figure 2.1 – Emplacement des provinces biogéochimiques telles que définies par Longhurst [1998]. Les provinces appartenant au biome côtier sont en vert, celles appartenant au biome polaire en gris, celles appartenant au biome des coups de vent d'ouest en bleu et celles appartenant au biome des alizés en violet. La légende des numéros des provinces est indiquée dans le Tableau 2.1.

Tableau 2.1 – Provinces biogéochimiques [Longhurst, 1998].

Province #	Biome	Province
1	Polaire	Province polaire boréale
2	Polaire	Province Atlantique Arctique
3	Polaire	Province Atlantique Subarctique
4	Coups de vent d'Ouest	Province de la Dérive Nord-Atlantique
5	Coups de vent d'Ouest	Province du Gulf Stream
6	Coups de vent d'Ouest	Province du Gyre Subtropical Nord-Atlantique (Ouest)
7	Alizés	Province du Gyre Tropical Nord-Atlantique
8	Alizés	Province de l'Atlantique Tropical Ouest
9	Alizés	Province de l'Atlantique Tropical Est
10	Alizés	Province du Gyre Sud-Atlantique
11	Côtier	Province des Plateaux Atlantique Nord-Est
12	Côtier	Province côtière des Canaries
13	Côtier	Province côtière du courant de Guinée
14	Côtier	Province côtière de Guyane
15	Côtier	Province des Plateaux Atlantique Nord-Ouest
16	Coups de vent d'Ouest	Province de la Mer Méditerranée et de la Mer Noire
17	Alizés	Province des Caraïbes
18	Coups de vent d'Ouest	Province du Gyre Subtropical Nord-Atlantique (Est)
19	Côtier	Province de la baie de Chesapeake
20	Côtier	Province côtière du Courant du Brésil
21	Côtier	Province côtière des Plateaux Atlantique Sud-Ouest
22	Côtier	Province côtière du Courant du Benguela
30	Alizés	Province du Gyre de la Mousson Indienne
31	Alizés	Province du Gyre Subtropical Indien Sud
32	Côtier	Province côtière de l'Afrique de l'Est
33	Côtier	Province de la Mer Rouge et du Golfe Persan
34	Côtier	Province de l'upwelling arabe Nord-Ouest
35	Côtier	Province côtière Indienne Est
36	Côtier	Province côtière Indienne Ouest
37	Côtier	Province de l'Australie
50	Polaire	Province Epicontinentale Pacifique Nord
51	Coups de vent d'Ouest	Province du Gyre Subarctique Pacifique (Est)
52	Coups de vent d'Ouest	Province du Gyre Subarctique Pacifique (Ouest)
53	Coups de vent d'Ouest	Province du Courant Kuroshio
54	Coups de vent d'Ouest	Province du Front Polaire Pacifique Nord
56	Coups de vent d'Ouest	Province du Gyre Subtropical Pacifique Nord
58	Coups de vent d'Ouest	Province de la Mer de Tasmanie
59	Coups de vent d'Ouest	Province du Gyre Subtropical Pacifique Sud
60	Alizés	Province du Gyre Tropical Pacifique Nord
61	Alizés	Province du Contre-courant Equatorial Pacifique Nord
62	Alizés	Province de la Divergence Equatoriale Pacifique
63	Alizés	Province de la Warm Pool Pacifique Ouest
64	Alizés	Province des Bassins Profonds Archipélagiques
65	Côtier	Province du downwelling d'Alaska
66	Côtier	Province de l'upwelling de Californie
67	Côtier	Province de l'Amérique Centrale
68	Côtier	Province du Chili
69	Côtier	Province de la Mer de Chine
70	Côtier	Province de la Sonde
71	Côtier	Province Australienne Est
72	Côtier	Province de la Nouvelle-Zélande
80	Coups de vent d'Ouest	Province de la Convergence Subtropicale Sud
81	Coups de vent d'Ouest	Province Subantarctique
82	Polaire	Province Antarctique
83	Polaire	Province Australe
99	Lacs	Lacs, Mer Caspienne, Mer d'Aral

Certaines provinces ont des caractéristiques similaires. La classification de la géographie écologique des océans suggérée par Longhurst [1998] s'appuie sur la définition de la couche de mélange océanique : couche de la surface de la mer, homogénéisée par la turbulence induite par la tension du vent, se situant proche de l'interface océan-atmosphère. Elle se fait selon 4 « biomes » (Tableau 2.1) :

-  le **biome côtier** : où la profondeur de la couche de mélange est forcée par divers processus côtiers,
-  le **biome polaire** : où la profondeur de la couche de mélange est figée par une couche de surface saumâtre¹ qui se forme chaque printemps dans les zones de glace,
-  le **biome des coups de vent d'Ouest** : où la profondeur de la couche de mélange est largement forcée par des vents locaux et le rayonnement solaire,
-  le **biome des alizés** : où la profondeur de la couche de mélange est forcée par un ajustement géostrophique² de l'échelle d'un bassin océanique par un forçage de vents souvent éloignés.

Intéressons-nous maintenant à la répartition particulière des données de concentration en chlorophylle-a.

2 - Une répartition lognormale

De nombreuses mesures *in situ* en chlorophylle, généralement utilisées pour tester des algorithmes de production primaire [Balch *et al.*, 1992], semblent être distribuées **lognormalement** (Figure 2.2), c'est-à-dire que le logarithme de ces valeurs a une distribution normale [Campbell, 1995].

Néanmoins, la carte de localisation des stations de mesure [Balch *et al.*, 1992 ; Figure 2.2] montre que la plupart d'entre elles sont dans l'Hémisphère Nord et que les mesures près des côtes sont échantillonnées de manière disproportionnée comparé aux centres des gyres océaniques.

¹ Caractéristique d'une eau dont la salinité est inférieure à celle de l'eau de mer (eau qui se forme dans les zones où la mer se mêle à l'eau douce).

² Equilibre géostrophique : équilibre entre les gradients horizontaux de pression et les forces de Coriolis dans l'océan.

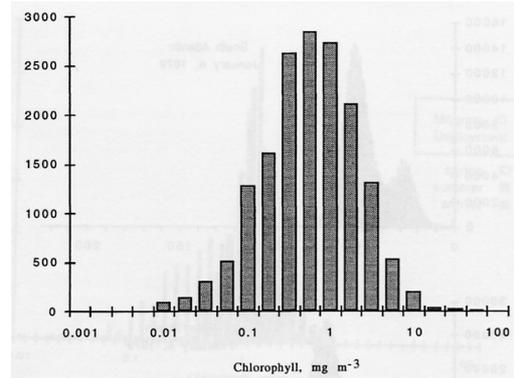


Figure 2.2 – Histogramme de répartition de 16364 mesures *in situ* de concentration en chlorophylle dans l'océan, à partir d'une compilation de Balch *et al.* [1992]. Ces mesures ont été réalisées sur l'océan global, mais l'échantillonnage est plus important au niveau des latitudes moyennes de l'Hémisphère Nord, et les régions centrales des gyres océaniques ont été sous-échantillonnées [Campbell, 1995].

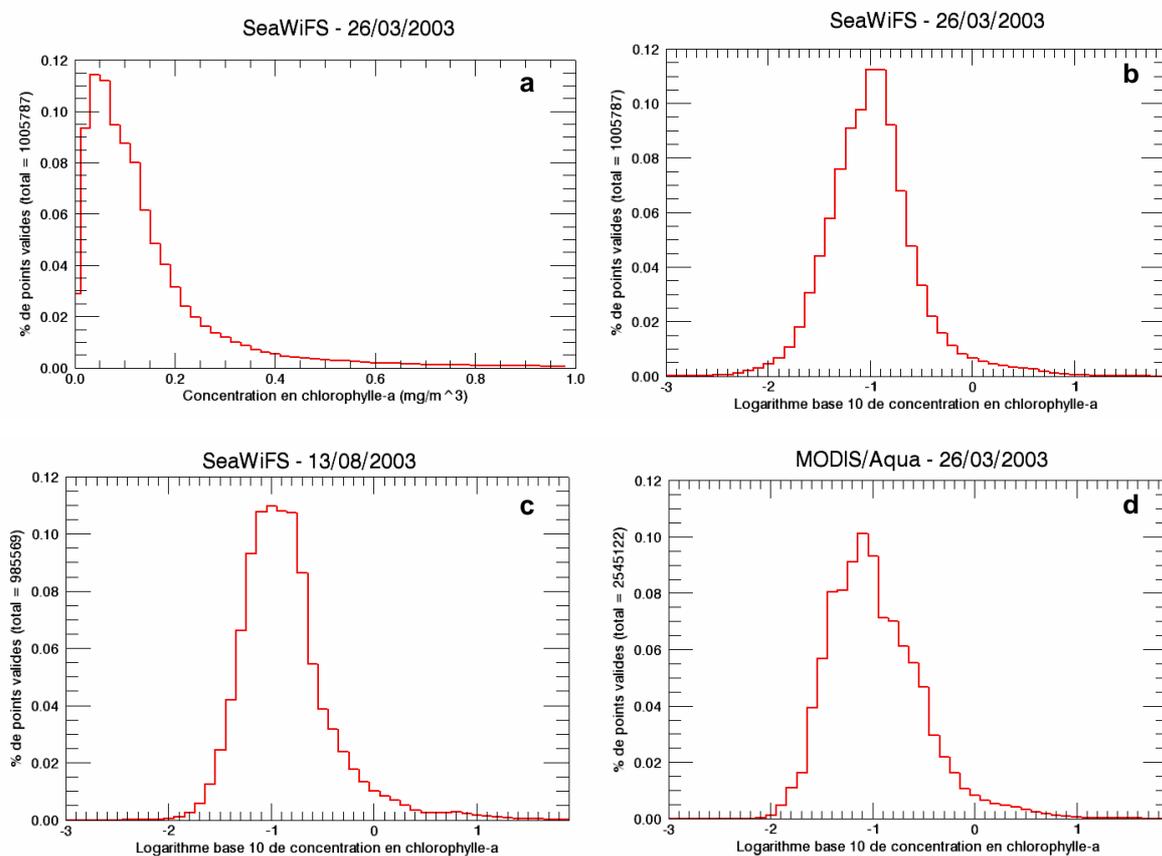


Figure 2.3 – (a) Distribution de la concentration en chlorophylle-*a* de SeaWiFS pour le 26 mars 2003. (b) Distribution du logarithme base 10 de la concentration en chlorophylle-*a* de SeaWiFS pour le 26 mars 2003. (c) Distribution du logarithme base 10 de la concentration en chlorophylle-*a* de SeaWiFS pour le 13 août 2003. (d) Distribution du logarithme base 10 de la concentration en chlorophylle-*a* de MODIS/Aqua pour le 26 mars 2003.

La Figure 2.3 montre la distribution des données globales de concentration en chlorophylle-*a* provenant du capteur SeaWiFS pour le 26 mars 2003 (a) ainsi que la

distribution du logarithme base 10 de ces mêmes données **(b)**. Il est donc clair que cette distribution est lognormale, confirmant les résultats de Campbell [1995]. La Figure 2.3c confirme ce résultat pour le même capteur mais un autre jour, le 13 août 2003. La Figure 2.3d confirme aussi ce résultat pour le même jour (26 mars 2003) mais pour un autre capteur, MODIS/Aqua. On peut noter que la courbe de distribution pour le 26 mars 2003 n'est pas identique pour les capteurs SeaWiFS et MODIS/Aqua, laissant présager une différence notable entre les données issues de ces deux capteurs. Cette étude d'intercomparaison entre les capteurs est menée dans la partie 4.

3 - Comparaison des données satellitaires avec les mesures *in situ*

3.1 - Introduction

Pour déterminer la qualité d'un jeu de données satellitaires ou d'un jeu de données combinées de chlorophylle-*a*, la méthode généralement utilisée consiste à comparer ces données avec des mesures *in situ* (ou « terrains »), collectées à partir de bateaux ou plateformes océaniques (voir le chapitre précédent pour les détails concernant les différents types de mesure).

Suivant la méthode proposée par Gregg et Casey [2004], les mesures *in situ* sont tout d'abord mises sur la grille régulière correspondant au jeu de données à analyser. Pour cela, les mesures *in situ* coïncidentes (i.e. prises le même jour) et localisées sur le même pixel de la grille d'étude sont moyennées. Ensuite, il s'agit de comparer chaque donnée satellitaire à la mesure *in situ* coïncidente dans le temps et dans l'espace, si elle existe : cette opération est appelée « matchup ».

3.2 - Définition des formules de comparaison par l'exemple

Pour bien appréhender la notion de statistiques sur des données lognormalement distribuées, regardons le petit exemple suivant. Le Tableau 2.2 montre dix matchups sélectionnés régulièrement parmi les matchups SeaWiFS de l'année 2003 sur l'océan global : une fois les matchups rangés selon les données *in situ* croissantes, on divise en 10 parties et on prend la valeur centrale de chaque groupe. Il s'agit donc d'un échantillon représentatif de la distribution de la concentration en chlorophylle. Cette partie a pour objet

de préciser les formulations relatives à 3 quantités statistiques qui seront nécessaires par la suite : le coefficient de corrélation, l'erreur de mesure et le biais.

Tableau 2.2 - Dix des matchups relatifs au capteur SeaWiFS.

Matchups #	1	2	3	4	5	6	7	8	9	10
Données <i>in situ</i> (en mg/m ³)	0.02	0.03	0.05	0.08	0.12	0.19	0.30	0.52	1.09	24.71
Données satellitaires (en mg/m ³)	0.04	0.03	0.10	0.10	0.18	0.29	1.39	0.74	1.42	13.70

3.2.1 - Calcul sur les données « brutes »

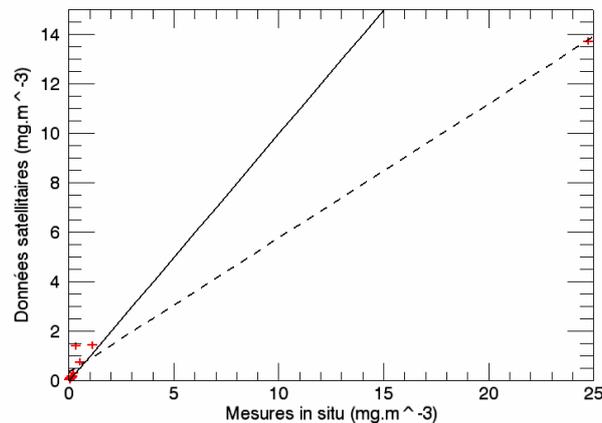


Figure 2.4 – Représentation des données satellitaires en fonction des mesures *in situ*. La droite en trait plein est la droite de régression linéaire idéale, obtenue quand les données satellitaires sont égales aux données *in situ* ; la droite en pointillés est la droite de régression linéaire réelle.

Le coefficient de détermination (i.e. le carré du coefficient de corrélation) renseigne sur le degré de similitude entre deux séries, soit, dans le cas qui nous concerne, les données satellitaires S_n et les mesures *in situ* I_n correspondantes. Sa formulation pour les valeurs brutes est la suivante :

$$r_d^2 = \frac{\left(\sum_{n=1}^N (S_n - \bar{S})(I_n - \bar{I}) \right)^2}{\left(\sum_{n=1}^N (S_n - \bar{S})^2 \right) \left(\sum_{n=1}^N (I_n - \bar{I})^2 \right)} \quad \text{avec } \bar{S} = \frac{1}{N} \sum_{n=1}^N S_n \quad \text{et } \bar{I} = \frac{1}{N} \sum_{n=1}^N I_n$$

où S_n est la $n^{\text{ième}}$ mesure satellitaire, I_n la $n^{\text{ième}}$ mesure *in situ* et N le nombre de matchups. Dans notre exemple, $r_d^2 = 0.98$.

On cherche à exprimer une relation linéaire entre S_n et I_n . On cherche donc les paramètres α_d et β_d tels que $S_n = \alpha_d I_n + \beta_d + \varepsilon_n$, où ε_n est l'erreur. Dans le cas idéal, c'est-

à-dire pour $S_n = I_n$, la pente de la droite de régression est $\alpha_d = 1$ et l'ordonnée à l'origine $\beta_d = 0$. Pour déterminer α_d et β_d , on cherche à minimiser ε_n par moindres carrés, c'est-à-dire à minimiser la somme des carrés des écarts résiduelle définie par :

$$SS_r = \sum_{n=1}^N (\varepsilon_n)^2 = \sum_{n=1}^N (S_n - \alpha_d I_n - \beta_d)^2 = \Phi(\alpha_d, \beta_d).$$

Lorsque cette fonction est minimale, les dérivées partielles par rapport à ces paramètres s'annulent, soit :

$$\begin{cases} \frac{\partial \Phi}{\partial \alpha_d} = -2 \sum_{n=1}^N I_n (S_n - \alpha_d I_n - \beta_d) = 0 \\ \frac{\partial \Phi}{\partial \beta_d} = -2 \sum_{n=1}^N (S_n - \alpha_d I_n - \beta_d) = 0 \end{cases} \quad \text{soit} \quad \begin{cases} \alpha_d \sum_{n=1}^N I_n^2 + \beta_d \sum_{n=1}^N I_n = \sum_{n=1}^N I_n S_n \\ \alpha_d \sum_{n=1}^N I_n + \beta_d N = \sum_{n=1}^N S_n \end{cases}.$$

C'est ce que l'on appelle communément le « système des équations normales ». Il admet pour solution :

$$\alpha_d = \frac{N \sum_{n=1}^N I_n S_n - \sum_{n=1}^N I_n \sum_{n=1}^N S_n}{N \sum_{n=1}^N I_n^2 - \left(\sum_{n=1}^N I_n \right)^2} : \text{ dans notre exemple, la pente de la droite de régression}$$

linéaire est $\alpha_d = 0.54$.

$$\beta_d (\text{mg/m}^3) = \frac{\sum_{n=1}^N I_n^2 \sum_{n=1}^N S_n - \sum_{n=1}^N I_n \sum_{n=1}^N I_n S_n}{N \sum_{n=1}^N I_n^2 - \left(\sum_{n=1}^N I_n \right)^2} : \text{ dans notre exemple, l'ordonnée à l'origine de la}$$

droite de régression linéaire est $\beta_d = 0.33 \text{ mg/m}^3$.

L'erreur de mesure est considérée comme étant l'écart-type entre la donnée satellitaire et la mesure *in situ* correspondante. Cette quantité est fournie par l'erreur RMS (Root Mean Square) dont la formulation est la suivante :

$$\text{RMS}_d (\text{mg/m}^3) = \sqrt{\frac{1}{N} \sum_{n=1}^N (S_n - I_n)^2}$$

Dans notre exemple, $\text{RMS}_d = 3.50 \text{ mg/m}^3$.

Quant au biais, il est défini comme étant la moyenne des écarts des données satellitaires par rapport aux mesures *in situ*. C'est aussi l'ordonnée à l'origine de la droite de régression si on impose que la pente de celle-ci soit égale à 1 :

$$\text{BIAIS}_d(\text{mg}/\text{m}^3) = \frac{1}{N} \sum_{n=1}^N (S_n - I_n).$$

Dans notre exemple, $\text{BIAIS}_d = -0.91 \text{ mg}/\text{m}^3$.

Par cohérence physique, on devrait obtenir α_d proche de 1 et β_d proche de 0, avec une erreur de mesure et un biais petits. Or, ce n'est pas le cas : le modèle linéaire de dépendance entre S et I n'est pas physiquement cohérent. Il n'est donc pas correct de réaliser nos calculs statistiques sur les valeurs brutes.

Nous avons vu, dans la partie précédente, que les données de concentration en chlorophylle ont une distribution lognormale. Regardons donc maintenant les calculs de statistiques sur le logarithme base 10 des données.

3.2.2 - Calculs sur le logarithme base 10 des données

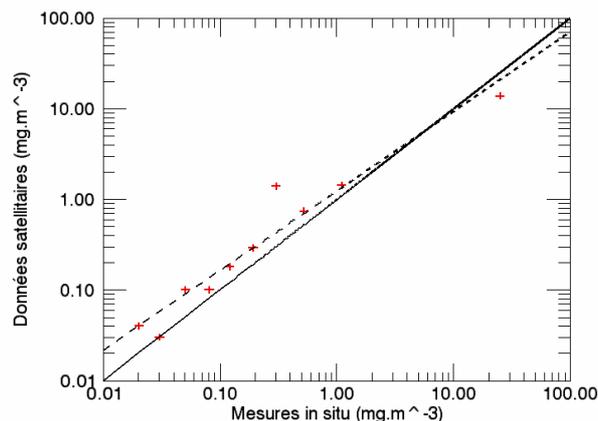


Figure 2.5 – Représentation du logarithme base 10 des données satellitaires en fonction de celui des mesures *in situ*. La droite en trait plein est la droite de régression linéaire idéale, obtenue quand les données satellitaires sont égales aux données *in situ* ; la droite en pointillés est la droite de régression linéaire réelle.

La formulation du coefficient de détermination entre le logarithme base 10 des deux jeux de données est la suivante :

$$r_1^2 = \frac{\left(\sum_{n=1}^N (\log_{10}(S_n) - \bar{S})(\log_{10}(I_n) - \bar{I}) \right)^2}{\left(\sum_{n=1}^N (\log_{10}(S_n) - \bar{S})^2 \right) \left(\sum_{n=1}^N (\log_{10}(I_n) - \bar{I})^2 \right)} \quad \text{avec } \bar{S} = \frac{1}{N} \sum_{n=1}^N \log_{10}(S_n) \text{ et } \bar{I} = \frac{1}{N} \sum_{n=1}^N \log_{10}(I_n)$$

Dans notre exemple, $r_1^2 = 0.935$.

Les données de concentration en chlorophylle ont une distribution lognormale. Or, les variables d'une distribution du type lognormal ne sont pas additives (i.e. la somme de deux variables lognormales n'est pas une variable de type lognormal) mais multiplicatives (le produit de deux variables lognormales est une variable de type lognormal). Ainsi, l'erreur recherchée est plutôt multiplicative qu'additive et s'exprime en pourcentage de la valeur : $S_n = (1 + \varepsilon_n) I_n$. La transformée en logarithme base 10 de cette équation donne $\log_{10}(S_n) = \log_{10}(I_n) + \log_{10}(1 + \varepsilon_n)$. On cherche, là encore, à exprimer la relation linéaire liant $\log_{10}(S_n)$ et $\log_{10}(I_n)$, c'est-à-dire déterminer α_1 et β_1 tels que $\log_{10}(S_n) = \alpha_1 \log_{10}(I_n) + \beta_1 + \log_{10}(1 + \varepsilon_n)$. De manière analogue à précédemment, on cherche à minimiser l'erreur $\log_{10}(1 + \varepsilon_n)$ et on obtient :

 α_1 , la pente de la droite de régression linéaire :

$$\alpha_1 = \frac{N \sum_{n=1}^N \log_{10}(I_n) \log_{10}(S_n) - \sum_{n=1}^N \log_{10}(I_n) \sum_{n=1}^N \log_{10}(S_n)}{N \sum_{n=1}^N (\log_{10}(I_n))^2 - \left(\sum_{n=1}^N \log_{10}(I_n) \right)^2}$$

Dans notre exemple, $\alpha_1 = 0.88$.

 β_1 , l'ordonnée à l'origine de la droite de régression linéaire :

$$\beta_1 = \frac{\sum_{n=1}^N \log_{10}(I_n^2) \sum_{n=1}^N \log_{10}(S_n) - \sum_{n=1}^N \log_{10}(I_n) \sum_{n=1}^N \log_{10}(I_n) \log_{10}(S_n)}{N \sum_{n=1}^N \log_{10}(I_n^2) - \left(\sum_{n=1}^N \log_{10}(I_n) \right)^2}$$

Dans notre exemple, $\beta_1 = 0.09$.

Contrairement à précédemment, ces valeurs sont proches des valeurs idéales, à savoir $\alpha_1 = 1$ et $\beta_1 = 0$, nous indiquant que nos considérations sont correctes.

L'erreur de mesure calculée sur le logarithme base 10 des données s'obtient suivant la formule de l'erreur RMS :

$$\text{RMS}_1 = \sqrt{\frac{1}{N} \sum_{n=1}^N (\log_{10}(S_n) - \log_{10}(I_n))^2}$$

Dans notre exemple, $\text{RMS}_1 = 0.28$.

La relation liant l'erreur de mesure calculée sur le logarithme base 10 des données RMS_I à l'erreur de mesure exprimée en pourcentage de valeur de chlorophylle $RMS_{d\%}$ est la suivante :

$$RMS_I = \log_{10}(1 + RMS_{d\%}) \text{ soit } RMS_{d\%} = 10^{RMS_I} - 1$$

Et donc : $RMS_{d\%} = 10^{\sqrt{\frac{1}{N} \sum_{n=1}^N (\log_{10}(S_n) - \log_{10}(I_n))^2}} - 1$, soit $RMS_{d\%} = 92\%$ dans notre exemple.

Le biais calculé sur le logarithme base 10 des données s'obtient de manière similaire aux données brutes :

$$BIAIS_I = \frac{1}{N} \sum_{n=1}^N (\log_{10}(S_n) - \log_{10}(I_n))$$

Dans l'exemple, $BIAIS_I = 0.17$.

De même que pour l'erreur de mesure, il existe une relation liant le biais calculé sur le logarithme base 10 des données et le biais exprimé en pourcentage de valeur de chlorophylle :

$$BIAIS_I = \log_{10}(1 + BIAIS_{d\%}) \text{ soit } BIAIS_{d\%} = 10^{BIAIS_I} - 1$$

Soit $BIAIS_{d\%} = 10^{\frac{1}{N} \sum_{n=1}^N (\log_{10}(S_n) - \log_{10}(I_n))} - 1$. L'application numérique pour notre exemple donne : $BIAIS_{d\%} = 49\%$.

3.2.3 - Conclusion

En raison de la distribution lognormale des données de concentration en chlorophylle-a, les statistiques d'erreur de mesure et de biais seront calculées, dans la suite de cette étude, sur le logarithme base 10 des données. De là, ces paramètres peuvent être exprimés en pourcentage de la valeur de concentration en chlorophylle-a.

3.3 - Résultat des « matchups »

3.3.1 - Méthodologie de création des « matchups »

Les mesures *in situ* **coïncidentes** (i.e. mesurées le même jour) et **co-localisées** (i.e. localisées sur l'emplacement d'un pixel de chlorophylle-a SeaWiFS ou MODIS/Aqua ; cf. Figure 1.16) sont tout d'abord moyennées, suivant la méthodologie de Gregg et Casey [2004]. Cela conduit à 9746 données *in situ* sur une grille $1/12^\circ \times 1/12^\circ$ et 14399 sur une grille $1/24^\circ \times 1/24^\circ$ (cf. Tableau 1.8). Après avoir considéré les moyennes coïncidentes et co-localisées ainsi que les pixels SeaWiFS et MODIS/Aqua valides, on obtient finalement 2410

points de comparaison pour SeaWiFS et 2789 pour MODIS/Aqua sur l'année 2003 et au niveau global. Les points de matchups pour les deux capteurs sont répartis comme indiqués Figure 2.6. Les analyses statistiques ont été réalisées globalement et régionalement, par provinces biogéochimiques telles que définies par Longhurst [1998] (cf. partie 1).

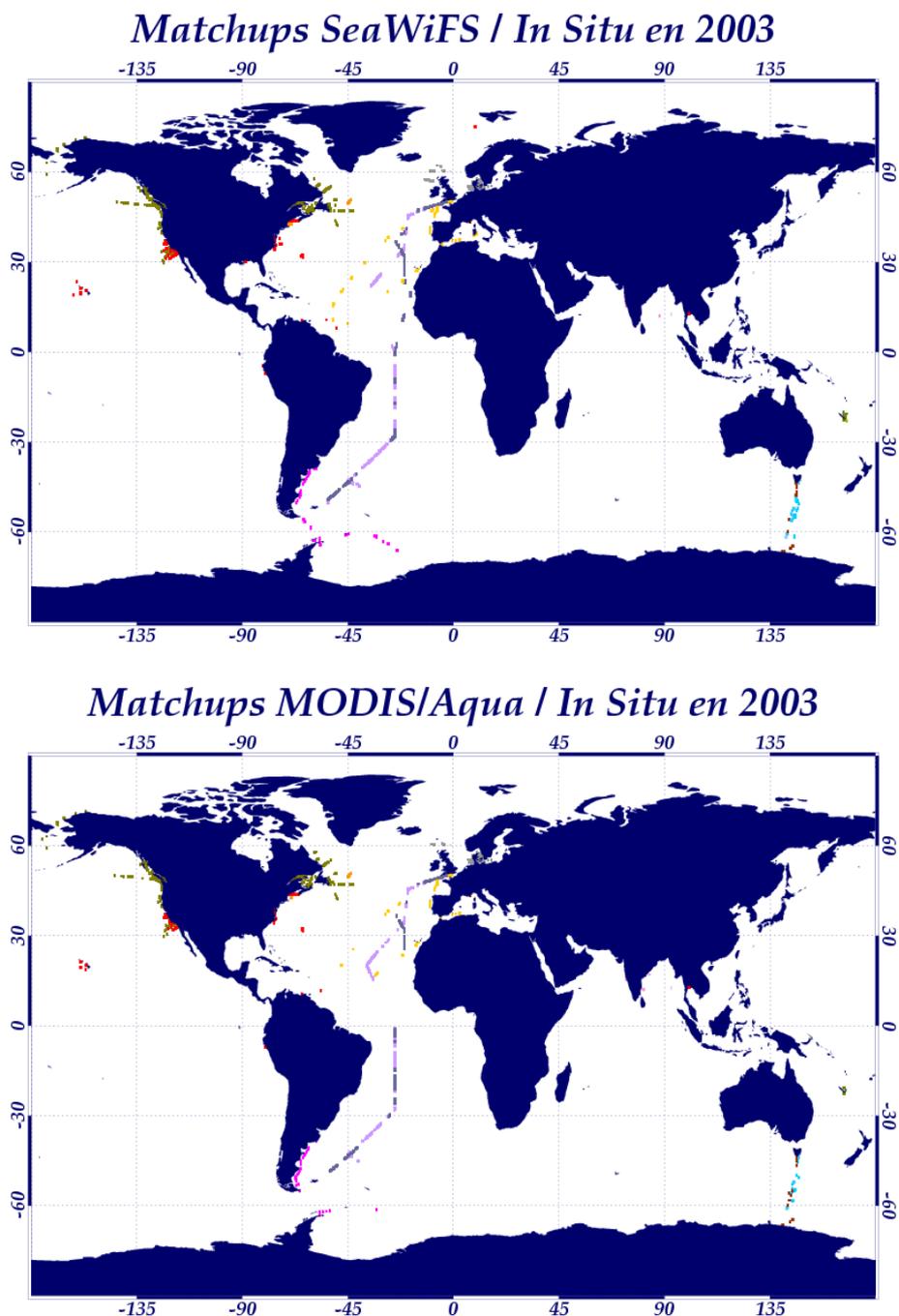


Figure 2.6 – Emplacement des 2410 matchups de SeaWiFS (en haut) et des 2789 matchups de MODIS/Aqua (en bas) sur l'année 2003. Les couleurs sont celles définies dans le Tableau 1.8 du chapitre précédent.

3.3.2 - Comparaison des données satellitaires et des observations *in situ*

On peut tout d'abord constater que les concentrations en chlorophylle mesurées par satellite et inférieures à 1 mg/m^3 , généralement trouvées en océan ouvert, concordent mieux avec les observations *in situ* que les concentrations en chlorophylle supérieures à 1 mg/m^3 , généralement trouvées dans les provinces côtières. En effet, le coefficient de détermination, pour les deux capteurs, est meilleur pour les matchups de l'océan du large ou hauturier (0.748 pour SeaWiFS et 0.829 pour MODIS/Aqua) que pour ceux de l'océan côtier (0.609 pour SeaWiFS et 0.586 pour MODIS/Aqua). Il en va de même pour l'erreur de mesure, qui est environ deux fois plus faible dans l'océan ouvert (88% pour SeaWiFS et 70% pour MODIS/Aqua) que dans le côtier (161% pour les deux capteurs). Par contre, les données issues du capteur SeaWiFS sont plus biaisées en océan ouvert (35%) qu'en côtier (7%) tandis que les données issues du capteur MODIS/Aqua sont entachées d'un biais identique de 24%.

Si on s'intéresse maintenant aux trois biomes composant l'océan ouvert (polaire, des coups de vent d'ouest et des alizés), on peut constater que les données satellitaires sont globalement meilleures dans le biome des coups de vent d'ouest, puis dans le biome polaire et enfin dans le biome des alizés. En effet, si on regarde tout d'abord le coefficient de détermination Figure 2.7, celui-ci est à peu près équivalent dans le biome polaire et dans le biome des coups de vents d'ouest, mais moins bon dans le biome des alizés. Par contre, l'erreur de mesure sur les données des deux capteurs est presque deux fois plus élevée dans le biome polaire (163% pour SeaWiFS et 155% pour MODIS/Aqua) que dans les biomes des coups de vent d'ouest (70% pour les deux capteurs) et des alizés (96% pour SeaWiFS et 60% pour MODIS/Aqua). Quant au biais, il est plus faible dans le biome des coups de vent d'ouest (16% pour SeaWiFS et 20% pour MODIS/Aqua), puis dans le biome polaire (17% pour SeaWiFS et 26% pour MODIS/Aqua) et, enfin, dans le biome des alizés (63% pour SeaWiFS et 29% pour MODIS/Aqua).

3.3.3 - Comparaison des comportements des capteurs

Les données de MODIS/Aqua sont légèrement meilleures que celles de SeaWiFS pour les reprocessings considérés (cf. Chapitre 1). En effet, globalement, les mesures de MODIS/Aqua correspondent mieux aux données *in situ* que celles de SeaWiFS : même si le biais est légèrement moins élevé pour SeaWiFS (22%) que pour MODIS/Aqua (24% ; cf. Figure 2.8), le coefficient de détermination (cf. Figure 2.7) et l'erreur de mesure (cf. Figure 2.8) sont légèrement meilleurs pour MODIS/Aqua (0.84 et 106% resp.) que pour SeaWiFS (0.82 et 120% resp.).

D'autre part, on peut remarquer que les données de SeaWiFS sont légèrement meilleures que celles de MODIS/Aqua dans l'océan côtier. En effet, si l'erreur de mesure est identique pour les deux capteurs (161%), le biais est beaucoup moins élevé pour le premier (7%) que pour le second (25%).

Pour l'océan ouvert, les données de MODIS/Aqua (erreur de mesure : 70% ; biais : 24%) sont meilleures que celles de SeaWiFS (erreur de mesure : 88% ; biais : 35%). Or, si on regarde plus en détail les résultats par biome, on peut s'apercevoir que, pour les biomes polaire et des coups de vent d'ouest, les erreurs de mesure des données MODIS/Aqua (resp. 155% et 70%) sont légèrement inférieures et égales à celles des données issues de SeaWiFS (resp. 163% et 70%). Par contre, le biais des données de SeaWiFS (biome polaire : 17% et biome des coups de vent d'ouest : 16%) est moins élevé que pour les données de MODIS/Aqua (resp. 26% et 20%). La différence vient donc du comportement des capteurs dans le biome des alizés : en effet, l'erreur de mesure et le biais des données issues de SeaWiFS (resp. 96% et 63%) sont nettement plus élevés que pour MODIS/Aqua (resp. 60% et 29%). Cette différence peut tout d'abord provenir du fait qu'il n'y ait pas de matchups pour MODIS/Aqua dans les provinces #17 (Province des Caraïbes) et #30 (Province du Gyre de la Mousson Indienne). Mais elle vient plus probablement de la moins bonne performance du capteur SeaWiFS dans l'Océan Atlantique Equatorial et Sud (provinces #7, #8 et #10). Comme l'avaient noté Gregg et Casey [2004], l'Atlantique Tropical (dont la Province #8, dite Province de l'Atlantique Tropical Ouest), surtout, est remarquable par son erreur de mesure et son biais élevé pour les données SeaWiFS. En effet, le Désert du Sahara, qui est l'une des plus grandes sources d'aérosols absorbants, compromet la précision des corrections atmosphériques dans cette région de l'Atlantique. De plus, deux rivières tropicales majeures (l'Amazone et l'Orénoque) ont leur embouchure dans cette région, apportant d'importantes quantités de matières terrigènes. La conséquence de ces deux sources d'erreurs est la surestimation de la concentration en chlorophylle-a par le capteur SeaWiFS (nettement visible en Annexe A.1), car les longueurs d'onde relatives au bleu sont alors plus absorbées.

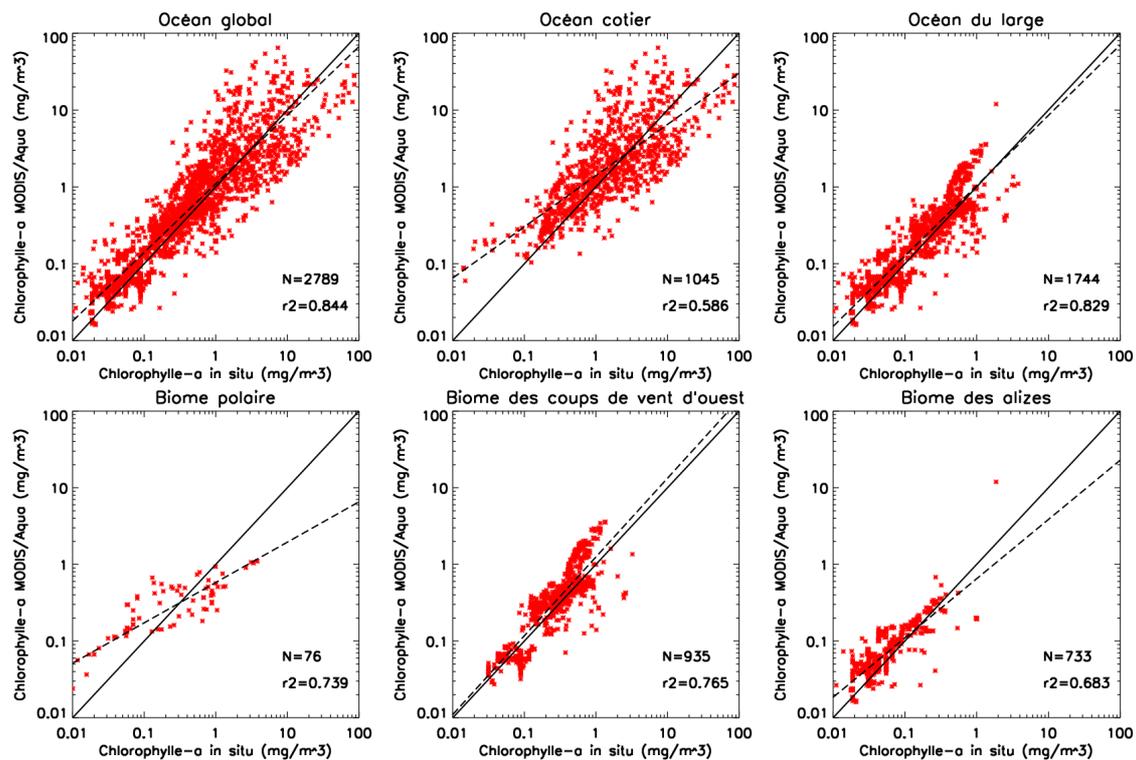
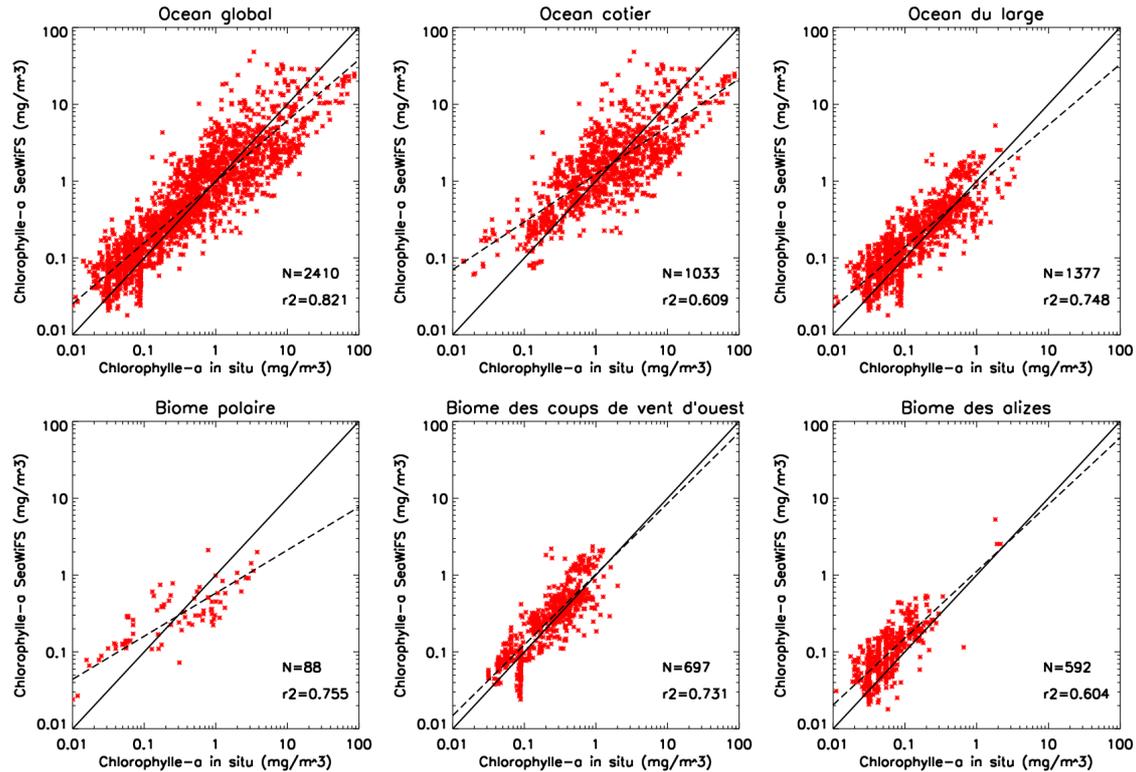


Figure 2.7 – Diagrammes des concentrations en chlorophylle-a SeaWiFS (en haut) et de MODIS/Aqua (en bas) en fonction des mesures *in situ* pour l'océan global, le côtier et le large (1^e ligne) et les biomes polaire, des coups de vent d'ouest et des alizés (2^e ligne). A noter que les biomes polaire, des coups de vent d'ouest et des alizés sont des sous-zones de l'océan du large (cf. partie 1). NB : les diagrammes par province sont disponibles en Annexe A.1.

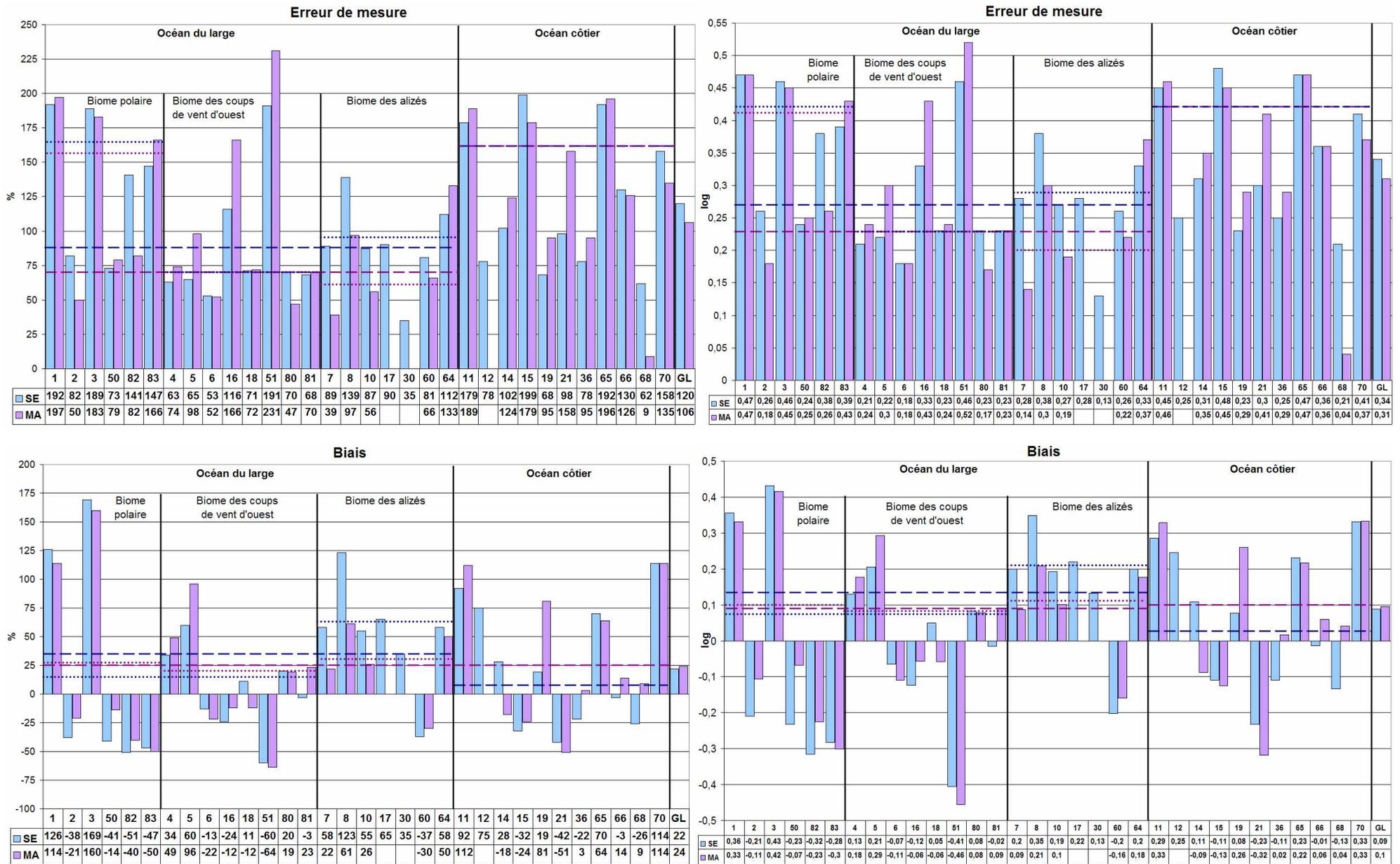


Figure 2.8 – Erreur de mesure (en haut) et biais (en bas) entre les données de concentration en chlorophylle-a de SeaWiFS (« SE » : bleu) et MODIS/Aqua (« MA » : violet) et les données *in situ*, en pourcentage de chlorophylle-a (à gauche) et en logarithme base 10 (à droite). Chaque barre d'historgramme correspond à une province biogéochimique, « GL » étant pour « Global ». Les lignes de traits tirés (bleu foncé pour SeaWiFS, violet foncé pour MODIS/Aqua) désignent les moyennes sur l'océan du large ou l'océan côtier ; les lignes pointillées désignent les moyennes par biome de l'océan du large (biome polaire, des coups de vent d'ouest et des alizés).

3.3.4 - Conclusion

Les données satellitaires de concentration en chlorophylle-*a*, pour les reprocessings utilisés, sont meilleures dans l'océan du large (ou hauturier) qu'en côtier. Cette différence est d'autant plus nette pour le capteur MODIS/Aqua ; c'est probablement pourquoi, il y a 3 ans environ, la NASA proposait trois produits différents pour MODIS/Aqua : un produit global, un produit pour l'océan du large et un produit pour le côtier. Les données sont ensuite plus fiables dans le biome des coups de vent d'ouest, puis dans les biomes polaire et des alizés.

Les données issues de MODIS/Aqua sont globalement moins entachées d'erreur que celles de SeaWiFS, mais leur biais par rapport aux observations *in situ* est, lui, plus important. La différence de qualité entre les deux jeux de données est principalement notable dans l'Atlantique Tropical, où SeaWiFS surestime nettement la concentration en chlorophylle-*a*.

3.4 - Construction des cartes d'erreur et de biais

Pour nos méthodes de combinaison de données, une carte des erreurs de mesure et / ou une carte des biais entre les données satellitaires et les mesures *in situ* est nécessaire. Les valeurs fournies Figure 2.8 sont donc utilisées. Pour les provinces biogéochimiques pour lesquelles il n'y avait pas de matchup, les valeurs d'erreur RMS et de biais appliquées sont celles de la province la plus proche. « Province la plus proche » signifie la plus proche géographiquement et biogéochimiquement. Ces associations de provinces, montrées dans le Tableau 2.3, sont donc réalisées suivant le biome (polaire, des coups de vent d'ouest, des alizés ou côtier ; cf. partie 1) auquel elles appartiennent.

Tableau 2.3 – Associations choisies entre les provinces biogéochimiques pour lesquelles il y a des matchups (indiquées par « X ») et celles pour lesquelles il n'y en a aucun.

Province #	Biome	Association avec la province #... ou le biome ...
1	Polaire	X
2	Polaire	X
3	Polaire	X
4	Coups de vent d'Ouest	X
5	Coups de vent d'Ouest	X
6	Coups de vent d'Ouest	X
7	Alizés	X
8	Alizés	X
9	Alizés	8
10	Alizés	X
11	Côtier	X
12	Côtier	X (SeaWiFS) – 11 (MODIS/Aqua)
13	Côtier	12 (SeaWiFS) – 11 (MODIS/Aqua)
14	Côtier	X
15	Côtier	X
16	Coups de vent d'Ouest	X

17	Alizés	X (SeaWiFS) – 7 (MODIS/Aqua)
18	Coups de vent d'Ouest	X
19	Côtier	X
20	Côtier	Moyenne (14,21)
21	Côtier	X
22	Côtier	12 (SeaWiFS) – 11 (MODIS/Aqua)
30	Alizés	X (SeaWiFS) – Biome alizés (MODIS/Aqua)
31	Alizés	30 (SeaWiFS) – Biome alizés (MODIS/Aqua)
32	Côtier	36
33	Côtier	36
34	Côtier	36
35	Côtier	36
36	Côtier	X
37	Côtier	70
50	Polaire	X
51	Coups de vent d'Ouest	X
52	Coups de vent d'Ouest	51
53	Coups de vent d'Ouest	Biome des coups de vent d'ouest
54	Coups de vent d'Ouest	Biome des coups de vent d'ouest
56	Coups de vent d'Ouest	Biome des coups de vent d'ouest
58	Coups de vent d'Ouest	Biome des coups de vent d'ouest
59	Coups de vent d'Ouest	Biome des coups de vent d'ouest
60	Alizés	X
61	Alizés	Biome des alizés
62	Alizés	Biome des alizés
63	Alizés	Biome des alizés
64	Alizés	X
65	Côtier	X
66	Côtier	X
67	Côtier	Moyenne (66,68)
68	Côtier	X
69	Côtier	Biome côtier
70	Côtier	X
71	Côtier	70
72	Côtier	70
80	Coups de vent d'Ouest	X
81	Coups de vent d'Ouest	X
82	Polaire	X
83	Polaire	X

Pour éviter des gradients trop brutaux aux frontières des provinces biogéochimiques dans les champs d'erreur de mesure et de biais, ceux-ci ont été lissés à 100 km par un filtrage médian. Les cartes utilisées par la suite sont montrées Figure 2.9 et Figure 2.10.

Le lecteur aura pu constater que, si l'Océan Atlantique (notamment l'Océan Atlantique Nord) est bien desservi par les matchups, ce n'est guère le cas pour le reste du monde, en particulier l'Océan Pacifique Sud et l'Océan Indien. Nous avons estimé aussi bien que possible les cartes d'erreur de mesure et de biais pour chacun des capteurs, étant donné le nombre de mesures *in situ* que nous avons en notre possession. Ces estimations donnent un ordre de grandeur suffisant pour les algorithmes utilisés.

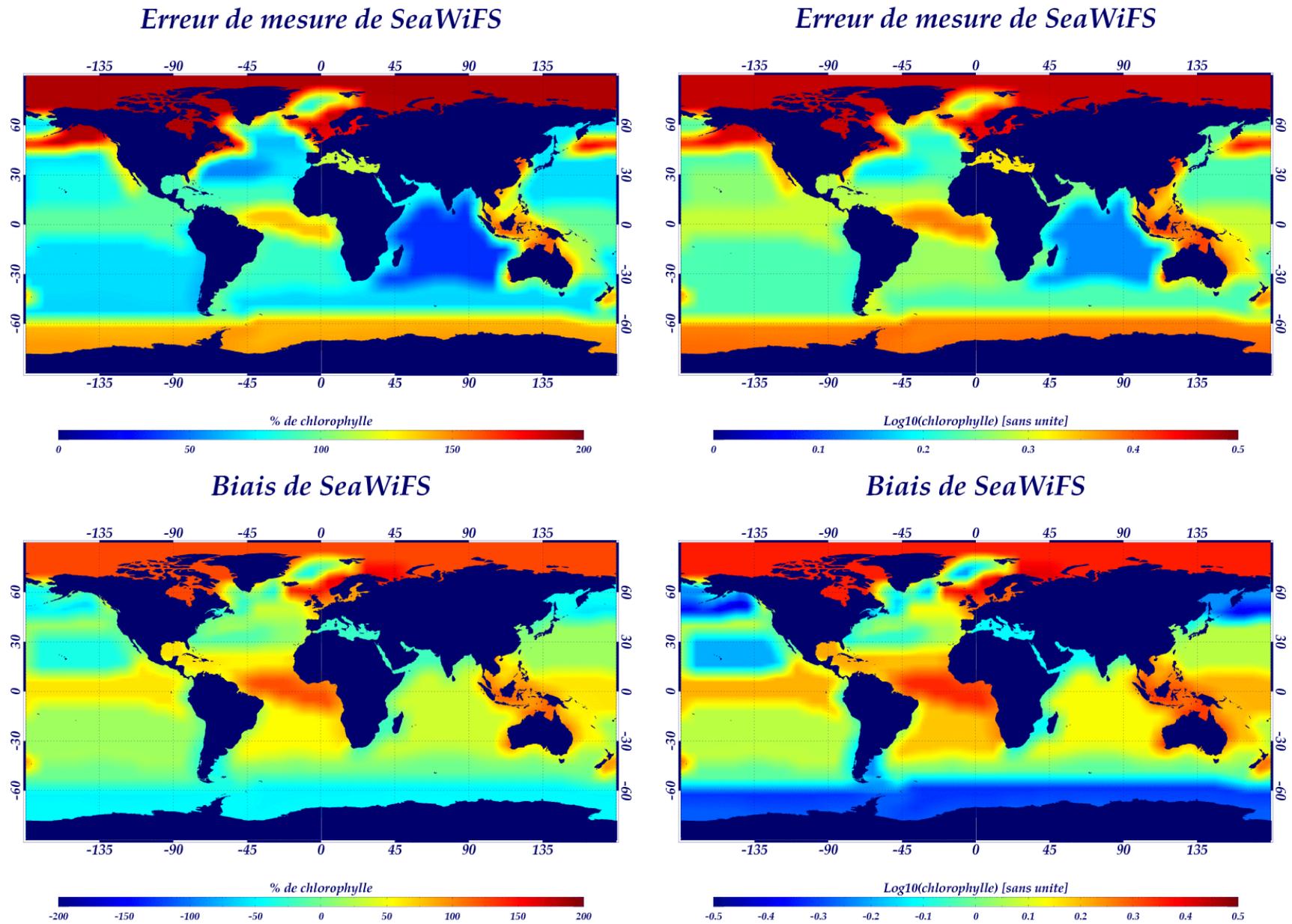


Figure 2.9 – Cartes des erreurs de mesure (en haut) et des biais (en bas) de SeaWiFS, en pourcentage de chlorophylle-*a* (à gauche) et en logarithme base 10 (à droite).

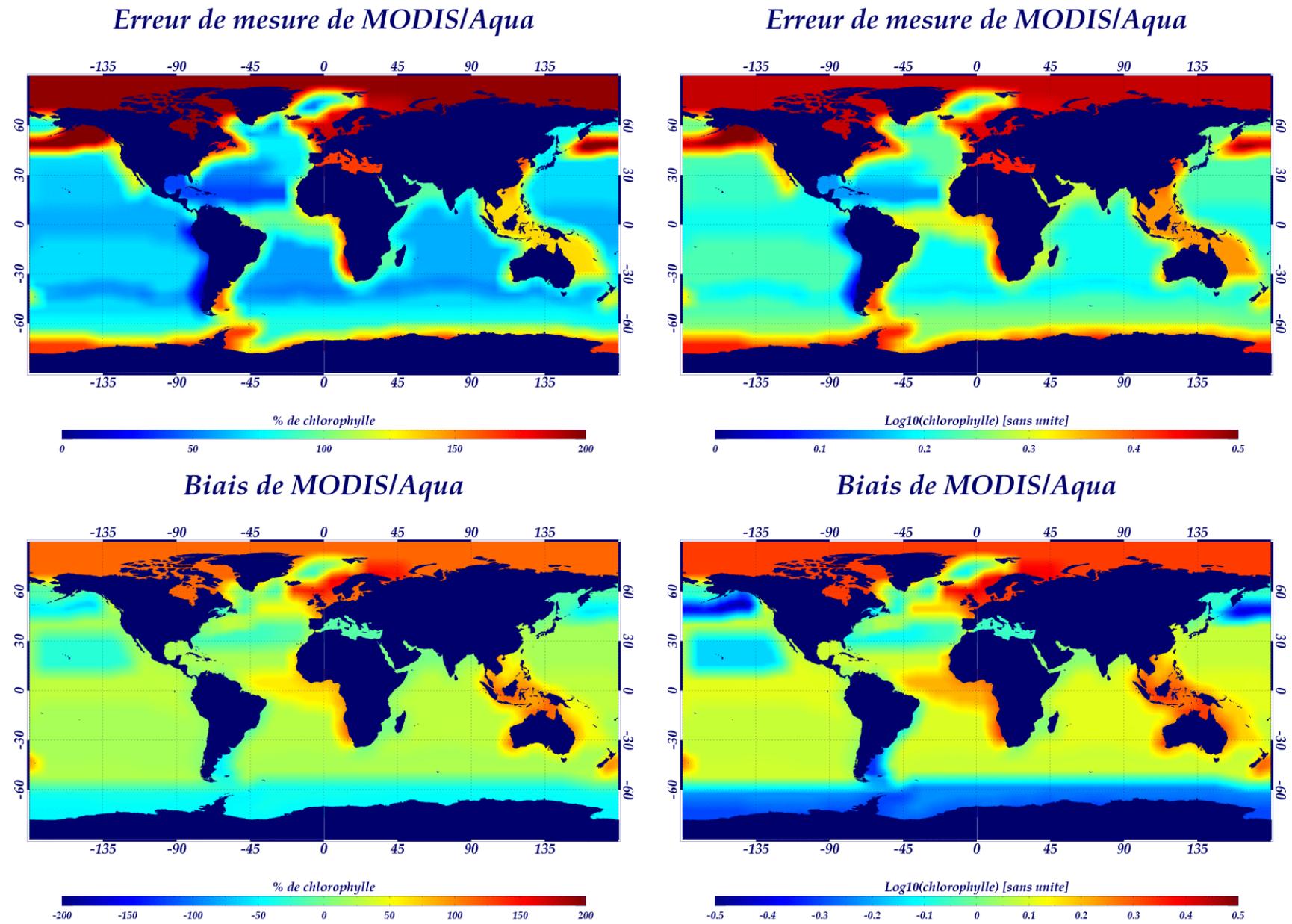


Figure 2.10 – Cartes des erreurs de mesure (en haut) et des biais (en bas) de MODIS/Aqua, en pourcentage de chlorophylle-a (à gauche) et en logarithme base 10 (à droite).

3.5 - Conclusion

Dans cette partie, une méthode d'étude qualitative des données satellitaires a tout d'abord été présentée : il s'agit de comparaisons statistiques par rapport aux observations *in situ*, appelées « matchups ». Nous avons vu, qu'étant donnée la distribution lognormale des données de concentration en chlorophylle-a, ces statistiques doivent être déterminées sur le logarithme base 10 des données, et non sur les valeurs brutes, comme à l'accoutumée. Enfin, les matchups sur les données SeaWiFS et MODIS/Aqua nous ont appris que, pour les reprocessings utilisés, les données de ce dernier capteur étaient globalement légèrement meilleures que celles du premier, tant au niveau de l'erreur de mesure qu'au niveau du biais.

Maintenant que nous avons comparé les données satellitaires aux données *in situ*, nous allons comparer les données satellitaires entre elles.

4 - Intercomparaison des données satellitaires

4.1 - Introduction

L'objectif de cette partie est de quantifier le biais existant entre les données de concentration en chlorophylle-a de SeaWiFS et celles de MODIS/Aqua. Nous allons tout d'abord formuler un tel calcul. Puis, nous étudierons les résultats obtenus.

4.2 - Méthodologie

Nous souhaitons quantifier le biais journalier moyen de SeaWiFS par rapport à MODIS/Aqua, qui est le moins entaché d'erreur de mesure et de biais. La formulation d'un tel calcul est proche de celle utilisée pour le calcul du biais entre les données satellitaires et les mesures *in situ* :

$$\text{BIAIS}_{\text{SE-MA}}(\text{long, lat}) = \frac{1}{T} \sum_{t=1}^T (\log_{10}(\text{SE}_t(\text{long, lat})) - \log_{10}(\text{MA}_t(\text{long, lat})))$$

où (long, lat) désigne les coordonnées du point auquel on s'intéresse, $\text{SE}_t(\text{long, lat})$ (resp. $\text{MA}_t(\text{long, lat})$) est la donnée de concentration en chlorophylle-a SeaWiFS (resp. MODIS/Aqua) aux coordonnées (long, lat) pour le jour t et T est le nombre de couples données SeaWiFS / données MODIS/Aqua que l'on a pour la durée souhaitée.

Là encore, ce biais peut s'exprimer en pourcentage de valeur de chlorophylle-a, selon la relation :

$$\text{BIAIS}_{\text{SE-MA}/\log_{10}} = \log_{10}(1 + \text{BIAIS}_{\text{SE-MA}/\%}) \text{ soit } \text{BIAIS}_{\text{SE-MA}/\%} = 10^{\text{BIAIS}_{\text{SE-MA}/\log_{10}}} - 1$$

$$\text{soit } \boxed{\text{BIAIS}_{\text{SE-MA}/\%}(\text{long, lat}) = 10^{\frac{1}{T} \sum_{t=1}^T (\log_{10}(\text{SE}_t(\text{long, lat})) - \log_{10}(\text{MA}_t(\text{long, lat})))} - 1}$$

4.3 - Résultats et discussion

La carte obtenue pour l'année 2003 à l'aide de la formule précédente est celle montrée Figure 2.11. On peut observer que le biais moyen maximal est de l'ordre de 20%. De plus, la concentration en chlorophylle-a de SeaWiFS est globalement plus élevée que celle de MODIS/Aqua, spécialement dans l'Atlantique Equatorial, l'océan Pacifique Sud et l'océan Indien. Par contre, c'est l'inverse dans la partie ouest de l'océan Pacifique Nord ainsi que dans le gyre subtropical pacifique sud. On peut noter un découpage ressemblant étrangement à celui des provinces biogéochimiques (Figure 2.1), laissant penser que la concentration en chlorophylle-a donnée par SeaWiFS est plus élevée que celle fournie par MODIS/Aqua pour les régions de forte concentration en chlorophylle, et moins élevée pour les régions de plus faible concentration en chlorophylle, comme le confirment les matchups montrés Figure 2.8.

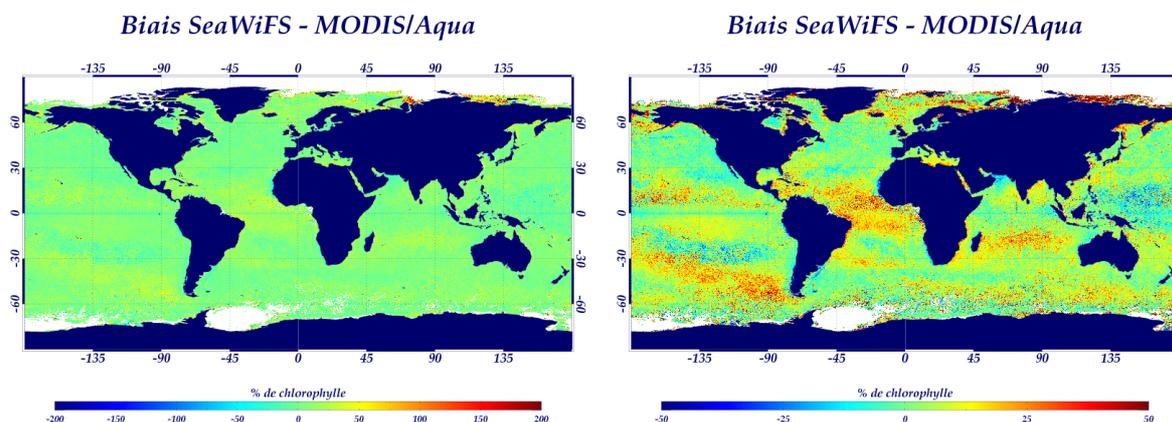


Figure 2.11 – Carte des biais journaliers moyens de SeaWiFS par rapport à MODIS/Aqua pour l'année 2003, avec l'échelle des Figures 2.9 et 2.10 (à gauche) et une échelle plus ajustée (à droite).

La Figure 2.12 montre les variations mensuelles du biais de SeaWiFS par rapport à celui de MODIS/Aqua. On peut constater que le biais global est plus important en mai et juillet (9%) et en décembre (6.5%), et moins important en janvier-mars (4.5%) et en octobre (2.5%). Cette tendance se fait aussi remarquer pour les 4 biomes, de façon moins nette pour

le biome des alizés. La figure de gauche montre que le biais dans l'océan du large est plus important que dans l'océan côtier, d'environ 3 à 4% pour toute l'année. La figure de droite montre que le biais le plus important se fait ressentir dans le biome polaire en mai.

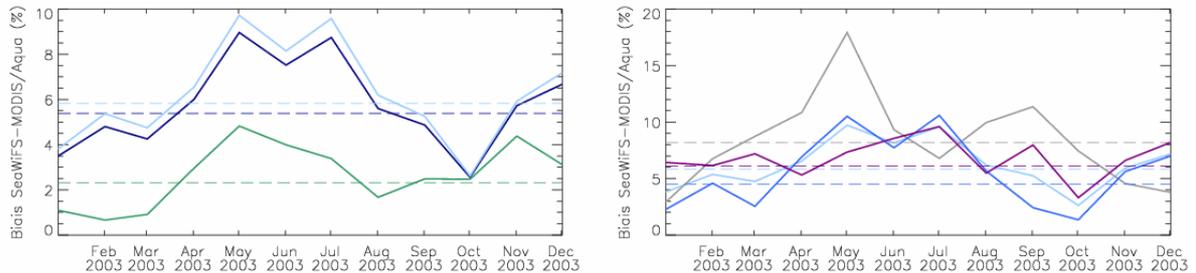


Figure 2.12 – A gauche : variations mensuelles du biais de SeaWiFS par rapport à MODIS/Aqua pour l'océan global (bleu foncé), l'océan côtier (bleu-vert) et l'océan du large (bleu clair). A droite : même chose pour l'océan du large (bleu clair), le biome polaire (gris), le biome des coups de vent d'ouest (bleu foncé) et le biome des alizés (violet).

4.4 - Conclusion

Dans cette partie, nous avons évalué le biais journalier moyen des données de concentration en chlorophylle-a de SeaWiFS par rapport à celles de MODIS/Aqua. Nous avons vu qu'il était lié aux provinces biogéochimiques : il semblerait que SeaWiFS surestime plus que MODIS/Aqua la chlorophylle là où elle est élevée et la sous-estime plus aussi là où elle est la plus faible. La variabilité temporelle de ce biais montre qu'il est plus important en mai-juillet et décembre, qu'en janvier-mars ou en octobre.

5 - Conclusion du chapitre

Ce chapitre nous a permis de qualifier et quantifier la concentration en chlorophylle-a à l'échelle de l'océan global et par province biogéochimique.

Nous avons tout d'abord étudié la distribution particulière de la concentration en chlorophylle-a, qui est lognormale, i.e. c'est le logarithme de la concentration en chlorophylle qui a une distribution gaussienne.

Ensuite, le concept de matchups, ou comparaison des données satellitaires avec les données *in situ*, a été introduit. Nous avons vu que la distribution particulière des données nous oblige à calculer les statistiques à partir du logarithme base 10 des données. Les matchups montrent que, pour les reprocessings étudiés, les données issues du capteur MODIS/Aqua sont globalement légèrement meilleures que celles issues de SeaWiFS. En effet, SeaWiFS semble surestimer plus que MODIS/Aqua les fortes concentrations en

chlorophylle-*a* et sous-estimer plus que MODIS/Aqua les faibles concentrations en chlorophylle-*a*.

Dans le chapitre suivant, nous allons nous pencher sur le cœur de notre problème : la combinaison des cartes journalières de concentration en chlorophylle-*a*. Les deux méthodes comparées sont la moyenne pondérée par l'erreur capteur et l'analyse objective. L'application est montrée pour la combinaison des produits SeaWiFS et MODIS/Aqua.

Chapitre 3

Combinaison de données de
chlorophylle : Moyenne Pondérée
et Analyse Objective

CHAPITRE 3

COMBINAISON DES DONNEES DE CHLOROPHYLLE : MOYENNE PONDEREE ET ANALYSE OBJECTIVE

Dans ce chapitre, deux approches pour la combinaison de données de concentration en chlorophylle sont testées et comparées : la moyenne pondérée par l'erreur capteur et l'analyse objective. Cette étude a été réalisée, dans un premier temps, pour la combinaison des données de concentration en chlorophylle issues de SeaWiFS (reprocessing #4.5) et MODIS/Aqua (reprocessing #1) dans l'océan Atlantique Nord et Equatorial, donnant lieu à une publication dans le journal *IEEE Transactions on Geoscience and Remote Sensing* [Pottier *et al.*, 2006 ; cf. Annexe B]. Dans ce chapitre, nous allons montrer l'application globale de ces deux méthodes de combinaison de données pour les derniers reprocessings de SeaWiFS (#5.1) et MODIS/Aqua (#1.1).

Dans une première partie, nous nous penchons sur la moyenne pondérée. Après avoir explicité les deux approches envisagées, les résultats de l'application de cette méthode sont montrés pour deux journées tests. Dans une seconde partie, l'analyse objective est présentée. Après avoir brièvement rappelé son principe, les paramètres d'entrée nécessaires sont exposés. La sensibilité de l'analyse objective à ceux-ci est ensuite vérifiée avant de montrer les résultats de l'application de cette méthode aux deux mêmes journées tests. Dans une dernière partie, les deux méthodes sont comparées : tout d'abord en comparant les données combinées avec des mesures *in situ*, puis en étudiant les tendances temporelles et latitudinales.

Les deux journées de test ont été choisies à deux saisons différentes, les 26 mars et 13 août 2003, montrés sur la Figure 3.1. Le 26 mars 2003, la couverture spatiale de SeaWiFS est de 16% de l'océan global, tandis que celle de MODIS/Aqua est de 10.2% (Figure 3.2). Le 13 août 2003, la couverture spatiale de SeaWiFS (15.8%) est aussi plus importante que celle de MODIS/Aqua (9.4%) (Figure 3.2).

Sommaire

1 – MOYENNE PONDEREE.....	98
2 – ANALYSE OBJECTIVE	107
3 – COMPARAISON ENTRE LES DEUX METHODES.....	133
4 – CONCLUSION DU CHAPITRE.....	139

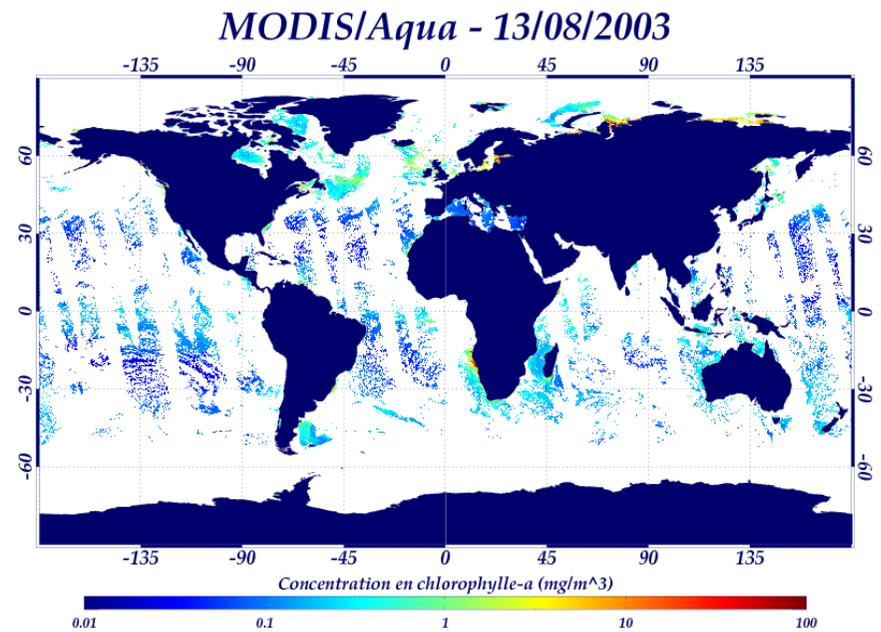
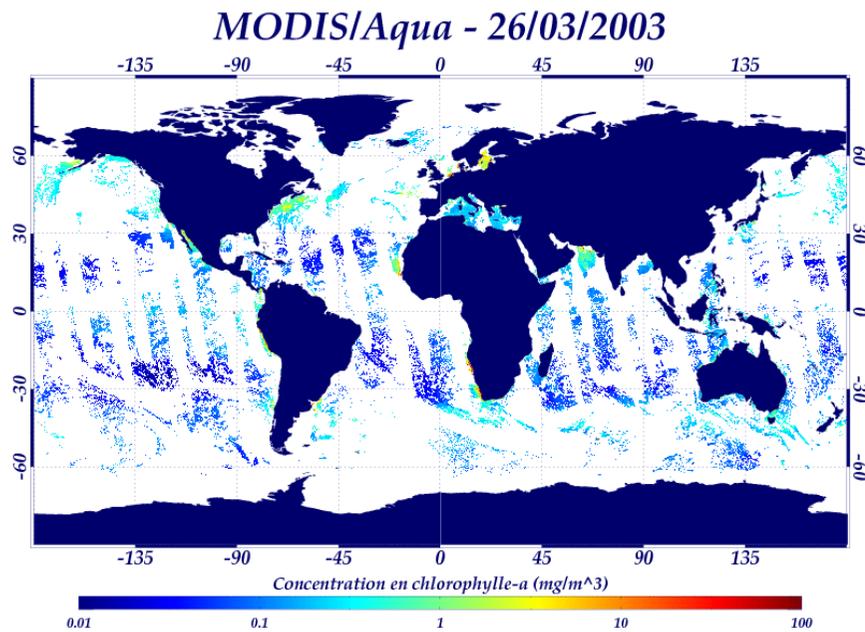
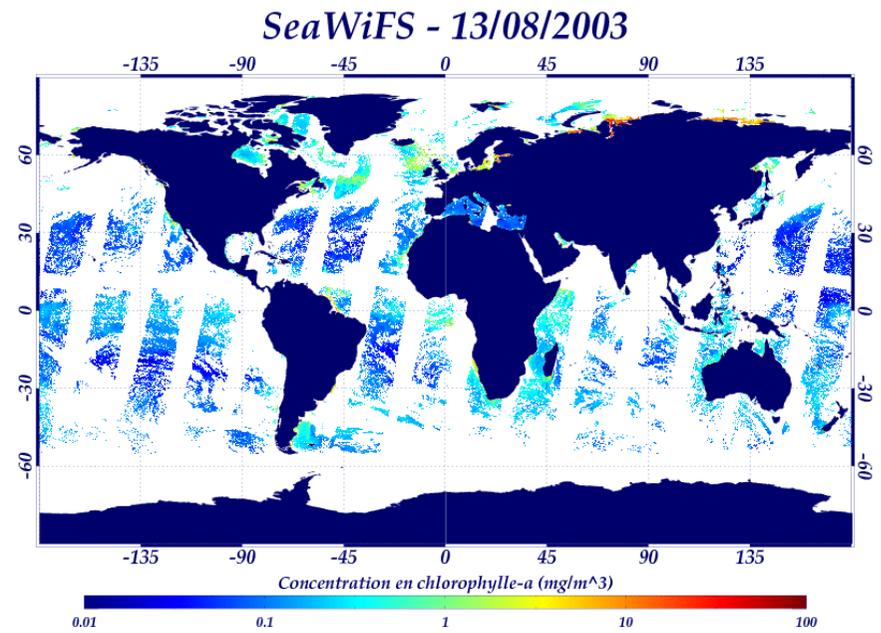
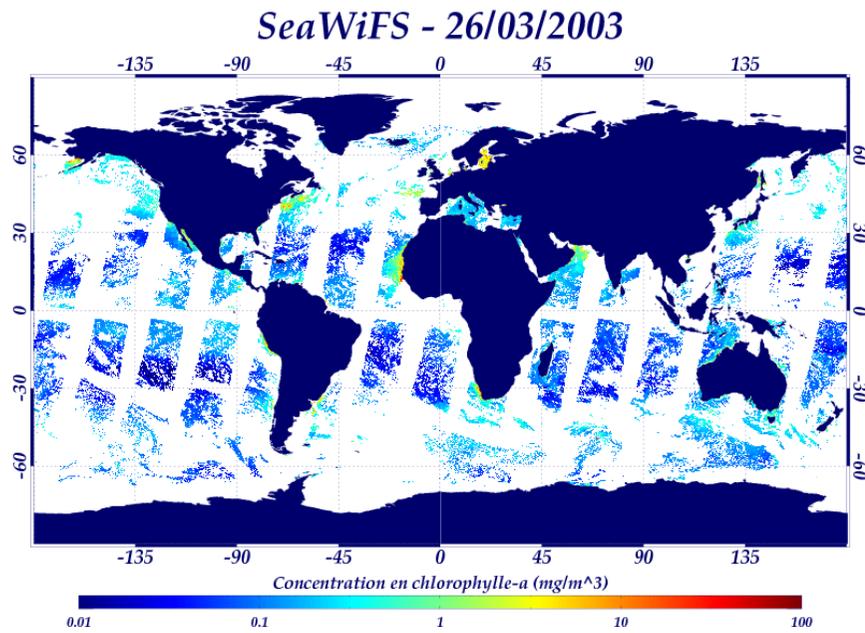
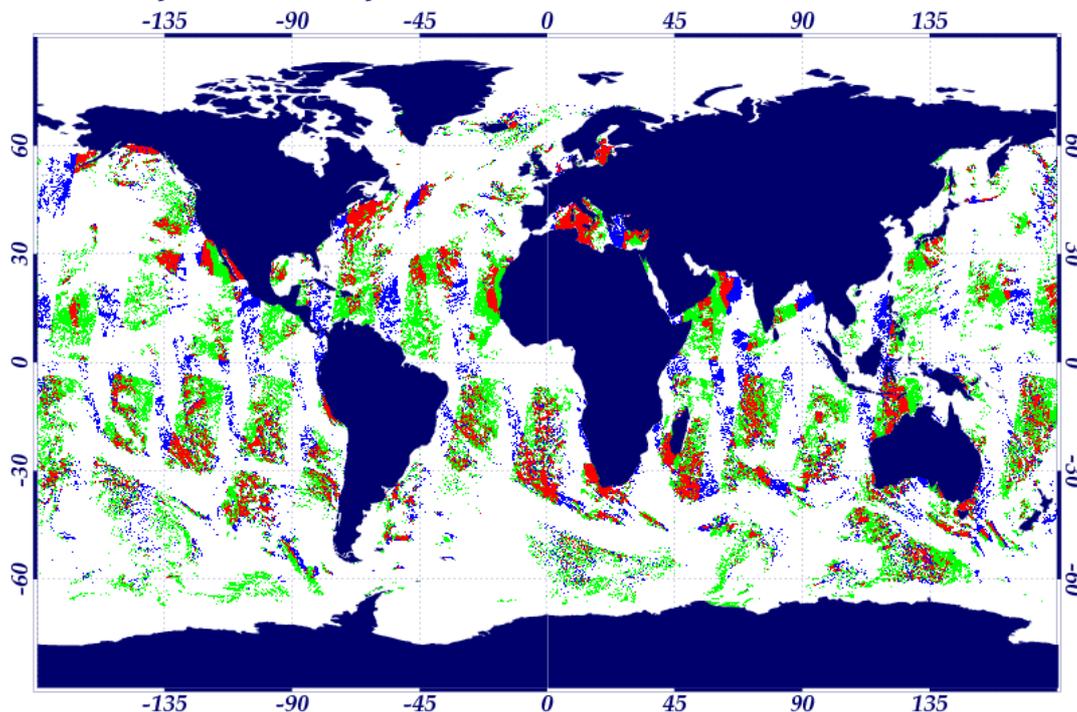


Figure 3.1 – Carte des concentrations en chlorophylle-a issues du capteur SeaWiFS (en haut) et MODIS/Aqua (en bas) pour le 26 mars 2003 (à gauche) et le 13 août 2003 (à droite).

Repartition spatiale des donnees - 26/03/2003



Repartition spatiale des donnees - 13/08/2003

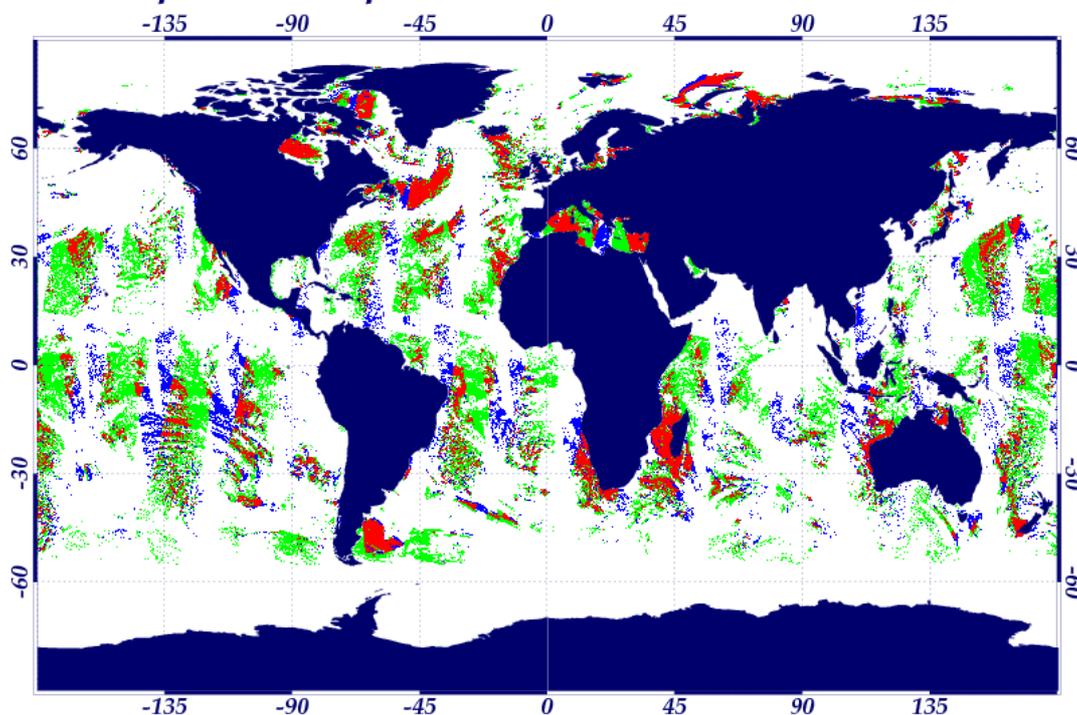


Figure 3.2 – Carte de la répartition spatiale des données pour le 26 mars 2003 (en haut) et le 13 août 2003 (en bas). La couleur blanche signifie qu'il n'y a pas de données, le vert qu'il s'agit de données issues de SeaWiFS uniquement, le bleu de MODIS/Aqua uniquement et le rouge qu'il y a une donnée pour les deux capteurs.

1 - Moyenne pondérée

1.1 - Introduction

La première méthode développée est la moyenne pondérée par l'erreur capteur. L'algorithme a été développé de façon à obtenir des données combinées à n'importe quelle résolution souhaitée, avec autant de capteurs que possible à n'importe quelle résolution. Dans une première partie, le principe de la méthode dans le cas particulier de la combinaison de SeaWiFS et MODIS/Aqua est présenté, ainsi que les deux approches envisageables. Puisque les données de concentration en chlorophylle sont lognormalement distribuées [Campbell, 1995], les deux algorithmes mis en place sont formulés pour les applications sur les données brutes et sur le logarithme des valeurs. Puis les résultats de ces deux algorithmes appliqués aux données journalières de concentration en chlorophylle ou à leur logarithme sont montrés.

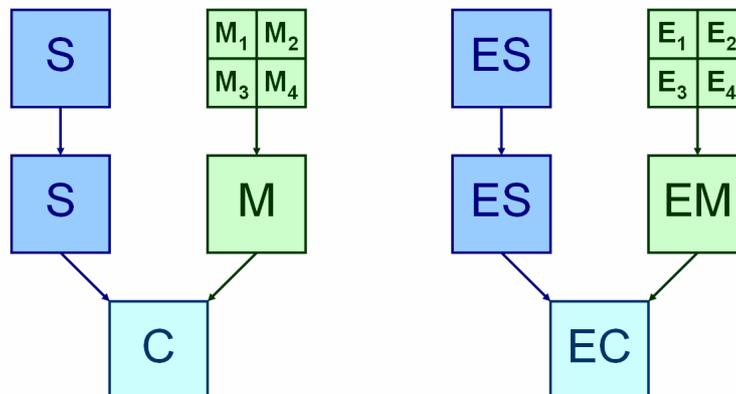
1.2 - Méthodologie

1.2.1 - Introduction

A partir de la carte de concentration en chlorophylle relative à un jour donné et de la carte des erreurs de mesure pour chaque capteur (i.e. l'erreur de mesure est l'erreur RMS calculée entre les données de chaque capteur et les mesures *in situ*, cf. Chapitre 2), on obtient la carte des données combinées de chlorophylle pour ce jour et la carte des erreurs associées. La valeur d'un pixel combiné est égale à la moyenne pondérée des pixels issus des différents capteurs mis en jeu et localisés au même endroit. Les poids sont la confiance que l'on accorde à chaque capteur par rapport aux autres, calculée à partir des erreurs RMS.

Dans cette partie, on s'intéresse à la combinaison des données SeaWiFS et MODIS/Aqua, dont les résolutions sont de $1/12^\circ$ et $1/24^\circ$ respectivement (la projection utilisée ici a été calculée par le second algorithme montré au Chapitre 1-2.2.5). Deux choix sont alors possibles pour combiner les données : la résolution des données combinées peut être celle du produit SeaWiFS ($1/12^\circ$: Approche MP1) ou celle du produit MODIS/Aqua ($1/24^\circ$: Approche MP2).

1.2.2 - Approche MP1



Légende:

- S : pixel SeaWiFS de concentration en chlorophylle
- ES : erreur associée au pixel SeaWiFS S
- M_i : pixel MODIS/Aqua de concentration en chlorophylle
- E_i : erreur associée au pixel MODIS/Aqua M_i
- M : pixel MODIS/Aqua de résolution 1/12°
- EM : erreur associée au pixel MODIS/Aqua M
- C : pixel combiné
- EC : erreur associée au pixel combiné C

Figure 3.3 – Description de la 1^{ère} approche pour la moyenne pondérée (Approche MP1).

Commençons par étudier le cas de valeurs brutes. La 1^{ère} étape de l'Approche MP1 (cf. Figure 3.3) consiste à sous-échantillonner les pixels du produit chlorophylle issu de MODIS/Aqua pour les amener à la résolution de ceux du produit SeaWiFS (1/12°). Ainsi, un pixel candidat de MODIS/Aqua M est obtenu à partir de 4 pixels de chlorophylle originaux à l'aide de la formule suivante :

$$M = \frac{1}{4}(M_1 + M_2 + M_3 + M_4) = \sum_{i=1}^4 \lambda_i M_i \text{ avec } \sum_{i=1}^4 \lambda_i = 1.$$

Les coefficients λ_i représentent la fraction d'aire du pixel candidat MODIS/Aqua que couvre chaque pixel de chlorophylle MODIS/Aqua original. Bien entendu, seuls les pixels existants sont considérés.

Pour effectuer le calcul de l'erreur EM associée au pixel candidat MODIS/Aqua M, considérons la distribution de moyenne M et d'écart-type EM : mathématiquement, il s'agit du résultat de la combinaison linéaire des 4 distributions de moyenne M_i et d'écart-type EM_i (i.e. l'erreur associée au pixel de chlorophylle M_i , qui est l'erreur RMS en mg/m^3 associée au point d'étude – c'est-à-dire le produit de M_i et de l'erreur RMS en pourcentage – et calculée à l'aide des matchups au Chapitre 2). Puisqu'il n'existe aucune règle générale pour calculer

l'écart-type d'une combinaison linéaire de variables lognormalement distribuées, l'erreur (c'est-à-dire l'écart-type) EM de M doit être calculée de manière basique :

$$EM = \sqrt{\sum_{n=1}^N X_n^2} \text{ avec } X_n = \sum_{i=1}^4 \lambda_i Y_{i,n} .$$

Dans un premier temps, on construit 4 distributions lognormales $Y_i, i = \{1,2,3,4\}$ de N points (après quelques tests, nous avons choisi $N=1000$: bon rapport entre la précision du résultat et le temps de calcul), de moyenne M_i et d'écart-type EM_i . Puis on calcule le résultat de leur combinaison linéaire X_n (les poids sont les coefficients λ_i). Enfin, l'erreur EM associée au pixel M est l'écart-type de la distribution ainsi calculée, X_n . L'absence de certains des 4 pixels MODIS/Aqua originaux amène si peu de différence que nous avons considéré que cela n'avait aucun impact réel sur la valeur de EM.

La seconde et dernière étape de l'Approche MP1 est la combinaison des candidats de SeaWiFS et MODIS/Aqua. Le pixel combiné C est calculé comme suit :

$$C = \left(1 - \frac{\%ES}{\%EM + \%ES}\right) S + \left(1 - \frac{\%EM}{\%EM + \%ES}\right) M$$

où $\%EX = \frac{EX}{X}$: les poids des pixels candidats S et M représentent le pourcentage de confiance que l'on a dans le pixel de chlorophylle correspondant en comparaison à l'autre. L'erreur associée EC est, là aussi, considérée comme étant l'écart-type de la combinaison linéaire des 2 distributions lognormales de moyenne S et M et d'écart-type ES et EM, les poids étant $1 - \frac{\%ES}{\%EM + \%ES}$ et $1 - \frac{\%EM}{\%EM + \%ES}$ respectivement.

Intéressons-nous maintenant à l'application de l'Approche MP1 au logarithme des concentrations en chlorophylle. La formulation du sous-échantillonnage de MODIS/Aqua, 1^e étape de l'algorithme, s'écrit maintenant :

$$M = 10^{\sum_{i=1}^4 \lambda_i \log M_i} .$$

Cette fois-ci, la distribution considérée (i.e. le logarithme des concentrations en chlorophylle) est normale. L'erreur associée au pixel candidat MODIS/Aqua se calcule donc simplement, à l'aide de la relation suivante :

$$EM = \sqrt{\sum_{i=1}^4 (\lambda_i EM_i)^2} .$$

Notons, au passage, que dans ce cas, EM_i est l'erreur associée à la valeur $\log M_i$: EM_i est l'erreur RMS « log » calculée au Chapitre 2.

Le calcul du pixel de chlorophylle combiné est formulé comme suit :

$$C = 10^{\left(1 - \frac{\%ES}{\%EM + \%ES}\right) \log S + \left(1 - \frac{\%EM}{\%EM + \%ES}\right) \log M}$$

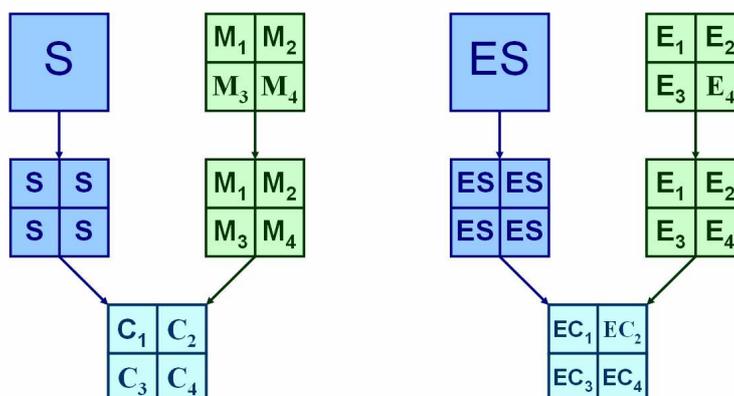
où $\%EX = \frac{EX}{\log X}$. L'erreur associée EC est, là aussi, considérée comme étant l'écart-type de

la combinaison linéaire des 2 distributions normales et se calcule donc simplement :

$$EC = \sqrt{\left(\left(1 - \frac{\%ES}{\%EM + \%ES}\right) \log S\right)^2 + \left(\left(1 - \frac{\%EM}{\%EM + \%ES}\right) \log M\right)^2}$$

1.2.3 - Approche MP2

La 1^{ère} étape de l'Approche MP2 (cf. Figure 3.4) consiste à amener les pixels de chlorophylle SeaWiFS à la résolution de ceux de MODIS/Aqua. Chaque pixel SeaWiFS est donc divisé en 4 parties d'aire identique, l'erreur associée étant égale à l'erreur du pixel original. La combinaison des pixels candidats ainsi obtenus se fait de la même manière que pour l'Approche MP1, que l'on considère les valeurs brutes ou leur logarithme.



Légende:

- S : pixel SeaWiFS de concentration en chlorophylle
- ES : erreur associée au pixel SeaWiFS S
- M_i : pixel MODIS/Aqua de concentration en chlorophylle
- E_i : erreur associée au pixel MODIS/Aqua M_i
- C_i : pixel combiné
- EC_i : erreur associée au pixel combiné C_i

Figure 3.4 – Description de la 2^{ème} approche pour la moyenne pondérée (Approche MP2).

1.2.4 - Implémentation de l'algorithme

L'organigramme du logiciel de combinaison de données par moyenne pondérée par l'erreur capteur est montré Figure 3.5.

Les paramètres d'initialisation sont les suivants :

- ☞ nombre de capteurs mis en jeu,
- ☞ caractéristiques des capteurs :
 - répertoire des fichiers d'entrée,
 - type de l'erreur de mesure : décimale (mg/m^3 ou log) ou en pourcentage, constante ou variable sur la zone,
 - valeur de l'erreur de mesure ou fichier des erreurs de mesure
- ☞ grille d'intérêt pour les données combinées (zone + échantillonnage spatial + dates),
- ☞ type des données traitées : chlorophylle ou logarithme de chlorophylle,
- ☞ nom général des fichiers de sortie

En sortie, on récupère, pour chaque jour étudié, la carte des données combinées et la carte des erreurs associées (en pourcentage de la donnée combinée).

1.2.5 - Conclusion

Dans cette partie, nous avons expliqué le principe de la méthode de combinaison par moyenne pondérée par l'erreur capteur. Dans le cas particulier de la combinaison des données de concentration en chlorophylle de SeaWiFS et MODIS/Aqua, nous avons vu que l'idée générale de la méthode se scinde en deux approches envisageables : soit les données combinées ont la résolution de SeaWiFS (i.e. $1/12^\circ$: Approche MP1), soit elles ont la résolution de MODIS/Aqua (i.e. $1/24^\circ$: Approche MP2). De plus, la méthode peut s'appliquer aux données brutes ou au logarithme des données.

Nous allons voir le résultat de l'application de ces 4 cas dans la partie suivante.

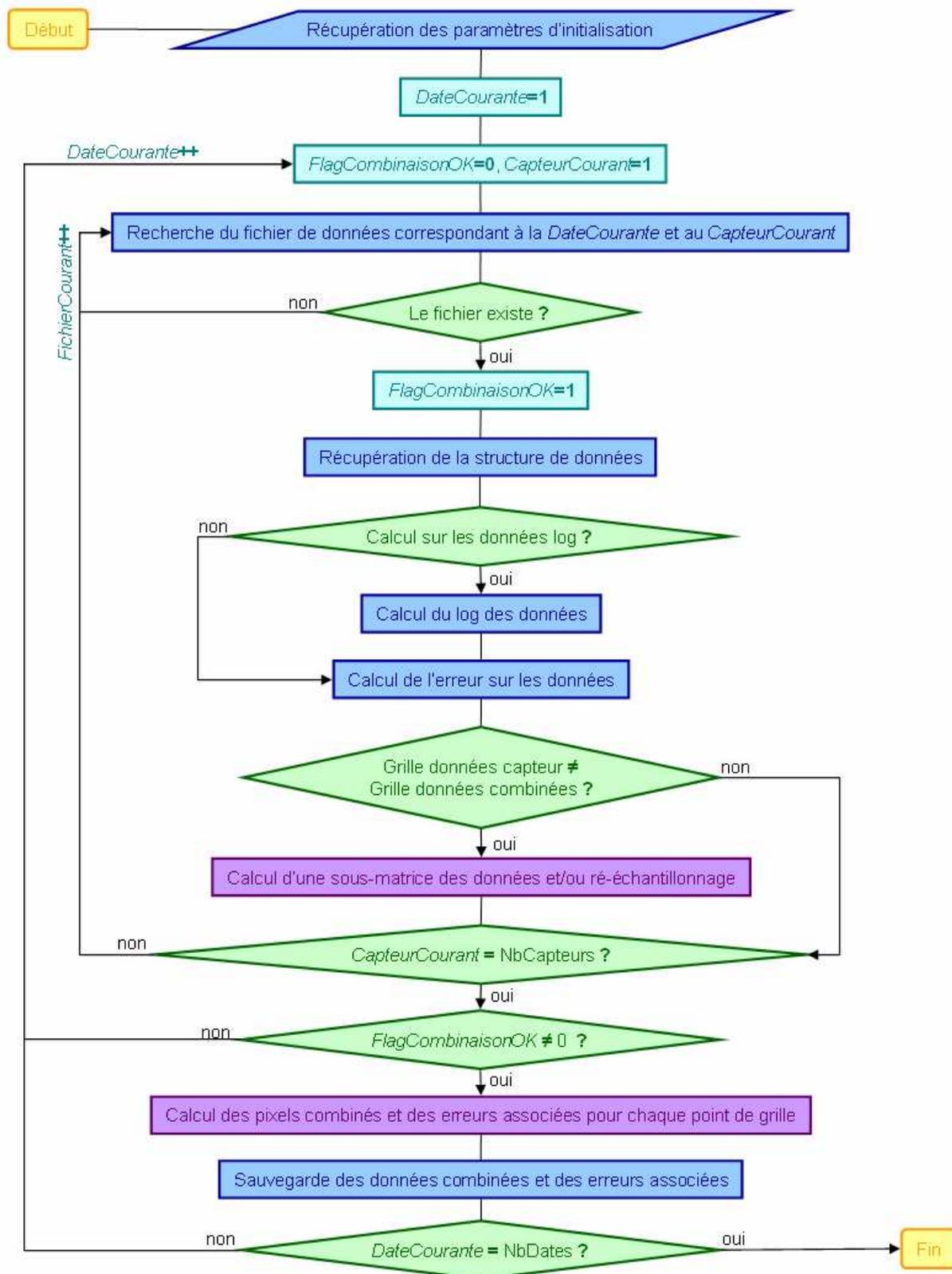


Figure 3.5 – Organigramme du logiciel de moyenne pondérée.

1.3 - Application aux données de chlorophylle

1.3.1 - Introduction

Nous venons de voir que, pour l'algorithme de moyenne pondérée par l'erreur capteur, 4 cas sont à considérer. En effet, les approches MP1 (la résolution des données combinées est celle du produit SeaWiFS, soit $1/12^\circ$; le suffixe « 12 » sera utilisé par la suite) et MP2 (la résolution des données combinées est celle du produit MODIS/Aqua, soit $1/24^\circ$; suffixe « 24 » ci-après) peuvent être appliquées aux valeurs brutes (préfixe « DEC » ci-après) ou au logarithme base 10 des données de concentration en chlorophylle (préfixe « LOG » ci-après).

Ces 4 cas ont été utilisés pour la combinaison des données globales de concentration en chlorophylle de SeaWiFS et MODIS/Aqua sur toute l'année 2003. Les résultats sont présentés dans cette partie.

1.3.2 - Les paramètres d'entrée

En entrée du logiciel, sont nécessaires la carte des données de concentration en chlorophylle pour le jour étudié pour chaque capteur (Figure 3.2) ainsi que les cartes des erreurs associées, en pourcentage de la valeur de concentration en chlorophylle pour chaque pixel (Figures 2.11 et 2.12).

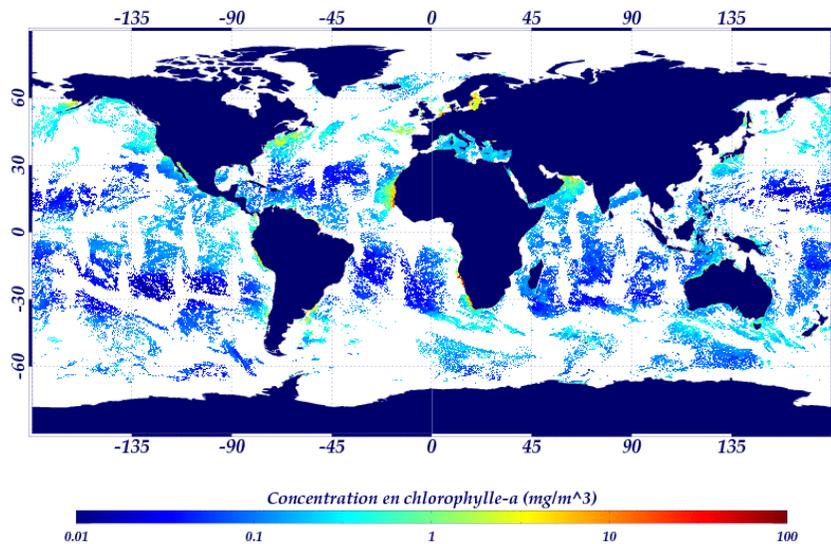
1.3.3 - Les résultats

Les résultats de l'application de l'algorithme MP1 (la résolution des données combinées est celle du produit SeaWiFS, soit $1/12^\circ$) au logarithme base 10 des données de concentration en chlorophylle-a (noté « LOG12 » par la suite) pour le 26 mars 2003 et le 13 août 2003 sont montrés Figure 3.6.

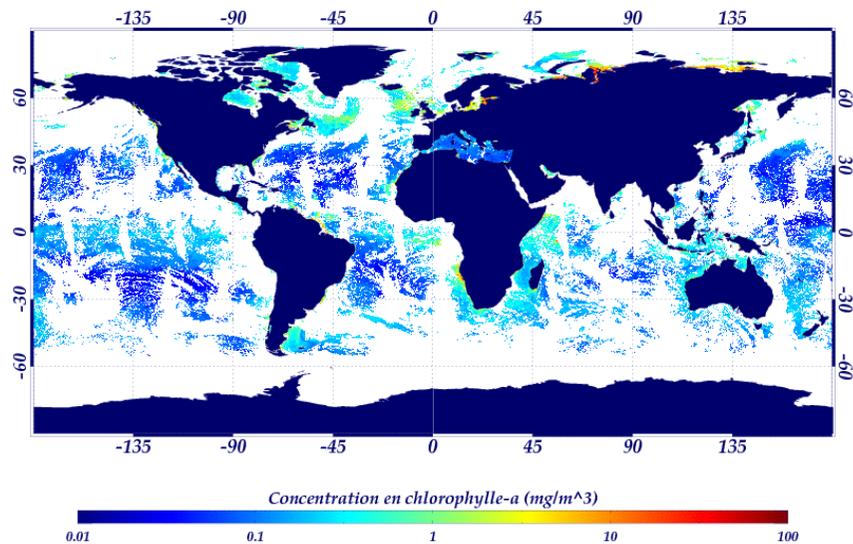
La couverture spatiale des données combinées est de 19.6% pour le 26 mars 2003 et 18.6% pour le 13 août 2003, soit une amélioration de 3.6% et 2.8% respectivement, comparé à la couverture initiale de SeaWiFS, ou 9.4% et 9.2% respectivement, si l'on fait la comparaison par rapport à la couverture initiale de MODIS/Aqua.

La principale limite de cette méthodologie est la discontinuité dans l'erreur associée des données combinées, comme le montre la Figure 3.6. En effet, s'il n'y a qu'une valeur observée, l'erreur de la donnée combinée sera celle de la valeur en question. Par contre, s'il y a une valeur observée par capteur (c'est-à-dire deux valeurs observées pour un même emplacement), l'erreur de la donnée combinée est moins élevée : par exemple, deux pixels d'erreur 102% et 86% donnent un pixel combiné d'erreur ~67%.

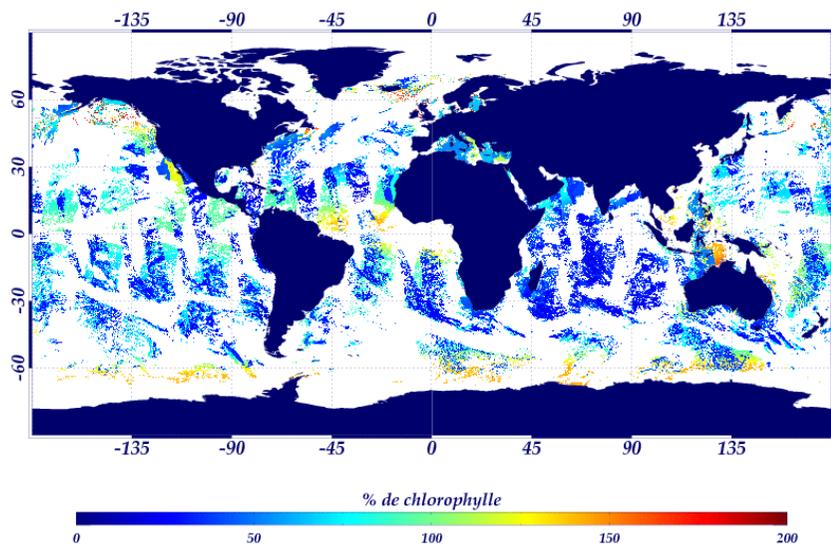
LOG12 - Données combinées - 26/03/2003



LOG12 - Données combinées - 13/08/2003



LOG12 - Erreur - 26/03/2003



LOG12 - Erreur - 13/08/2003

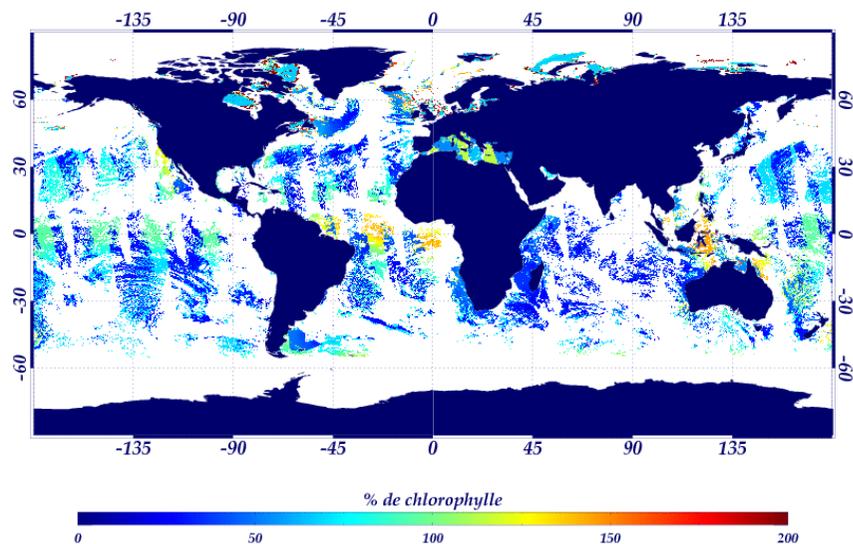


Figure 3.6 – Combinaison des données SeaWiFS et MODIS/Aqua par moyenne pondérée : résultats de l'application de l'algorithme MP1 au logarithme base 10 des données de concentration en chlorophylle-a (algorithme « LOG12 ») pour le 26 mars (à gauche) et le 13 août 2003 (à droite). Pour chaque jour, la carte des données combinées est donnée (en haut), ainsi que la carte des erreurs associées (en bas).

1.3.4 - Comparaison des 4 cas : choix de l'option à retenir

Dans cette partie, nous allons déterminer quelle option de l'algorithme de moyenne pondérée nous allons retenir. Pour cela, trois points sont à considérer.

En premier lieu, l'erreur sur les données combinées et leurs biais par rapports aux données *in situ*. Par conséquent, pour chacun des 4 cas, les comparaisons entre les données combinées et les observations *in situ* ont été examinées, et les erreurs RMS et les biais ont été calculés, suivant les relations décrites au chapitre 2. Les résultats sont montrés dans le Tableau 3.1. On peut noter dans un premier temps que, pour chacun des cas, les valeurs des erreurs RMS et des biais des données combinées sont situés entre ceux de SeaWiFS et MODIS/Aqua étudiés séparément (cf. Chapitre 2, Figure 2.10). Ensuite, les erreurs de mesure et les biais globaux sont très proches dans chacun des 4 cas, les options DEC24 et LOG24 donnant une erreur de mesure légèrement moins élevée que les autres et l'option LOG12 un biais moindre.

Tableau 3.1 – Erreur de mesure et biais (en log et en pourcentage) entre les observations *in situ* et les données combinées pour les 4 cas envisagés pour la moyenne pondérée par l'erreur capteur.

Option	N	r ²	Erreur RMS		Biais	
			log	%	log	%
DEC12	3230	83%	0.33	116	0.089	23
DEC24	4650	84%	0.31	106	0.095	24
LOG12	3230	83%	0.33	116	0.088	22
LOG24	4650	84%	0.31	106	0.095	24

Le second point à considérer, qui est essentiel pour l'objectif d'océanographie opérationnelle, est la durée de calcul. Les durées d'exécution des différents algorithmes sont montrées dans le Tableau 3.2. L'application de la moyenne pondérée aux données brutes (algorithmes DEC12 et DEC24) demande une grande durée d'exécution car il faut calculer beaucoup de distributions lognormales (cf. partie 1.2). L'application sur le logarithme base 10 des données est donc la plus rapide, comme on peut le voir dans le Tableau 3.2.

Tableau 3.2 – Durée d'exécution des différents algorithmes de moyenne pondérée pour un processeur 700 MHz.

Option	Durée moyenne de calcul pour une journée	Durée de calcul pour toute l'année 2003
DEC12	2h	~1 mois (30 jours)
DEC24	1h50	~1 mois (28 jours)
LOG12	18 min	5 jours
LOG24	23 min	6 jours

Le dernier point à considérer est la structure même des données combinées. L'approche MP2 ne fonctionne bien que là où l'on a au moins un pixel MODIS/Aqua ; en effet, si seul SeaWiFS est présent dans une région, la résolution à $1/24^\circ$ des données combinées n'est pas vraiment exacte, puisque chaque pixel SeaWiFS à $1/12^\circ$ est divisé en 4 pour avoir 4 pixels à $1/24^\circ$... L'algorithme MP1 est donc à préférer à l'algorithme MP2.

En conclusion, si on considère les trois critères de choix que sont la qualité (i.e. la comparaison par rapport aux observations *in situ*) et la structure des données combinées, ainsi que la durée de calcul, l'algorithme LOG12 (données combinées de la résolution la plus grossière (celle de SeaWiFS : $1/12^\circ$) et une application sur le logarithme base 10 des valeurs de concentration en chlorophylle-a) est le plus adéquat.

1.4 - Conclusion

Dans cette partie, l'algorithme de combinaison de données par moyenne pondérée par l'erreur capteur a été décrit. Nous avons vu que 4 cas avaient été envisagés pour la combinaison des mesures de SeaWiFS et MODIS/Aqua, en fonction de la résolution des données combinées ($1/12^\circ$ ou $1/24^\circ$) et du type de valeurs sur lesquelles l'algorithme est appliqué (concentrations en chlorophylle-a ou leur logarithme base 10).

Ces 4 cas ont été appliqués pour la combinaison des données SeaWiFS et MODIS/Aqua de concentration en chlorophylle-a à l'échelle globale pour l'année 2003. Une comparaison des résultats avec les observations *in situ* nous a montré que les 4 cas donnaient des erreurs et des biais très proches. Mais dans un objectif d'océanographie opérationnelle, c'est l'algorithme LOG12 (données combinées de la résolution la plus grossière (celle de SeaWiFS : $1/12^\circ$) et une application sur le logarithme base 10 des valeurs de concentration en chlorophylle-a) qui a été choisi comme étant le plus adéquat.

2 - Analyse objective

2.1 - Introduction

La seconde approche étudiée est basée sur l'interpolation optimale, utilisée en océanographie pour la première fois par Bretherton *et al.* [1976]. Cette méthode a déjà été appliquée avec succès aux données satellitaires altimétriques dans diverses applications [Le Traon *et al.*, 1998], la température de surface (Reynolds et Smith [1994] ; Ruiz et Larnicol

[2004] dans la cadre du projet Medspiration) et la salinité (dans le cadre de la mission ESA SMOS ; <http://earth.esa.int/symposia/smos04/presentations/SynergySMOSAqua.pdf>).

Dans cette partie, nous allons tout d'abord formuler le principe de la méthode. Des calculs préliminaires définissent les paramètres d'entrée nécessaires à l'analyse objective : leurs résultats seront montrés dans un second temps. Des tests de sensibilité seront ensuite réalisés, dans l'objectif d'établir la robustesse de l'algorithme. Enfin, nous montrons l'application aux données de concentration en chlorophylle.

2.2 - Méthodologie

2.2.1 - Introduction

L'objectif est de déterminer la valeur d'un champ θ en un point de l'espace et du temps, étant donné N mesures du champ irrégulièrement réparties dans l'espace et le temps Φ_{obs}^i avec $i \in [1, N]$. Une valeur en un point de la grille est interpolée comme étant la combinaison linéaire des observations connues. Les poids sont calculés de manière optimale et prennent en compte la distribution spatiale des observations les unes par rapport aux autres (i.e. la corrélation spatiale), la variance du champ, l'erreur instrumentale et les biais (calculés à partir des matchups, cf. Chapitre 2-3). Par conséquent, cette méthode compense les différences entre capteurs, que ce soit au niveau de leur conception ou de leurs caractéristiques, des particularités de calibration et du traitement des données. Ainsi, deux valeurs observées à des coordonnées spatio-temporelles identiques ne sont pas forcément identiques. De même, une valeur interpolée n'est pas forcément identique à une valeur observée à des coordonnées spatio-temporelles identiques.

2.2.2 - Formulation

Le champ à estimer se calcule comme suit :

$$\theta_{est}(x) = \sum_{i=1}^N \sum_{j=1}^N A_{ij}^{-1} C_{xj} \Phi_{obs}^i$$

où :

$\Phi_{obs}^i = \Phi_i + \varepsilon_i$ est la valeur mesurée, Φ_i étant la « vraie » valeur et ε_i l'erreur de mesure,

A_{ij} est la matrice de covariance entre les observations :

$$A_{ij} = \langle \Phi_{obs}^i, \Phi_{obs}^j \rangle = \langle \Phi_i, \Phi_j \rangle + \langle \varepsilon_i, \varepsilon_j \rangle,$$

C_{xj} est le vecteur de covariance entre les observations et le point devant être estimé :

$$C_{xj} = \langle \theta(x), \Phi_{\text{obs}^j} \rangle = \langle \theta(x), \Phi_j \rangle.$$

La variance de l'erreur associée est donnée par :

$$e^2 = C_{xx} - \sum_{i=1}^N \sum_{j=1}^N C_{xi} C_{xj} A_{ij}^{-1}.$$

L'analyse objective utilisée par Le Traon *et al.* [1998] décompose l'erreur en un bruit blanc et un biais, en exprimant $\langle \varepsilon_i, \varepsilon_j \rangle$ de la manière suivante:

$\langle \varepsilon_i, \varepsilon_j \rangle = \delta_{ij} b^2$ pour des points de coordonnées (i, j) provenant de capteurs différents,

$\langle \varepsilon_i, \varepsilon_j \rangle = \delta_{ij} b^2 + E$ pour des points de coordonnées (i, j) provenant du même capteur,

où b^2 est la variance du bruit de mesure (i.e. le carré des erreurs RMS calculées au Chapitre 2) et E la variance du biais (i.e. le carré des moyennes des différences calculées au Chapitre 2).

En raison du grand nombre de mesures, l'interpolation est réalisée avec des observations incluses dans une « bulle d'influence », dont la taille est définie par les échelles de corrélation zonales R_x et méridionales R_y .

D'autre part, cette méthode requiert une hypothèse de champ gaussien. Or, nous avons vu au Chapitre 2 que les données de concentration en chlorophylle sont lognormalement distribuées, c'est-à-dire que c'est le logarithme des données qui a une distribution gaussienne. L'analyse objective a néanmoins été appliquée directement sur les données de concentration en chlorophylle et les résultats obtenus sont mauvais. En effet, des valeurs irréalistes peuvent être obtenues à des endroits où la différence de variance entre les points de la bulle d'influence est très importante ; la matrice de covariance est ainsi numériquement mal conditionnée et son inversion pose problème. Un tel problème n'apparaît pas si on utilise le logarithme des données car de telles différences n'existent pas. Par la suite, seule l'étude sur le logarithme base 10 des données sera montrée.

2.2.3 - Implémentation de l'algorithme

Au cours de ce travail de thèse, le logiciel d'analyse objective utilisé par CLS pour la température de surface dans le cadre du projet Medspiration a été adapté pour être appliqué aux données de concentrations en chlorophylle-a. L'organigramme du logiciel est montré Figure 3.7.

Deux fichiers de paramètres sont nécessaires à son bon fonctionnement. Un fichier de configuration, contenant les paramètres propres au fonctionnement de l'analyse :

- ☒ le sous-échantillonnage, i.e. la sous-grille pour le calcul de la matrice de covariance,
- ☒ le fichier contenant le masque terre / mer, nécessaire à la détermination des positions d'étude valides,
- ☒ le fichier de distance à la mer : dans le logiciel, une zone de terre séparant deux zones d'océan dans la bulle d'influence empêche de prendre en compte les observations de l'autre côté de la terre pour la position d'étude.
- ☒ le fichier de bathymétrie, pour limiter le calcul trop près des côtes,
- ☒ le fichier de variance du signal,
- ☒ le flag indiquant la méthode pour centrer les données,
- ☒ l'indice de choix pour la fonction de corrélation,
- ☒ le flag indiquant si les coefficients de la fonction de corrélation sont fixes ou variables spatialement, ainsi que la valeur ou le fichier associé,
- ☒ le flag indiquant si les rayons de corrélation spatio-temporels sont fixes ou variables spatialement, ainsi que la valeur ou le fichier associé,
- ☒ la valeur maximale de chlorophylle à prendre en compte dans les résultats,
- ☒ l'erreur maximale autorisée,
- ☒ le nombre de points minimum qui doivent être présents dans la bulle d'influence pour que le calcul soit réalisé,
- ☒ le nombre de points maximum dans la bulle d'influence gardés pour calculer la matrice de corrélation.

Le fichier de commande du logiciel d'analyse objective contient les paramètres suivants :

- ☒ grille d'intérêt pour les données combinées (zone + échantillonnage spatial + dates),
- ☒ les fichiers relatifs aux observations nécessaires, contenant les anomalies, i.e. la différence entre le logarithme base 10 de la carte de chlorophylle du jour donné et le logarithme base 10 de la carte de climatologie du mois correspondant,
- ☒ pour chaque fichier d'observation, le flag indiquant si l'erreur de mesure est fixe ou variable spatialement, ainsi que la valeur ou le fichier associé,
- ☒ pour chaque fichier d'observation, le flag indiquant si le biais est fixe ou variable spatialement, ainsi que la valeur ou le fichier associé,
- ☒ le fichier de climatologie mensuel correspondant au mois de la date de traitement,
- ☒ le fichier texte qui contiendra les résultats.

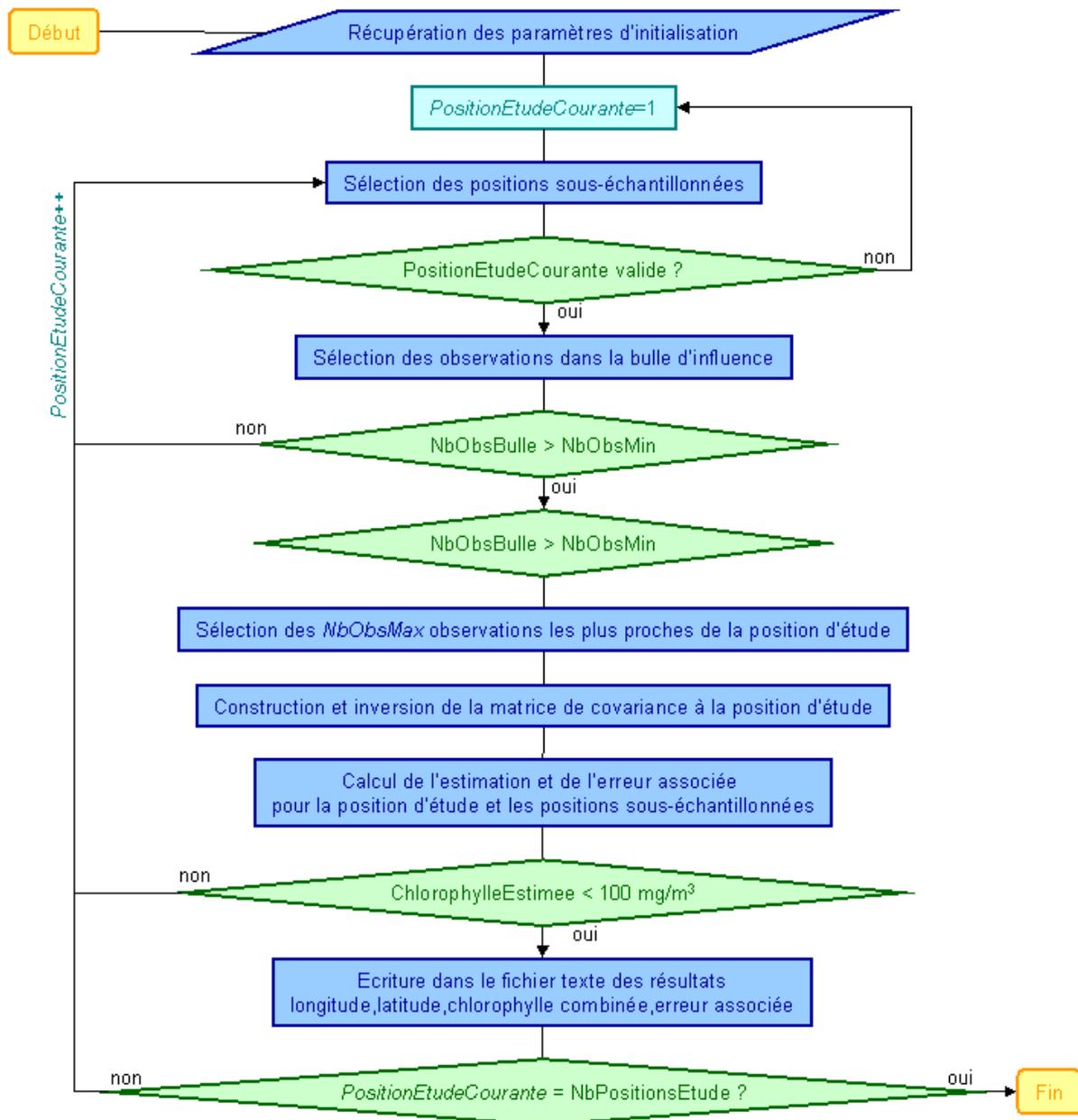


Figure 3.7 – Organigramme du logiciel d'analyse objective.

2.2.4 - Conclusion

Dans cette partie, le concept de l'analyse objective a été introduit, de même que l'implémentation de cet algorithme.

Nous avons vu que certains fichiers à mettre en entrée du logiciel nécessitent des calculs préliminaires. Nous allons les découvrir dans la partie suivante.

2.3 - Calculs préliminaires

2.3.1 - Introduction

Des calculs préliminaires sont nécessaires pour déterminer les paramètres à mettre en entrée de l'analyse objective : une estimation *a priori* du champ, la variance temporelle, un modèle de fonction de corrélation et les échelles de corrélation spatiales.

Ces paramètres, calculés sur le logarithme des données de concentration en chlorophylle puisque l'analyse objective est elle-même appliquée à ce format de données, sont montrés dans cette partie.

2.3.2 - Climatologie mensuelle

L'objectif de l'analyse objective est de corriger la première estimation d'un champ (« first guess field »), plus ou moins connue, par l'apport des données journalières. En effet, les données en entrée de l'algorithme (i.e. les $\Phi_{obs,i}$), sont des anomalies de logarithme de chlorophylle (appelée par la suite « log-anomalies »), c'est-à-dire le logarithme du champ de concentrations en chlorophylle moins le logarithme de cette estimation *a priori*. Plusieurs choix sont possibles quant à ce champ de base : il peut être la carte de chlorophylle du jour précédent, la moyenne des cartes de chlorophylle de la semaine précédente, une climatologie mensuelle, etc. Reynolds et Smith [1994] se sont déjà penchés sur ce problème pour l'application de l'analyse objective à la température de surface de la mer (SST : Sea Surface Temperature). Une climatologie mensuelle ou la carte de chlorophylle du jour précédent sont deux solutions envisageables, avec leurs avantages et leurs inconvénients. L'analyse du jour précédent a l'avantage d'être plus précise qu'une climatologie. Néanmoins, la climatologie est plus homogène en terme de statistiques. Nous avons donc choisi la solution de la climatologie mensuelle. Ainsi, l'estimation de la fonction de corrélation est plus robuste et plus facile à développer. De plus, il est aussi plus raisonnable d'utiliser un champ *a priori* le plus rempli spatialement possible.

La climatologie mensuelle a été construite à partir des données mensuelles de SeaWiFS comprises entre octobre 1997 et septembre 2004, soit 7 ans de données. Plusieurs méthodes ont été envisagées :

- ☒ moyenne mois par mois : la moyenne des mois de janvier donne le mois de janvier de la climatologie, etc.,
- ☒ moyenne glissante sur 3 mois (par exemple : janvier 2002 équivaut à la moyenne de décembre 2001, janvier 2002 et février 2002) puis moyenne mois par mois,
- ☒ filtrage temporel à un an à l'aide d'une transformée de Fourier.

La Figure 3.8 montre le résultat des 3 méthodes pour des pixels présentant différents taux de présence temporelle (a : 100% ; b : 95% ; c : 70% ; d : 50%). La première méthode (courbe orange) a été éliminée, car toute valeur « extraordinaire » est répercutée dans la climatologie : pour le point de coordonnées 47°W-1° N par exemple, la très grande valeur de concentration en chlorophylle de février 1999 se retrouve dans le mois de février de la climatologie. La 3^e méthode (courbe bleue) a été préférée à la 2nde (courbe verte) en raison de sa régularité ; en effet, quand le taux de présence temporel n'est plus de 100%, la courbe verte devient plus irrégulière (Figure 3.8c et d).

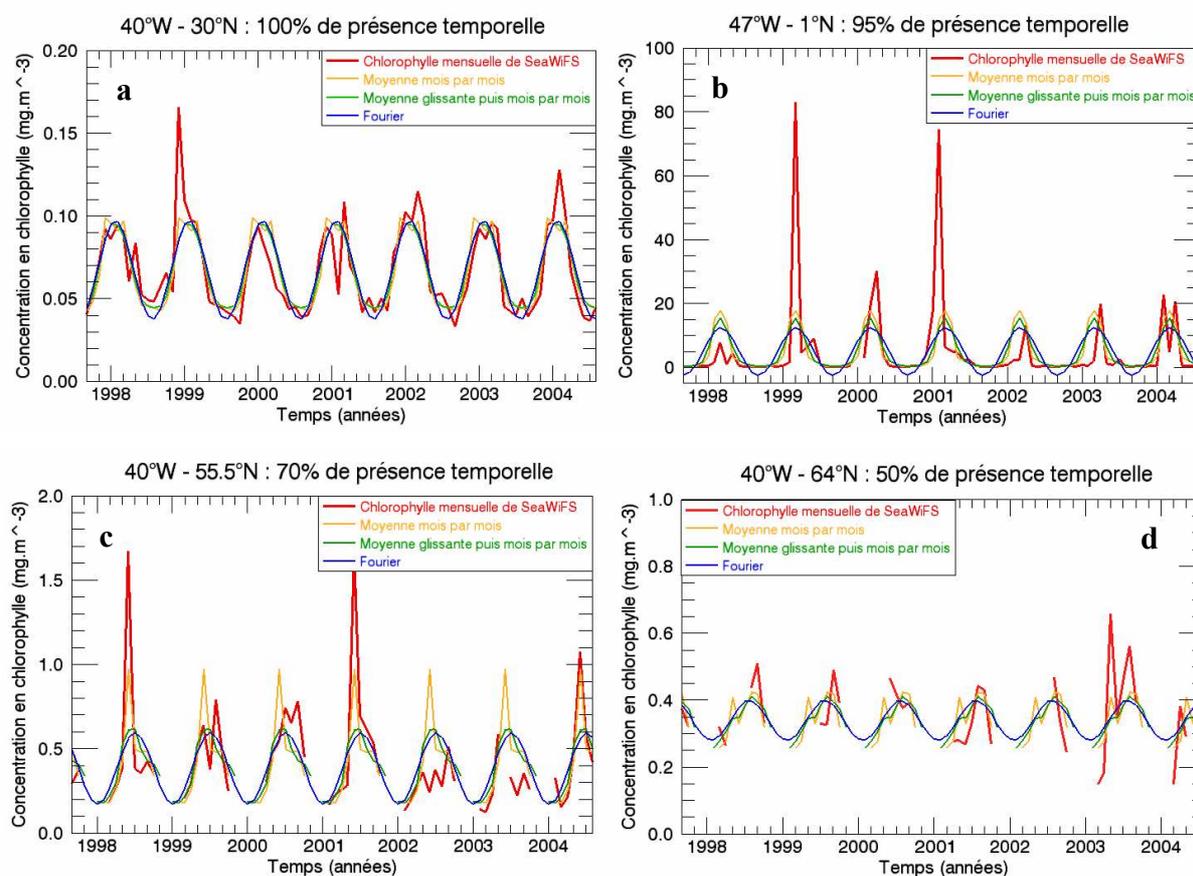


Figure 3.8 – Comparaison des différentes méthodes de calcul de la climatologie (moyenne mois par mois en orange, moyenne glissante sur 3 mois puis mois par mois en vert et filtrage à 1 an par transformée de Fourier en bleu foncé) pour (a) le point 40°W-30°N (taux de présence temporelle de 100%), (b) le point 47°W-1°N (taux de présence temporelle de 95%), (c) le point 40°W-55.5°N (taux de présence temporelle de 70%) et (d) le point 40°W-64°N (taux de présence temporelle de 50%).

Néanmoins, la Figure 3.8b montre que, pour une trop grande variabilité de chlorophylle (de l'ordre de 10^{-2} à 10^2 mg.m⁻³ par exemple), le filtrage à 1 an par transformée de Fourier amène des concentrations en chlorophylle « négatives ». Le moyen d'éviter ce problème est

d'appliquer le filtrage sur le logarithme des données et de repasser dans le domaine physique. Ainsi, les concentrations en chlorophylle obtenues sont toujours positives, comme le montre la Figure 3.9b pour le point situé à 47°W-1°N. Pour ce qui est des pixels ayant une variabilité moins importante, le résultat est quasiment identique (Figure 3.9a).

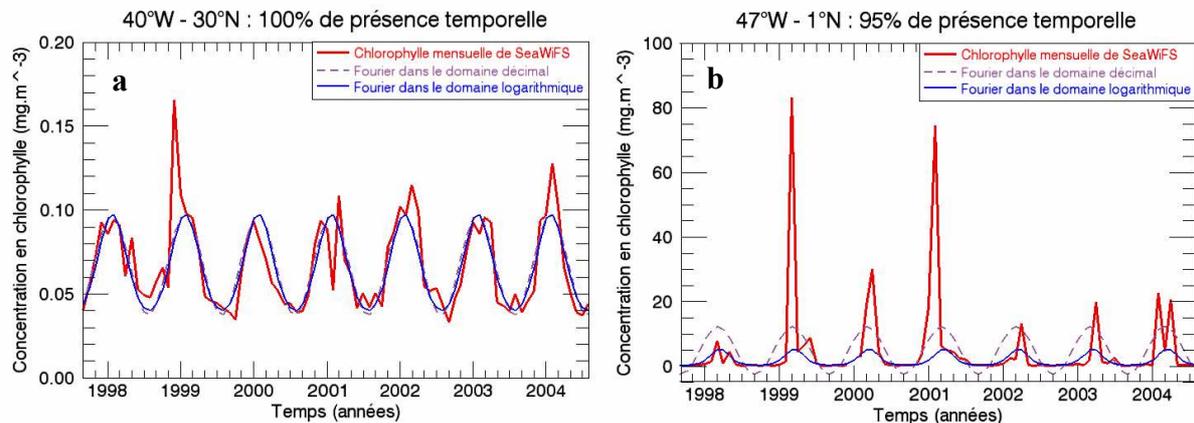


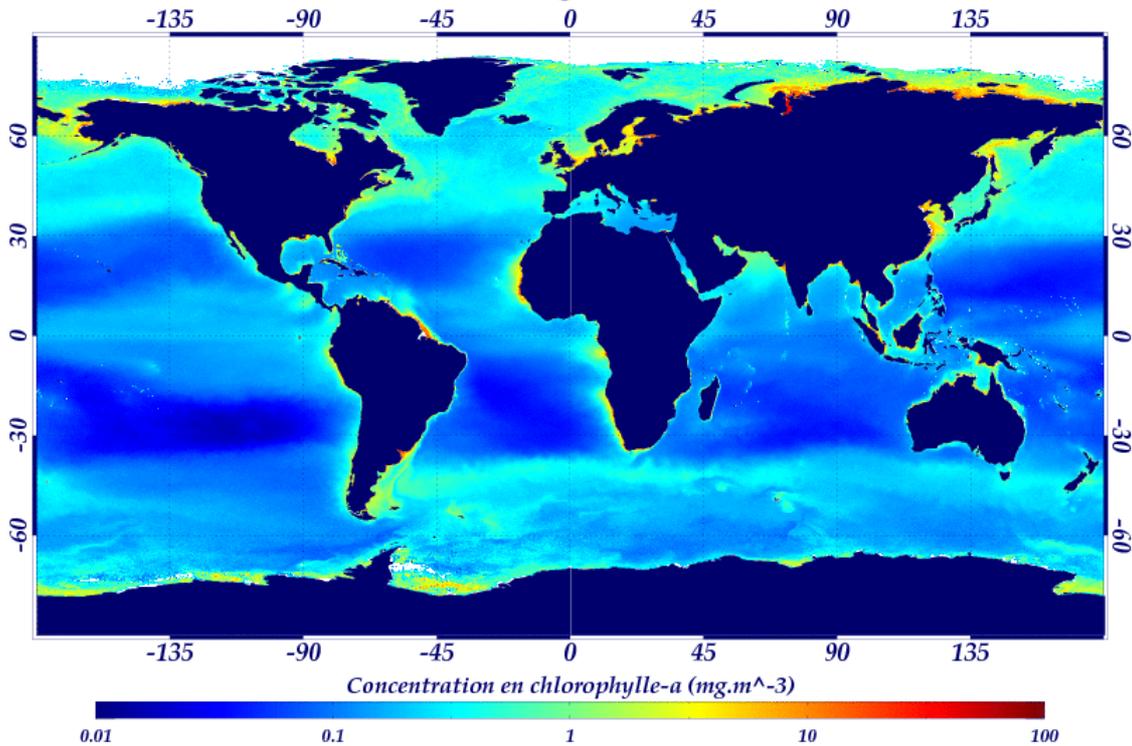
Figure 3.9 - Comparaison de la méthode de filtrage à 1 an par transformée de Fourier dans le domaine décimal (courbe violette en pointillés) et dans le domaine logarithmique (courbe bleue) pour (a) le point 40°W-30°N (100% de présence temporelle) et (b) le point 47°W-1°N (95% de présence temporelle).

Les cartes de climatologie mensuelle obtenues pour les mois de mars et août sont montrées Figure 3.10. Les autres mois sont montrés en Annexe A-2.1.

2.3.3 - Variance spatio-temporelle

La variance temporelle calculée sur le logarithme base 10 des données journalières de concentration en chlorophylle de l'année 2003 provenant du capteur SeaWiFS est montrée Figure 3.11. Pour éviter des différences marquées entre les points voisins, ce qui est inadéquat pour l'analyse objective, un filtrage médian à 50 km a été appliqué sur le champ. La Figure 3.11 montre que la variabilité est particulièrement forte près des côtes. Ceci est principalement dû au fait que les eaux côtières contiennent des substances aux propriétés optiques très différentes [Uz et Yoder, 2004], surtout à l'embouchure des fleuves (l'Amazone, l'Orénoque, le Congo, ...). En océan ouvert, la variabilité méso-échelle est associée à une activité importante, comme la réflexion des Aiguilles, la confluence des courants du Brésil et des Malouines ou le plume du fleuve Amazone. La région de plus forte variabilité dans l'Océan Pacifique est celle du gyre oligotrophe du Pacifique Sud, région contenant les plus faibles concentrations de chlorophylle jamais observées. La variance y est élevée du fait que son calcul a tendance à amplifier la variabilité quand la moyenne est faible [Uz et Yoder, 2004].

Climatologie en Mars



Climatologie en Aout

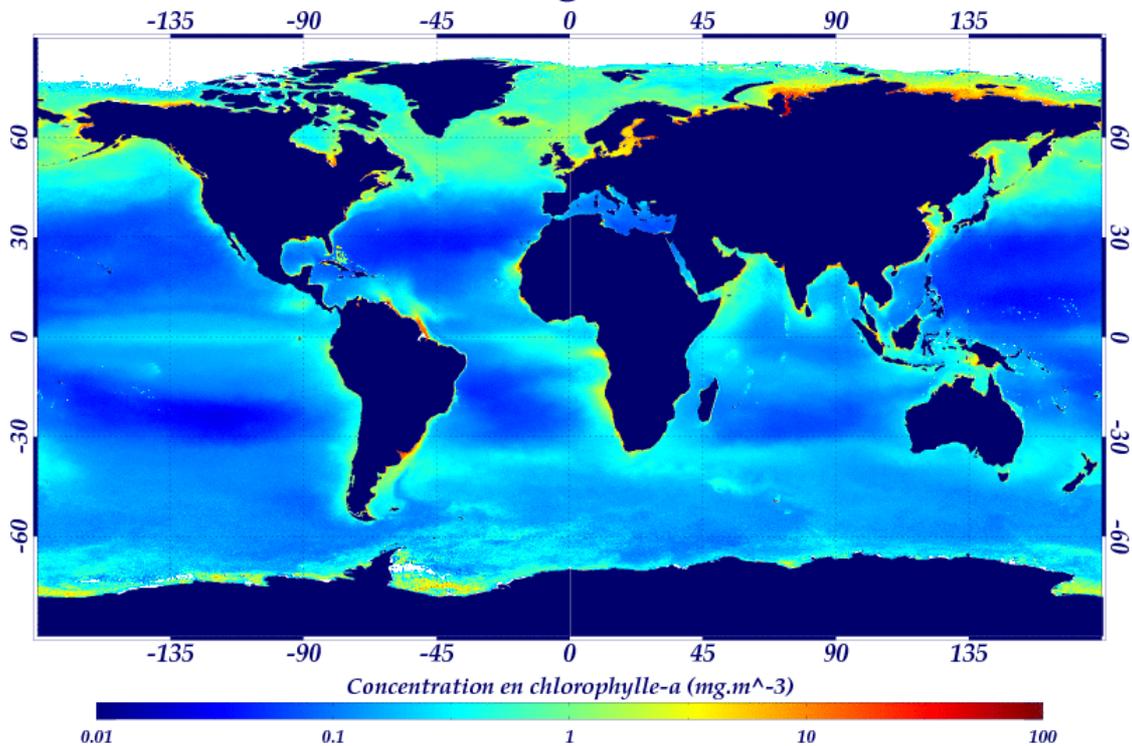


Figure 3.10 - Les cartes de climatologie mensuelle obtenues par filtrage à 1 an par transformée de Fourier appliquée au logarithme base 10 des données.

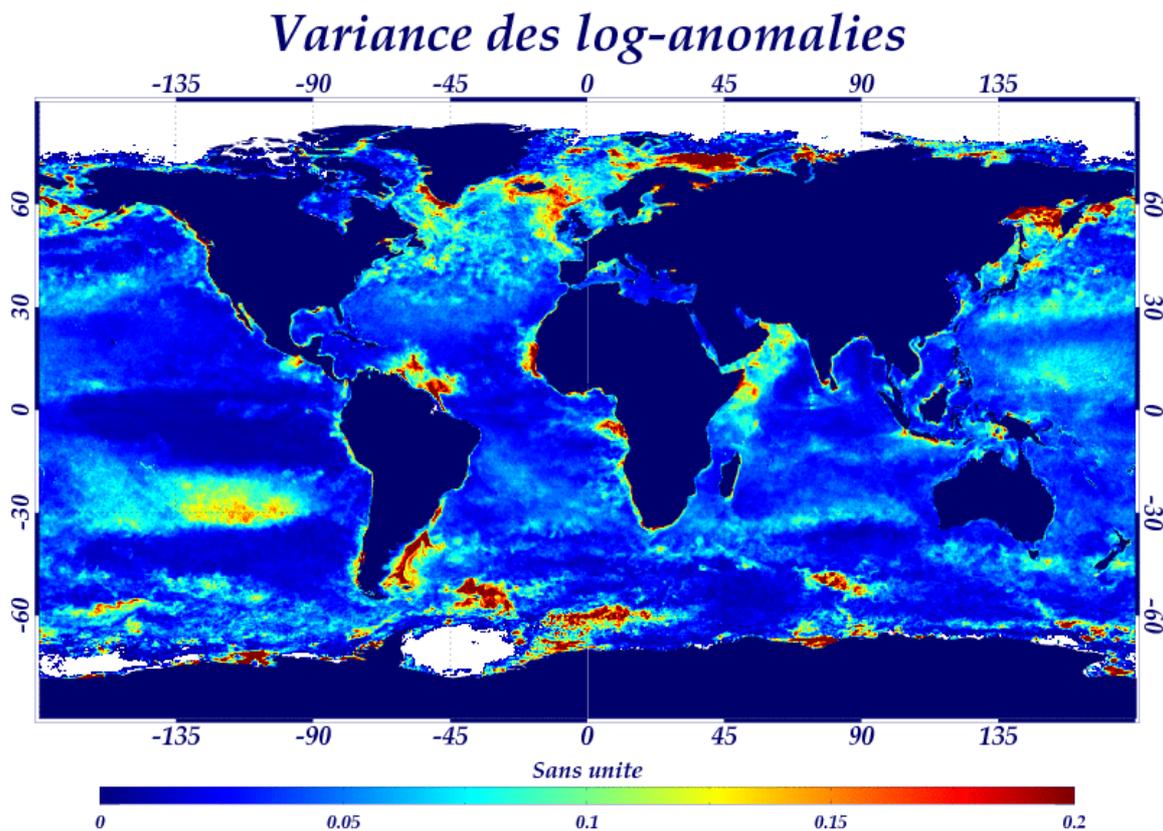


Figure 3.11 – Variance temporelle calculée sur le logarithme des données journalières de concentration en chlorophylle de l’année 2003 provenant du capteur SeaWiFS. A noter que les zones en blanc indiquent des régions pour lesquelles il n’y a pas de donnée sur toute l’année 2003.

2.3.4 - Fonction de corrélation

La fonction de corrélation spatio-temporelle empirique a tout d’abord été calculée sur les log-anomalies SeaWiFS en chaque point d’une grille 1° x 1° en utilisant l’équation ci-après. Seules les valeurs incluses dans un sous-domaine centré autour de chaque emplacement étudié (SousDomaine) sont considérées. Le calcul est réalisé seulement si au moins 50% du SousDomaine est complété.

$$\text{CORR}_{\text{emp}}(\Delta\text{long}, \Delta\text{lat}, \Delta t) = \frac{\sum_{\text{carte}=1}^{\text{NombreDeCartes}} \frac{\sum_{\text{carte}=1}^{\text{NombreDeCartes}} \text{PROD}(\text{carte}, \text{paire}, \Delta\text{long}, \Delta\text{lat}, \Delta t)}{\text{VAR}(\text{SousDomaine}) \times \text{NombreDePaires}}}{\text{NombreDeCartes}}$$

où:

\otimes paire = (long1, lat1, 0) ↔ (long2, lat2, Δt)

\otimes Δlong = long2 – long1 ,

\otimes Δlat = lat2 – lat1

☞ NombreDePaires est le nombre de paires de valeurs incluses dans le SousDomaine de calcul

☞ NombreDeCartes est le nombre de cartes de log-anomalies utilisées

☞ le produit PROD est calculé comme suit :

$$\text{PROD}(\text{carte}, \text{paire}, \Delta\text{long}, \Delta\text{lat}, \Delta t) = [\text{ANO}(\text{carte}, \text{long1}, \text{lat1}, 0) - \text{MOYENNE}(\text{SousDomaine})] \times [\text{ANO}(\text{map}, \text{long2}, \text{lat2}, \Delta t) - \text{MOYENNE}(\text{SousDomaine})]$$

où ANO est la log-anomalie,

☞ la variance dans le SousDomaine est :

$$\text{VAR}(\text{SousDomaine}) = \frac{\sum_{i\text{Long}=1}^{\text{NbLong}} \sum_{i\text{Lat}=1}^{\text{NbLat}} \sum_{i\text{Tps}=1}^{\text{NbTps}} [\text{ANO}(\text{carte}, i\text{Long}, i\text{Lat}, i\text{Tps}) - \text{MOYENNE}(\text{SousDomaine})]^2}{\text{NombreDePts}}$$

où NbLong, NbLat et NbTps sont respectivement le nombre de points en longitude, latitude et temps dans le SousDomaine, et NombreDePts est le nombre de points inclus dans le SousDomaine.

Le choix de la taille du SousDomaine dans un tel calcul est important. Un petit SousDomaine donne des informations très locales. Une grande zone donne des informations plus lisses mais plus précises car le calcul est réalisé sur plus de paires. Pour avoir une première idée de la taille du SousDomaine, nous avons utilisé les échelles de corrélation spatio-temporelle issues d'une étude précédente : appliquant la méthode du semi-variogramme à des images SeaWiFS journalières de niveau 3 (SMI, reprocessing #2) pour l'année 1998, Doney *et al.* [2003] ont trouvé, pour l'océan global, un rayon de corrélation zonal maximum de 250 km et un rayon de corrélation méridional maximum de 200 km. Nous avons donc choisi ces dimensions, 250 km x 200 km, pour la taille du SousDomaine gardée constante sur tout le globe pour calculer nos rayons de corrélation empiriques. Considérant deux images séparées de 1 jour, Uz et Yoder [2004] ont trouvé des corrélations comprises entre 10% et 40%. Nous avons choisi dans un premier temps de travailler en 2 dimensions (longitude, latitude) et donc d'obtenir une fonction de corrélation spatiale. Nous avons aussi calculé la fonction de corrélation avec une échelle temporelle de 5 jours, mais les résultats n'ont pas été concluants.

Les fonctions de corrélation empirique sont trop bruitées pour être mises telles quelles en paramètres d'entrée : il faut donc les modéliser. Pour ce faire, deux fonctions ont été étudiées : la fonction exponentielle et la fonction inverse.

Regardons, en premier lieu, le modèle exponentiel, utilisé par Kwiatkowska et Fargion [2002a]. La forme générale de cette fonction s'écrit :

$$\text{CORR}_{\text{exp}}(r) = A_1 \exp(A_2 r) + A_3$$

où :

$$r = \sqrt{\left(\frac{x}{R_x}\right)^2 + \left(\frac{y}{R_y}\right)^2}, \text{ avec } R_x \text{ et } R_y \text{ les rayons de corrélation zonal et méridional}$$

respectivement (il s'agit de la première intersection de la fonction CORR_{emp} avec les axes des abscisses et des ordonnées respectivement),

A_1 , A_2 et A_3 sont des coefficients à déterminer.

La fonction doit remplir les conditions suivantes :

elle doit être strictement décroissante,

$\text{CORR}_{\text{exp}}(r=0) = 1$: un coefficient de corrélation maximal pour un rayon nul (normal...),

$\text{CORR}_{\text{exp}}(r=1) = 0$: un coefficient de corrélation nul aux échelles de corrélation.

On obtient ainsi la relation : $\text{CORR}_{\text{exp}}(r) = (1 - \alpha) \left(\frac{\alpha}{\alpha - 1}\right)^r + \alpha$ avec $\alpha \in]0,1[$. Plus $\alpha \rightarrow 0$, plus

la fonction est pointue. Plus $\alpha \rightarrow 1$, plus la fonction tend vers une droite.

Le second modèle est basé sur une fonction inverse, dont la formulation générale est (A_1 , A_2 , A_3 et A_4 étant les coefficients à déterminer) :

$$\text{CORR}_{\text{inv}}(r) = \frac{A_1}{A_2 r + A_3} + A_4.$$

Considérant les mêmes contraintes que précédemment, on obtient :

$$\text{CORR}_{\text{inv}}(r) = \beta + \frac{\beta(1-\beta)}{\beta-r} \text{ avec } \beta < 0. \text{ Là encore, plus } \beta \rightarrow 0, \text{ plus la fonction est pointue.}$$

Plus β est petit (i.e. grand en valeur absolue), plus la fonction tend vers une droite.

Le spectre de ces deux modèles de corrélation est positif pour chacun des nombres d'onde. Par conséquent, ils sont tous deux définis non-négatifs, condition nécessaire à l'analyse objective [Bretherton, 1976].

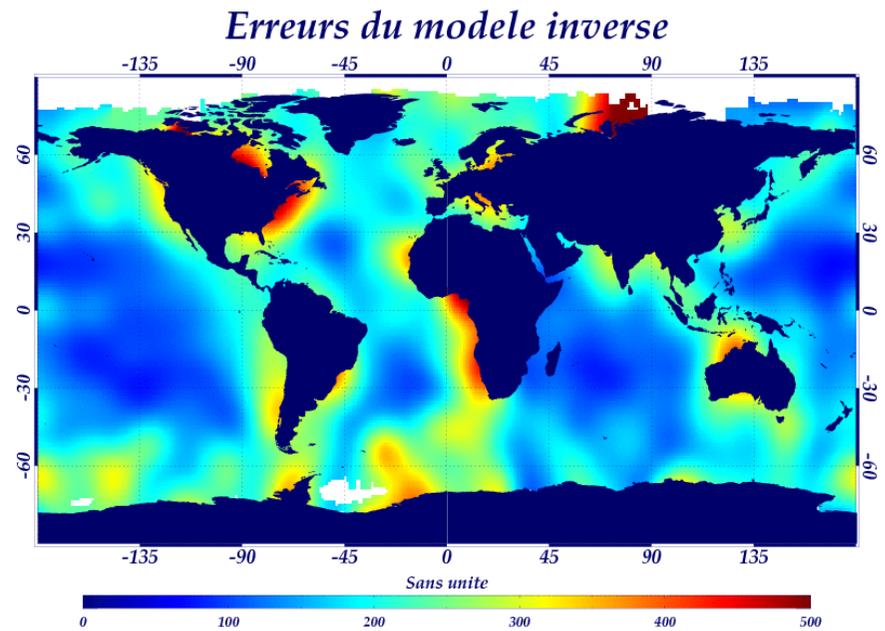
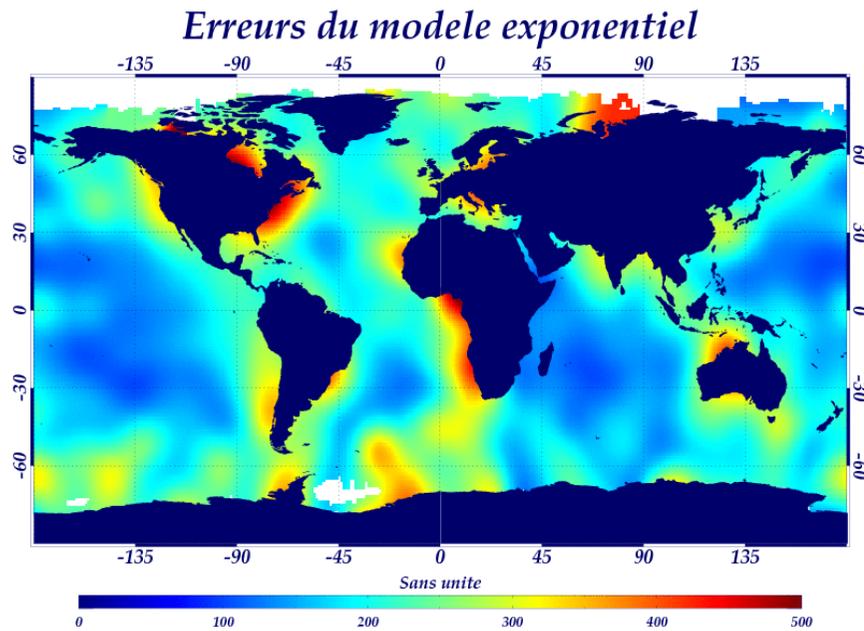
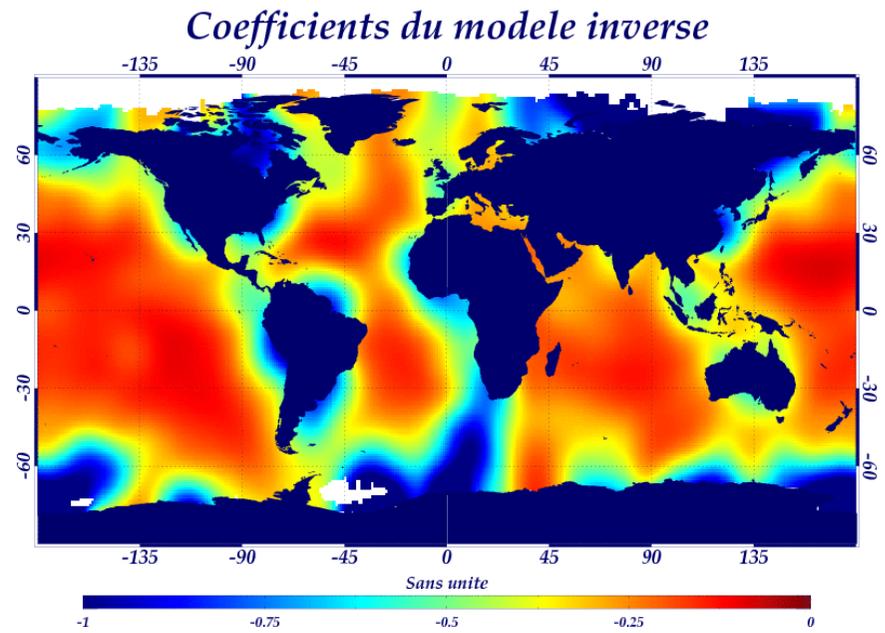
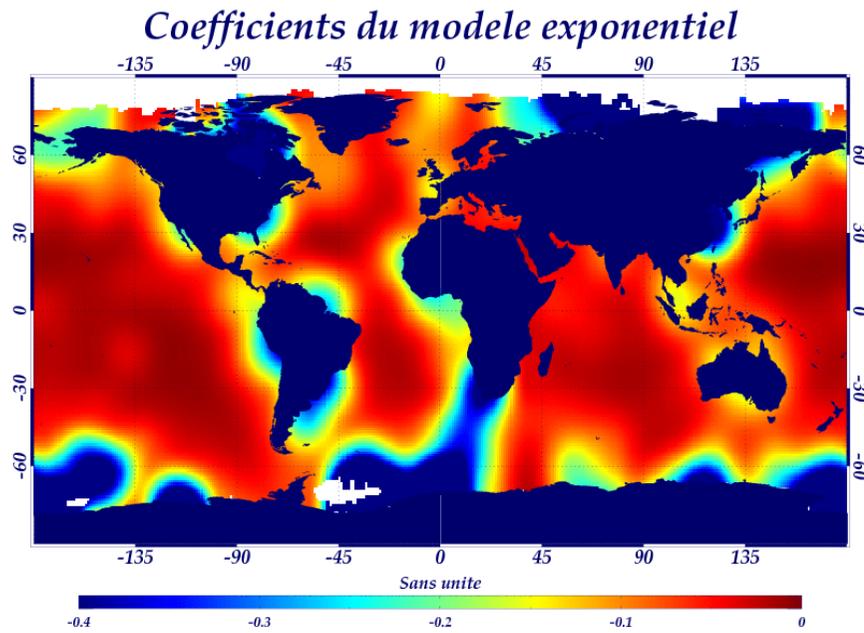


Figure 3.12 – A gauche : distribution spatiale des coefficients α pour le modèle exponentiel (en haut) et l'erreur absolue entre le modèle et la fonction empirique (en bas). A droite, même chose pour le modèle inverse.

En chaque point de la grille $1^\circ \times 1^\circ$, les coefficients α et β ont été calculés pour correspondre au mieux à la fonction de corrélation empirique. La Figure 3.12 montre la répartition globale de ces coefficients. Pour chacun des deux modèles, on peut apercevoir une dépendance océan côtier / océan ouvert. En océan ouvert, les coefficients sont proches de 0, indiquant une fonction de corrélation très pointue et donc des points peu corrélés avec leurs voisins. En océan côtier, les coefficients sont plus élevés (en valeur absolue), le modèle de corrélation est donc plus évasé, indiquant que les points sont mieux corrélés avec leurs voisins.

La Figure 3.13 compare les erreurs modèle / fonction empirique obtenues pour chaque modèle. On voit que la fonction inverse est plus adaptée que la fonction exponentielle à la modélisation des fonctions de corrélation empiriques. C'est donc le modèle inverse qui sera utilisé par la suite.

Comparaison modele inverse / modele exponentiel

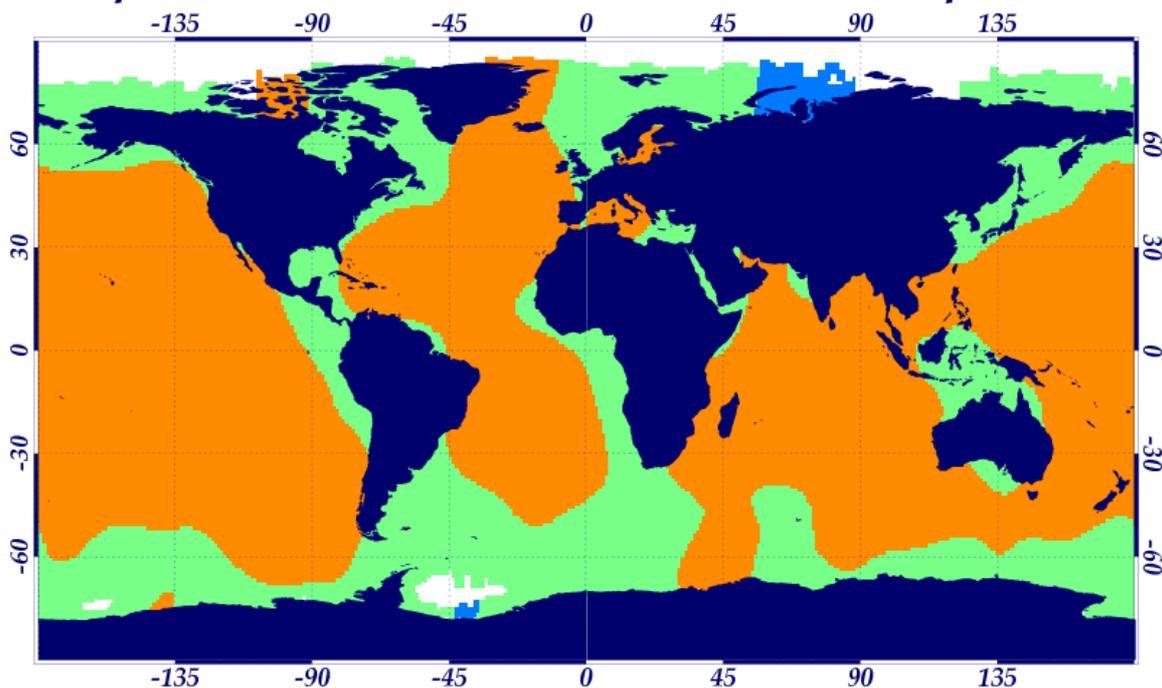
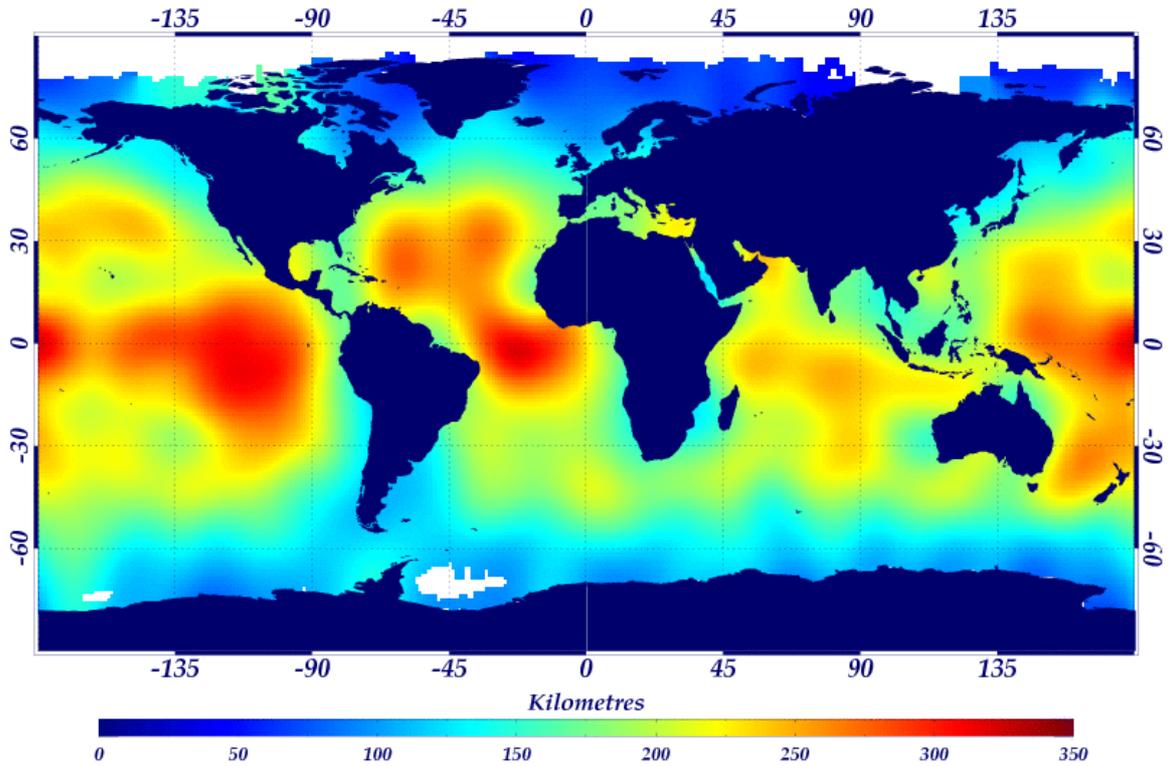


Figure 3.13 – Comparaison entre les modèles inverse et exponentiel. La couleur orange correspond aux emplacements pour lesquels le modèle inverse est plus adapté, la couleur verte aux emplacements pour lesquels les deux modèles sont à peu près équivalents et la couleur bleue aux emplacements pour lesquels le modèle exponentiel est le plus adapté.

Rayons de corrélation zonaux



Rayons de corrélation méridionaux

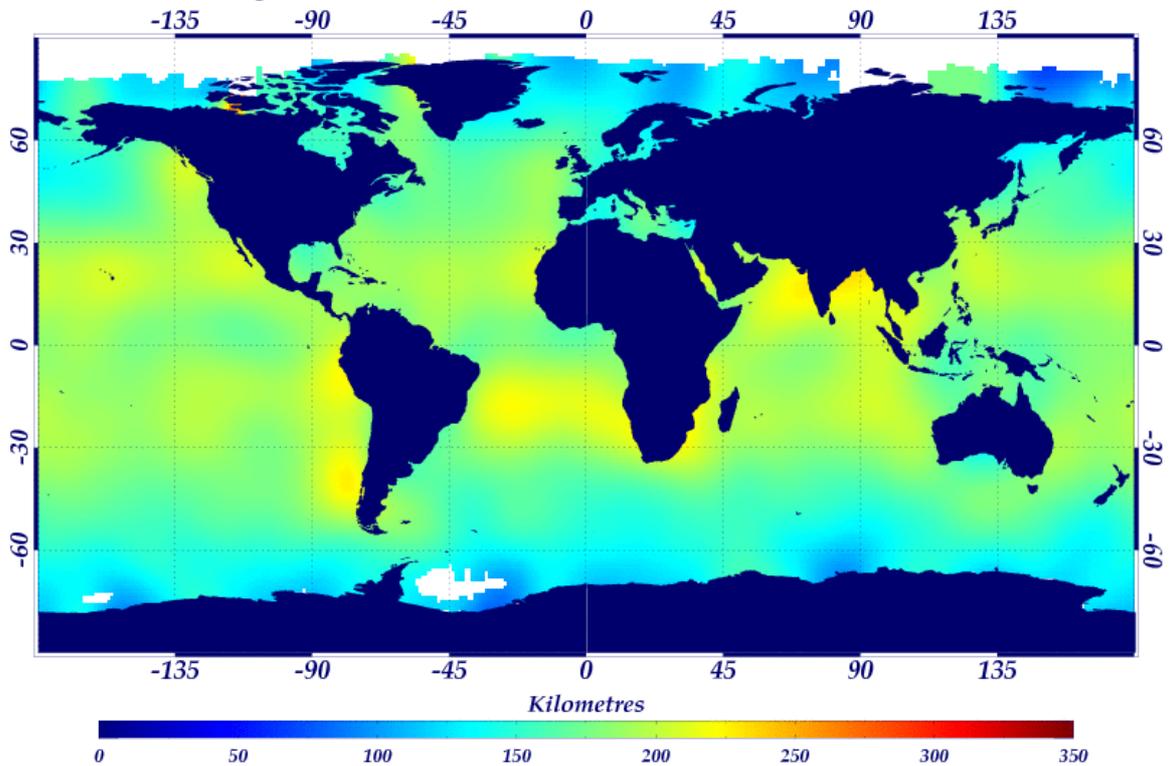


Figure 3.14 – Répartition spatiale des rayons de corrélation zonaux (à gauche) et méridionaux (à droite).

2.3.5 - Rayons de corrélation spatiaux

Le rayon de corrélation zonal (resp. méridional) est le premier point d'intersection de la fonction de corrélation empirique avec l'axe des abscisses (resp. l'axe des ordonnées). Ils définissent la taille de la bulle d'influence.

La Figure 3.14 montre les rayons de corrélation zonaux et méridionaux calculés pour toute l'année 2003. On y voit que les rayons de corrélation zonaux dépendent clairement de la latitude, alors que les rayons de corrélation méridionaux sont presque constants sur toute la zone (170 km en moyenne).

La Figure 3.15 montre la comparaison entre les rayons de corrélation zonaux et méridionaux. On peut constater que, généralement, la corrélation zonale est plus importante que la corrélation méridionale en plein océan tandis que c'est plutôt l'inverse dans les eaux côtières.

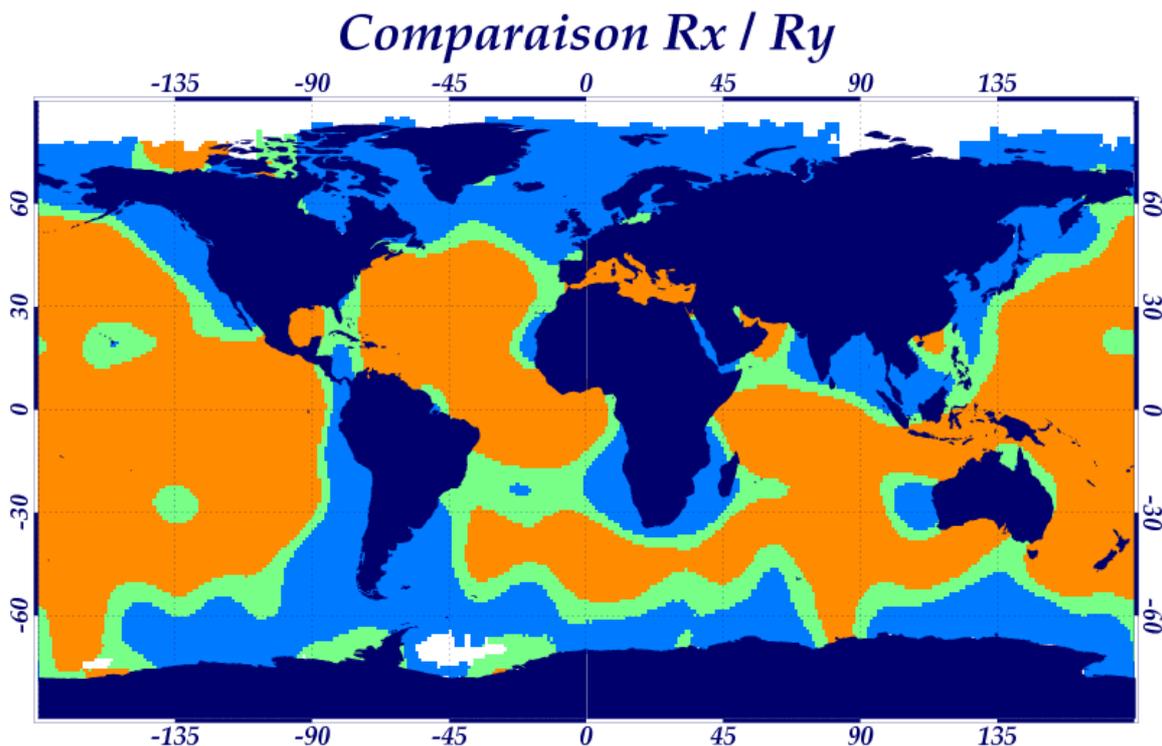


Figure 3.15 – Comparaison entre les rayons de corrélation zonaux et les rayons de corrélation méridionaux. La couleur orange correspond aux emplacements pour lesquels le rayon de corrélation zonal est plus élevé que le rayon de corrélation méridional (bulle d'influence ovale horizontalement), la couleur verte aux emplacements pour lesquels les rayons de corrélation zonal et méridional sont à peu près équivalents (bulle d'influence quasi ronde) et la couleur bleue aux emplacements pour lesquels le rayon de corrélation zonal est moins élevé que le rayon de corrélation méridional (bulle d'influence ovale verticalement).

2.3.6 - Conclusion

Dans cette partie, nous avons tout d'abord montré qu'il était adapté de choisir un champ *a priori* à partir d'une climatologie mensuelle calculée par filtrage à un an par transformée de Fourier. En effet, ce « first-guess » est alors plus étendu et, puisque plus lisse, il ne tient pas compte des éventuelles valeurs « extraordinaires » que peut prendre la chlorophylle certaines années. Puis nous avons constaté que la variabilité était importante dans les eaux côtières, ainsi que dans les eaux de forte activité (rétroflexion des Aiguilles, confluence des courants du Brésil et des Malouines ou embouchures de fleuves) et le gyre oligotrophe dans le Pacifique Sud. Nous avons ensuite comparé deux modèles pour la fonction de corrélation et choisi le modèle inverse, qui est plus proche du modèle empirique que le modèle exponentiel. Enfin, nous avons vu que les rayons de corrélation zonaux avaient une forte dépendance latitudinale contrairement aux rayons de corrélation méridionaux.

Nous allons maintenant tester la sensibilité de l'analyse objective aux paramètres qui viennent d'être calculés.

2.4 - Tests de sensibilité

2.4.1 - Introduction

Dans cette partie, nous testons la sensibilité de l'analyse objective aux paramètres de configuration listés dans la partie 2.2.3. Les résultats de ces tests ont été analysés pour le 26 mars 2003 dans la région du Gulf Stream, qui a une forte variabilité spatio-temporelle. La Figure 3.16 montre les données de SeaWiFS et MODIS/Aqua, ainsi que la répartition spatiale des données dans cette région et pour cette date.

2.4.2 - Rayon de corrélation temporel

Nous avons appliqué l'analyse objective avec des rayons de corrélation temporels de 1, 2, 3, 4 et 5 jours. On obtient les cartes montrées en Annexe A-2.2.2.

Nous avons choisi d'utiliser un **rayon de corrélation temporel de 2 jours** car chaque satellite couleur de l'eau a une répétitivité de 2 jours et, dans cette optique, une décroissance exponentielle temporelle a été ajoutée dans le modèle de corrélation. En effet, si on compare les cartes pour les rayons de corrélation temporels de 1 jour et 2 jours, on peut constater les différences encadrées en rouge sur les cartes de l'Annexe A-2.2.2 : il s'agit de zones qui ne contiennent pas de données pour le 26 mars 2003 mais où il y en a le jour précédent (le 25 mars 2003 : cf. cartes Annexe A-2.2.1). Si on ne considère que les

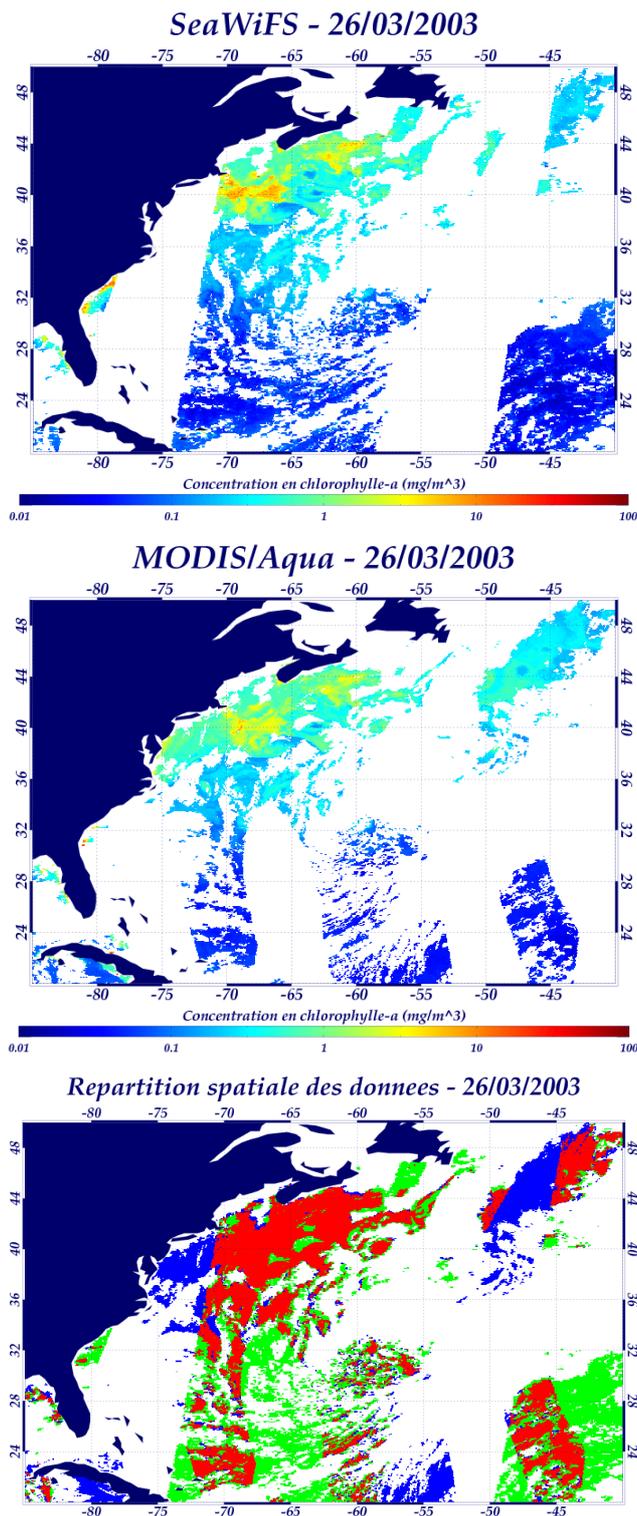


Figure 3.16 – Zone du Gulf Stream au 26 mars 2003 : zone d'étude pour les résultats des tests de sensibilité effectués sur l'analyse objective. En haut, à gauche : carte SeaWiFS. En haut à droite : carte MODIS/Aqua. En bas : carte de répartition spatiale des données (blanc : aucun capteur ; vert : uniquement SeaWiFS ; bleu : uniquement MODIS/Aqua ; rouge : les deux capteurs).

données du jour de traitement, on peut obtenir des résultats peu réalistes qui se rapprochent de la climatologie là où l'algorithme extrapole à partir des observations qu'il possède et qui peuvent se trouver sur les bords de la bulle d'influence. Par contre, si on considère le jour précédent, la répétitivité de 2 jours des satellites fait en sorte que ce phénomène disparaît. Il faut noter que prendre deux jours de données a pour conséquence d'allonger le temps de calcul, mais cela améliore notablement la qualité de la carte combinée.

En ce qui concerne les rayons de corrélation temporels s'élevant à 3 jours et plus, on peut constater un léger lissage, car le nombre d'observations de la bulle d'influence est limité par une valeur maximum pour le calcul de la fonction de corrélation, dont on parle plus loin, partie 2.4.5.

2.4.3 - Centrage des données

On peut montrer que, dans les méthodes de Gauss-Markov, il est possible de définir un estimateur optimal quand l'échantillon est « centré » : dans le contexte des moindres carrés, cela signifie que tout systématisme (biais, étalonnage, dérive d'un instrument, ...) soit être supprimé dans les données, d'où l'importance du retrait d'une « moyenne ». L'équation d'analyse objective s'écrit alors, $\tilde{\theta}$ étant cette valeur moyenne :

$$\theta_{\text{est}}(x) = \tilde{\theta} + \sum_{i=1}^N \sum_{j=1}^N A_{ij}^{-1} C_{xj} (\Phi_{\text{obs}^i} - \tilde{\theta})$$

Trois cas ont été testés. Tout d'abord, aucune moyenne n'est supprimée. Ensuite, la valeur considérée est obtenue par moyenne simple des observations de la bulle d'influence pondérées par leur biais. Enfin, la moyenne dans la bulle d'influence est calculée par la méthode de Bretherton [Bretherton *et al.*, 1976], qui revient à calculer une moyenne optimale (car considérant la matrice de covariance inversée) en tenant compte des erreurs des observations et donc du biais entre les capteurs. L'utilisation de la moyenne de Bretherton entraîne une modification sur le calcul de l'erreur :

$$e^2 = C_{xx} - \sum_{i=1}^N \sum_{j=1}^N C_{xi} C_{xj} A_{ij}^{-1} + \frac{\left(1 - \sum_{i=1}^N \sum_{j=1}^N C_{xi} A_{ij}^{-1}\right)^2}{\sum_{i=1}^N \sum_{j=1}^N A_{ij}^{-1}}$$

On obtient les résultats montrés en Annexe A-2.2.3. Il est clair que le centrage des données par moyenne des observations pondérées par le biais donne de très mauvais résultats à certains endroits (cf. cercles rouges sur notre image de l'Annexe A-2.2.3). Quant aux deux autres, très peu de différences sont visibles ; il s'agit essentiellement des zones pour lesquelles peu d'observations sont disponibles (cf. cercles verts sur notre image de l'Annexe A-2.2.3). Sachant qu'il est préférable, de par la théorie, de retirer une moyenne, nous avons choisi la **moyenne de Bretherton**.

2.4.4 - Nombre minimum d'observations dans la bulle d'influence

Le nombre minimum d'observations dans la bulle d'influence est utile pour éliminer les cas où seuls quelques points sont pris pour le calcul à une position donnée. Quatre configurations de minimum d'observations ont été testées : 1, 5, 10 et 30.

Les résultats sont montrés en Annexe A-2.2.4. Nous avons choisi un **minimum de 5 observations**, ce qui nous a semblé un bon compromis entre interpolation très tendancieuse (cas avec un minimum de 1 point) et limitation à l'extrapolation dans l'intertrace. Ce paramètre a très peu d'influence sur la durée du calcul.

2.4.5 - Nombre maximum d'observations dans la bulle d'influence

Le nombre maximum d'observations dans la bulle d'influence sert à contrôler la durée d'exécution de l'algorithme, puisqu'on se place dans un objectif d'océanographie opérationnelle. En effet, le facteur limitant est l'inversion de la matrice de covariance : plus la matrice comporte d'éléments, plus elle est longue à inverser. Evidemment, à l'inverse, si trop peu de points sont sélectionnés dans la bulle, l'estimation est moins correcte. Nous devons donc trouver le bon compromis entre vitesse de calcul et qualité de l'estimation. Cinq configurations ont été testées : 10, 50, 100, 150 et 300 observations maximum.

Les résultats obtenus sont montrés en Annexe-A-2.2.5. Les temps de calcul sont respectivement de 5 min, 12 min, 22 min, 30 min et 1h30 min (moins il y a de points dans la matrice de covariance, moins le temps de calcul est long). Qualitativement, un maximum de 10 points dans la bulle d'influence donne une carte combinée très bruitée. On peut constater qu'à partir de 50 points, les choses commencent à se stabiliser et à partir de 100 points, les différences entre les résultats deviennent très petites et le temps de calcul s'allonge beaucoup. Nous avons donc décidé de prendre un **maximum de 150 observations** dans la bulle d'influence, ce qui nous a semblé un bon compromis entre temps de calcul et qualité du résultat.

2.4.6 - Sous-échantillonnage des positions d'étude

La partie très gourmande en temps de calcul, dans l'analyse objective, est l'inversion de la matrice de covariance. Pour cela, le logiciel inclut la possibilité de ne pas faire calculer la matrice de covariance en tous les points de la grille des données combinées, mais seulement en certains (ces positions sont appelées « positions d'étude ») ; les coefficients de la matrice inversée sont alors aussi utilisés pour les estimations aux points de grilles environnants la position d'étude (appelés « positions sous-échantillonnées ») : jusqu'à un certain point, cette approximation peut se vérifier. Quatre configurations de sous-échantillonnage des positions d'étude ont été testées : 1 (la matrice de covariance est calculée en tous les points de grille), 3 (la matrice de covariance est calculée tous les 3 points, en longitude comme en latitude), 7 (la matrice de covariance est calculée tous les 7 points, en longitude comme en latitude) et 11 (la matrice de covariance est calculée tous les 11 points, en longitude comme en latitude).

Les résultats obtenus sont montrés en Annexe-A-2.2.5. Les temps de calcul sont respectivement de 2h30, 30 min, 14 min et 9 min (plus la grille de sous-échantillonnage est lâche, moins le temps de calcul est long). Qualitativement, un sous-échantillonnage de 1 ou de 3 donne des données combinées très proches (l'erreur RMS s'élève à 2%), la différence la plus nette étant les contours des zones non estimées. Par contre, à partir de 7, une pixellisation est très largement notable et les erreurs RMS pour les sous-échantillonnages de 7 et 9 sont respectivement de 7% et 13%.

Considérant les aspects temps de calcul et qualité des données combinées, nous avons choisi un sous-échantillonnage des positions d'étude de 3.

2.4.7 - Rayons de corrélation spatiaux

Des rayons de corrélation spatiaux dépend la taille des bulles d'influence parmi lesquelles sont sélectionnées les données pour réaliser les estimations. Leurs valeurs ont été déterminées empiriquement sur les anomalies journalières de SeaWiFS pour l'année 2003 (cf. partie 2.3.5). Nous avons souhaité étudier la sensibilité de l'analyse objective à ce paramètre et pour cela, nous avons testé les configurations suivantes : valeurs 4 fois plus petites que celles estimées empiriquement, valeurs 2 fois plus petites, valeurs 2 fois plus grandes et valeurs 4 fois plus grandes.

Les résultats obtenus sont montrés en Annexe A-2.2.7. Dans un premier temps, on peut constater que, si les rayons de corrélation spatiaux sont sous-estimés, la couverture spatiale du produit combiné est moins importante. D'autre part, une surestimation de ces rayons entraîne peut-être un gain en couverture spatiale, mais aussi un lissage du produit

combiné, ce phénomène de lissage étant limité par le paramètre limitant le nombre d'observations prises en compte dans le calcul de l'estimation (cf. partie 2.4.5).

2.4.8 - Erreur de mesure

L'erreur de mesure est l'un des deux facteurs (avec le biais) permettant à l'algorithme d'analyse objective d'accepter la possibilité d'avoir deux valeurs différentes en un même emplacement. C'est aussi l'un des deux facteurs (toujours avec le biais) imposant l'intervalle dans lequel doit se situer l'estimation. Leurs valeurs ont été déterminées empiriquement par comparaison avec les mesures *in situ* pour l'année 2003 (cf. Chapitre 2-3). Nous avons souhaité étudier la sensibilité de l'analyse objective à ce paramètre et pour cela, nous avons testé les configurations suivantes : valeurs 4 fois plus petites que celles estimées empiriquement, valeurs 2 fois plus petites, valeurs 2 fois plus grandes et valeurs 4 fois plus grandes.

Les résultats obtenus sont montrés en Annexe A-2.2.8. On voit très nettement que ce paramètre est un facteur de lissage du produit combiné. Plus l'erreur de mesure fixée est grande, plus le produit combiné est lissé et, au contraire, plus l'erreur de mesure est faible, plus le produit combiné est bruité. D'où l'importance d'estimer au mieux les erreurs de mesure associées aux capteurs.

2.4.9 - Biais

Le biais agit en complémentarité avec l'erreur de mesure. Il est aussi nécessaire pour « rapprocher » les données de plusieurs capteurs par rapport à un autre, considéré comme le plus fiable. Leurs valeurs ont été déterminées empiriquement par comparaison avec les mesures *in situ* pour l'année 2003 (cf. Chapitre 2-3).

Dans un premier temps, nous avons souhaité étudier la sensibilité de l'analyse objective à ce paramètre et pour cela, nous avons testé les configurations suivantes : valeurs 4 fois plus petites que celles estimées empiriquement, valeurs 2 fois plus petites, valeurs 2 fois plus grandes et valeurs 4 fois plus grandes. Les résultats obtenus sont montrés en Annexe A-2.2.9. Ce n'est pas très net à l'œil nu, mais plus ce paramètre est grand, plus les données sont « tirées » vers les valeurs de l'un ou l'autre des capteurs, en fonction de celui qui a le biais le plus petit.

D'autre part, nous avons aussi effectué deux autres tests en utilisant, du coup, le biais inter-capteurs déterminé au Chapitre 2-4. Nous avons tout d'abord considéré que SeaWiFS était le capteur le plus fiable, donc les données issues de MODIS/Aqua ont été ramenées à la hauteur de celles de SeaWiFS (cf. Annexe A-2.2.9, la figure intitulé « Biais MA-SE »). Puis vice-versa (cf. Annexe A-2.2.9, la figure intitulé « Biais SE-MA »). Dans chacun des cas, les

données du capteur considéré le moins fiable sont « tirées » vers celles du capteur considéré le plus fiable.

Dans nos calculs, nous avons décidé d'utiliser les cartes de biais par rapport aux observations *in situ*. En effet, il nous semble plus cohérent de recalibrer les données par rapport à l'un ou l'autre des capteurs en fonction du meilleur pour chaque localisation : nous avons vu au Chapitre 2-3 que ce n'est pas toujours le même capteur couleur de l'eau qui est le plus fiable, mais que ça dépend de l'endroit où on se place.

2.4.10 - Variance

La variance représente la variabilité possible du champ par rapport au « first-guess ». Leurs valeurs ont été déterminées empiriquement sur les anomalies journalières de SeaWiFS pour l'année 2003 (cf. partie 2.3.5). Nous avons souhaité étudier la sensibilité de l'analyse objective à ce paramètre et pour cela, nous avons testé les configurations suivantes : valeurs 4 fois plus petites que celles estimées empiriquement, valeurs 2 fois plus petites, valeurs 2 fois plus grandes et valeurs 4 fois plus grandes.

Les résultats obtenus sont montrés en Annexe A-2.2.10. A l'instar de l'erreur de mesure (cf. partie 2.4.8), on voit très nettement que ce paramètre est un facteur de lissage du produit combiné. Plus la variance est faible, plus le produit combiné est lissé et, au contraire, plus la variance est grande, plus le produit combiné est bruité. D'où l'importance d'estimer au mieux la variance de champ à cartographier.

2.4.11 - Coefficient de la fonction de corrélation

Le coefficient de la fonction de corrélation joue sur sa pente et donc sur l'influence du voisinage à une localisation donnée : ce coefficient est négatif et, plus il est proche de 0, plus la fonction de corrélation est pentue et donc l'influence du voisinage moindre. Les valeurs des coefficients ont été déterminées empiriquement sur les anomalies journalières de SeaWiFS pour l'année 2003 (cf. partie 2.3.4). Nous avons souhaité étudier la sensibilité de l'analyse objective à ce paramètre et pour cela, nous avons testé les configurations suivantes : valeurs 4 fois plus petites que celles estimées empiriquement, valeurs 2 fois plus petites, valeurs 2 fois plus grandes et valeurs 4 fois plus grandes.

Les résultats obtenus sont montrés en Annexe A-2.2.11. A l'instar de l'erreur de mesure (cf. partie 2.4.8) et de la variance (cf. partie 2.4.10), on voit que ce paramètre est un facteur de lissage du produit combiné. Plus le coefficient est petit (en valeur absolue), moins le voisinage lointain est considéré et donc plus le produit combiné est bruité. Au contraire, plus le coefficient est grand (en valeur absolue), plus le voisinage lointain est considéré et



donc plus le produit combiné est lissé. D'où l'importance d'estimer au mieux le coefficient de la fonction de corrélation.

2.4.12 - Conclusion

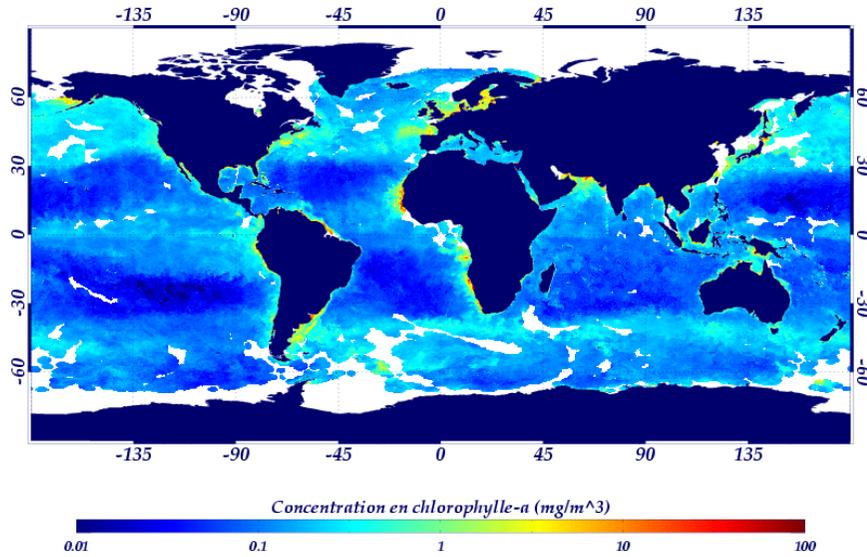
Dans cette partie, nous avons testé la sensibilité de l'algorithme d'analyse objective aux paramètres d'entrée. Nous avons vu que la mauvaise estimation de l'erreur de mesure, la variance du champ, du modèle de corrélation (dans une moindre mesure) et des coefficients associés à la fonction de corrélation avait pour conséquence un lissage ou un bruitage du produit combiné. Nous nous sommes ensuite aperçus que le rayon de corrélation temporel, le centrage des observations dans la bulle d'influence, le nombre minimum d'observations à prendre en compte pour le calcul de la matrice de covariance, les rayons de corrélation spatiaux et le biais influent sur la qualité du produit combiné. Enfin, nous avons montré que le nombre maximum d'observations pris dans la bulle d'influence pour le calcul de la matrice de covariance ainsi que le sous-échantillonnage des positions d'étude (i.e. les positions pour lesquelles on calcule la matrice de covariance) ont une grande influence sur la durée du calcul. Dans tous les cas, nous avons vu que seule une estimation rigoureuse des paramètres d'entrée permet d'obtenir un produit combiné le plus fiable possible.

2.5 - Application aux données de chlorophylle

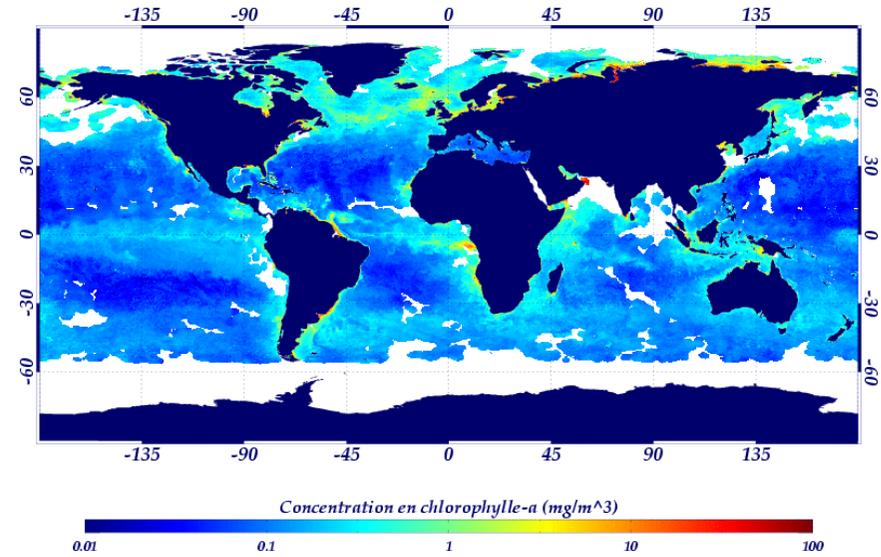
Nous avons appliqué l'algorithme d'analyse objective à la combinaison des données de concentration en chlorophylle-*a* issues de SeaWiFS et MODIS/Aqua en global sur toute l'année 2003. L'application s'est portée sur le logarithme base 10 des valeurs de concentration en chlorophylle-*a*. La résolution des produits combinés est celle des produits SeaWiFS, à savoir 1/12°.

La Figure 3.17 montre le résultat de l'application de l'analyse objective aux log-anomalies pour le 26 mars 2003 (à gauche) et le 13 août 2003 (à droite). La couverture spatiale du produit combiné est de 73.7% et 67.6%, soit une amélioration de 57.7% et 51.8% respectivement, par rapport à la couverture spatiale initiale de SeaWiFS.

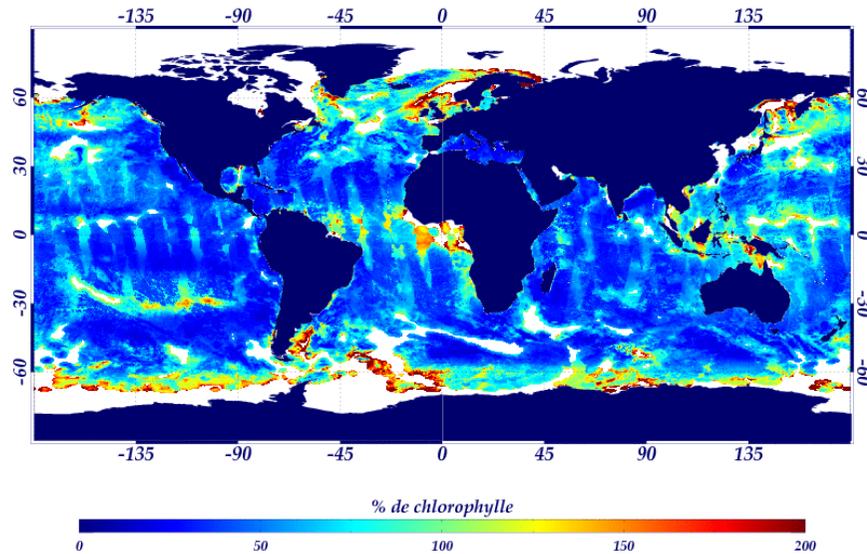
AO - Données combinées - 26/03/2003



AO - Données combinées - 13/08/2003



AO - Erreur - 26/03/2003



AO - Erreur - 13/08/2003

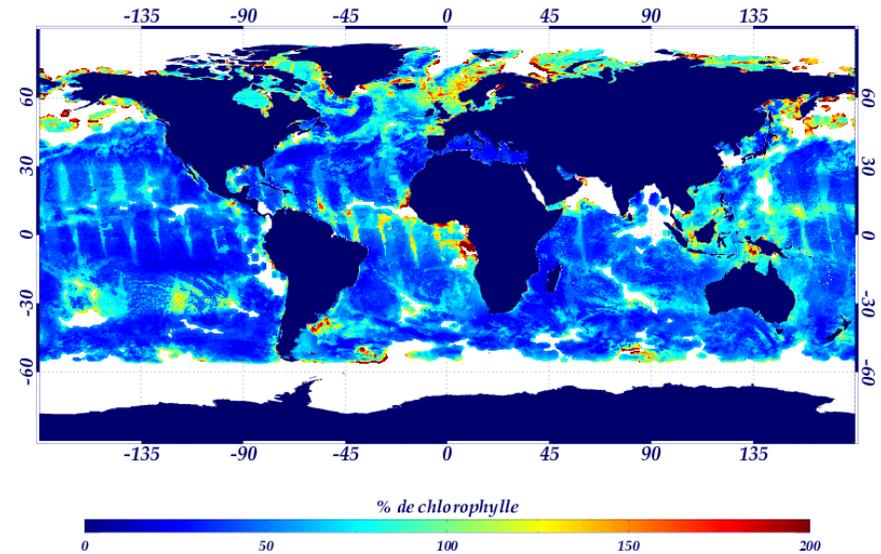


Figure 3.17 – Combinaison des données SeaWiFS et MODIS/Aqua par analyse objective : résultats de l'application au logarithme base 10 des données de concentration en chlorophylle-*a* pour le 26 mars (à gauche) et le 13 août 2003 (à droite). Pour chaque jour, la carte des données combinées est donnée (en haut), ainsi que la carte des erreurs associées (en bas).

Comme pour la moyenne pondérée, les matchups entre les données combinées et les mesures *in situ* ont été déterminés. On obtient les résultats du Tableau 3.3.

Tableau 3.3 – Erreur de mesure et biais (en log et en pourcentage) entre les observations *in situ* et les données combinées par analyse objective.

N	r ²	Erreur RMS		Biais	
		log	%	log	%
8113	82%	0.33	116	0.085	22

La durée du calcul s'élève à 10h30 pour une carte journalière avec un processeur à 3 GHz et 1h20 sur 8 processeurs à 3 GHz.

2.6 - Conclusion

Dans cette partie, l'algorithme de combinaison de données par analyse objective a été décrit. Nous avons vu que l'algorithme doit être appliqué au logarithme base 10 des données, et non aux valeurs brutes directement. Des calculs préliminaires sont nécessaires pour déterminer les paramètres à mettre en entrée du logiciel. Pour l'estimation *a priori* du champ, nous avons choisi une climatologie mensuelle, calculée par filtrage à 1 an par transformée de Fourier des données mensuelles SeaWiFS de concentration en chlorophylle-*a* entre 1997 et 2004. La variance temporelle a été déterminée sur les log-anomalies journalières de SeaWiFS pour l'année 2003, i.e. la différence entre le logarithme base 10 de la carte de chlorophylle journalière et le logarithme base 10 de la climatologie du mois associé. La fonction de corrélation empirique a été déterminée globalement sur une grille 1° x 1° pour l'année 2003 ; elle a été modélisée par une fonction inverse et des coefficients influant sur l'épaisseur du cône formé par la fonction de corrélation (et donc l'influence des observations du voisinage par rapport à un emplacement donné) ont été déterminés pour l'océan global. Les échelles de corrélation spatiales, à partir desquelles sont conditionnées les bulles d'influence, ont été déterminées globalement pour l'année 2003 : les rayons zonaux dépendent de la latitude tandis que les rayons méridionaux sont pratiquement constants sur l'océan global.

L'analyse objective a été appliquée à la combinaison des données de concentration en chlorophylle-*a* issues des capteurs SeaWiFS et MODIS/Aqua à l'échelle du globe pour l'année 2003.

3 - Comparaison entre les deux méthodes

3.1 - Introduction

L'objectif de cette partie est de comparer la moyenne pondérée avec l'analyse objective. Pour cela, trois méthodes ont été mises en œuvre : la comparaison des produits combinés avec les observations *in situ*, la comparaison des tendances temporelles et la comparaison des tendances latitudinales.

3.2 - Comparaison avec les données *in situ*

Le premier dispositif qui vient à l'esprit pour comparer la moyenne pondérée et l'analyse objective est la comparaison des produits combinés correspondant avec les observations *in situ*.

Le Tableau 3.4 récapitule les matchups effectués pour caractériser les erreurs et les biais des capteurs (SeaWiFS et MODIS/Aqua) ainsi que des produits combinés par moyenne pondérée (algorithme retenu, LOG12) et analyse objective. N, dans la 1^{er} ligne pour l'analyse objective, est plus important que pour la moyenne pondérée du fait de la couverture plus étendue des produits combinés. La 2^{er} ligne de matchups pour l'analyse objective correspond à la même configuration, exactement, que pour la moyenne pondérée : les résultats sont donc strictement comparables. On peut voir que les erreurs et biais des produits combinés sont quasiment les mêmes, quelle que soit la méthode de combinaison employée.

Tableau 3.4 – Récapitulatif des matchups. Pour l'analyse objective, la 2^e ligne correspond au même jeu de matchups que la moyenne pondérée.

	N	r ²	Erreur RMS		Biais	
			log	%	log	%
SeaWiFS	2410	82%	0.34	120	0.09	22
MODIS/Aqua	2789	84%	0.31	106	0.1	24
Moyenne pondérée (LOG12)	3230	83%	0.33	116	0.088	22
Analyse objective	8113	82%	0.33	116	0.085	22
	3230	82%	0.32	114	0.082	21

Néanmoins, il faut reconnaître que la distribution spatio-temporelle des mesures *in situ* est limitée. Il aurait été idéal de pouvoir différencier les matchups propres à l'analyse des données satellitaires originales et ceux propres à l'analyse des données combinées. Malheureusement, les matchups ne sont pas suffisants en nombre pour cela ni donc pour comparer les deux méthodes entre elles. Deux autres méthodes ont été développées, qui vont être présentées par la suite.

3.3 - Comparaison des tendances temporelles et latitudinales

3.3.1 - Méthodologie

Ces analyses sont basées sur les travaux réalisés au sein de l'OBPG (Ocean Biology discipline Processing Group) à la NASA pour comparer les capteurs entre eux [Franz, 2005].

L'analyse temporelle consiste à calculer journallement les valeurs moyennes et les écart-types à des échelles régionales et à regarder leur variation dans le temps, sur l'année 2003 en ce qui nous concerne. Les régions choisies (Figure 3.18, à gauche) sont celles utilisées par l'OBPG et détaillées dans le Tableau 3.5. Elles ont la particularité d'être remplies toute l'année, d'être relativement homogènes [Fougnie *et al.*, 2002] et dans des eaux relativement claires.

En complément des tendances temporelles globales et régionales, il est utile de regarder les moyennes latitudinales [Franz, 2005]. Cette analyse consiste à déterminer régionalement les valeurs moyennes et les écart-types pour une journée donnée. Les régions choisies (Figure 3.18, à droite) sont celles utilisées par l'OBPG et détaillées dans le Tableau 3.6. Nous avons réalisé cette évaluation pour nos deux journées de test : le 26 mars 2003 et le 13 août 2003.

Tableau 3.5 – Définitions des régions utilisées pour l'analyse des tendances temporelles.

Région	Latitude minimale	Latitude maximale	Longitude minimale	Longitude maximale
Hawaï	18	19.9	-158.5	-156.5
PacN	15	23	-180	-159.4
PacNW	10	22.7	139.5	165.6
PacSE	-44.9	-20.7	-130.2	-89
AtIN	17	27	-62.5	-44.2
AtIS	-19.9	-9.9	-32.3	-11
IndS	-29.9	-21.2	89.5	100.1

Tableau 3.6 – Définitions des régions utilisées pour l'analyse des tendances latitudinales.

Région	Latitude minimale	Latitude maximale	Longitude minimale	Longitude maximale
PacN50	40	50	-170	-150
PacN40	30	40	-170	-150
PacN30	20	30	-170	-150
PacN20	10	20	-170	-150
PacN10	0	10	-170	-150
PacS10	-10	0	-170	-150
PacS20	-20	-10	-170	-150
PacS30	-30	-20	-170	-150
PacS40	-40	-30	-170	-150
PacS50	-50	-40	-170	-150

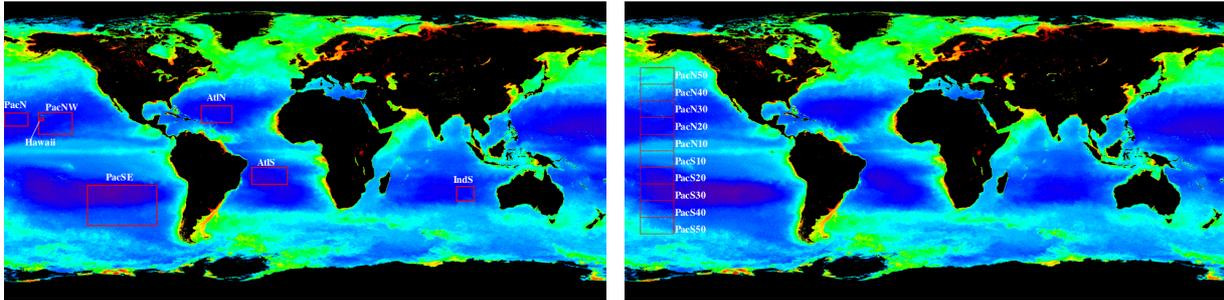
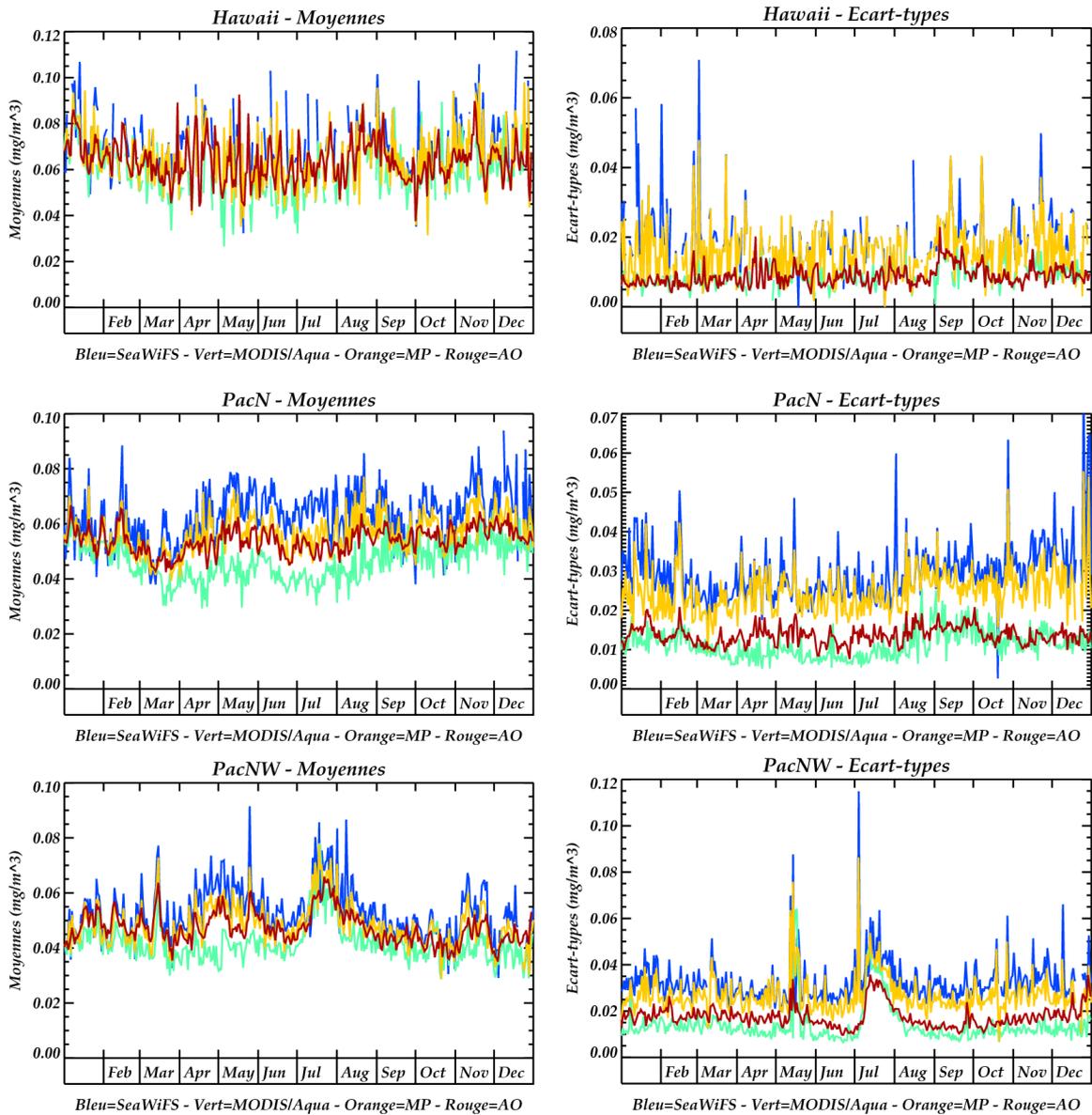


Figure 3.18 – Régions d'étude pour l'analyse temporelle (à gauche) et l'analyse latitudinale (à droite). [Source : Ocean Color Web].

3.3.2 - Résultats

Les résultats de l'analyse temporelle sont montrés sur la Figure 3.19, ci-après :



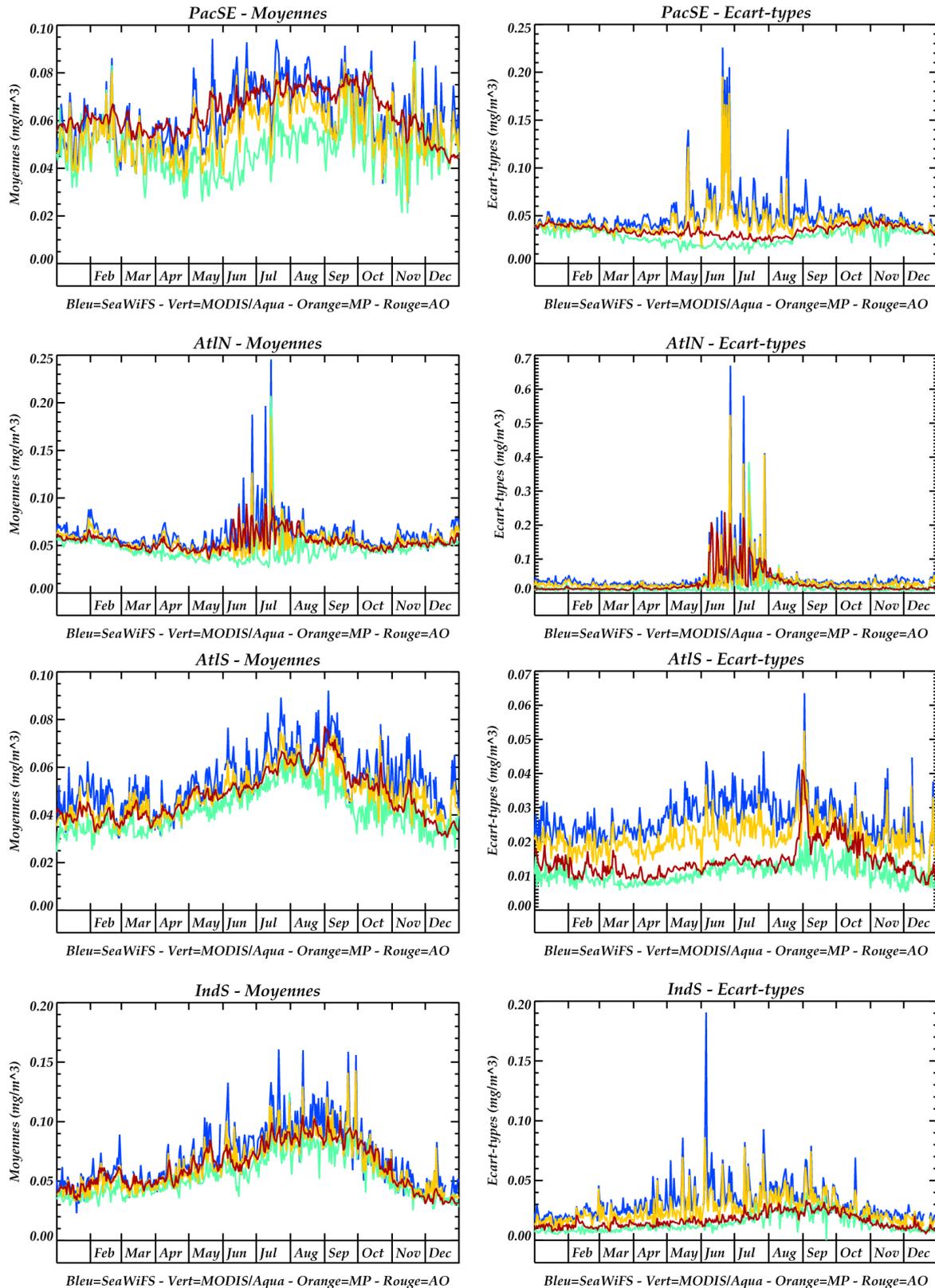


Figure 3.19 – Analyses temporelles régionales. A gauche : les moyennes sur chaque zone. A droite : les écart-types. La courbe correspondant à SeaWiFS est en bleu, celle relative à MODIS/Aqua est en vert, celle correspondant à la moyenne pondérée (algorithme LOG12) en orange et celle relative à l’analyse objective en rouge.

Les graphes de la Figure 3.19 montrent, dans un premier temps, que la concentration en chlorophylle-a issue de SeaWiFS est en moyenne supérieure à celle issue de MODIS/Aqua pour chacune des zones d'observation. Il en est de même pour l'écart-type, indiquant par là qu'à l'intérieur de chaque zone d'étude, la variation des valeurs des pixels de chlorophylle-a SeaWiFS par rapport à la valeur moyenne est plus importante que celle de MODIS/Aqua. On peut noter que ces différences sont moins flagrantes pour la zone d'Hawaï, qui est un point de calibration commun aux satellites [Eplee *et al.*, 2003].

Sur l'année 2003, les moyennes par zone des données combinées résultant de la moyenne pondérée d'une part et de l'analyse objective d'autre part sont comprises entre les moyennes de SeaWiFS et MODIS/Aqua. Les valeurs moyennes de ces deux données combinées sont proches, sauf sur la zone PacSE pour laquelle la moyenne des données combinées par analyse objective est généralement supérieure à celle des données combinées par moyenne pondérée sur toute l'année 2003, ainsi que pour la zone PacN pour laquelle la moyenne des données combinées par moyenne pondérée est supérieure à celle des données combinées par analyse objective.

En ce qui concerne les écart-types sur l'année 2003, on peut remarquer que ceux des données combinées par analyse objective sont systématiquement moins élevés que ceux des données combinées par moyenne pondérée : ceci s'explique par le phénomène de lissage résultant de la technique même de l'analyse objective, qui est une moyenne pondérée optimale d'observations incluses dans un voisinage de la position d'étude.

Les résultats de l'analyse latitudinale pour nos deux journées de test sont montrés Figure 3.20.

Pour chaque jour, on peut tout d'abord constater que les courbes de moyenne et d'écart-type relatives à l'algorithme de moyenne pondérée sont comprises entre les courbes correspondantes de SeaWiFS et de MODIS/Aqua, ce qui est naturel vu la nature du calcul.

En ce qui concerne les courbes relatives à l'analyse objective, elles sont globalement toujours légèrement en dessous des courbes correspondant à la moyenne pondérée. Au niveau de la moyenne, cela signifie que l'analyse objective a supprimé un certain biais dans les données de concentration en chlorophylle-a originales. Quant à l'écart-type, cela signifie que l'analyse objective a plus lissé les données que la moyenne pondérée.

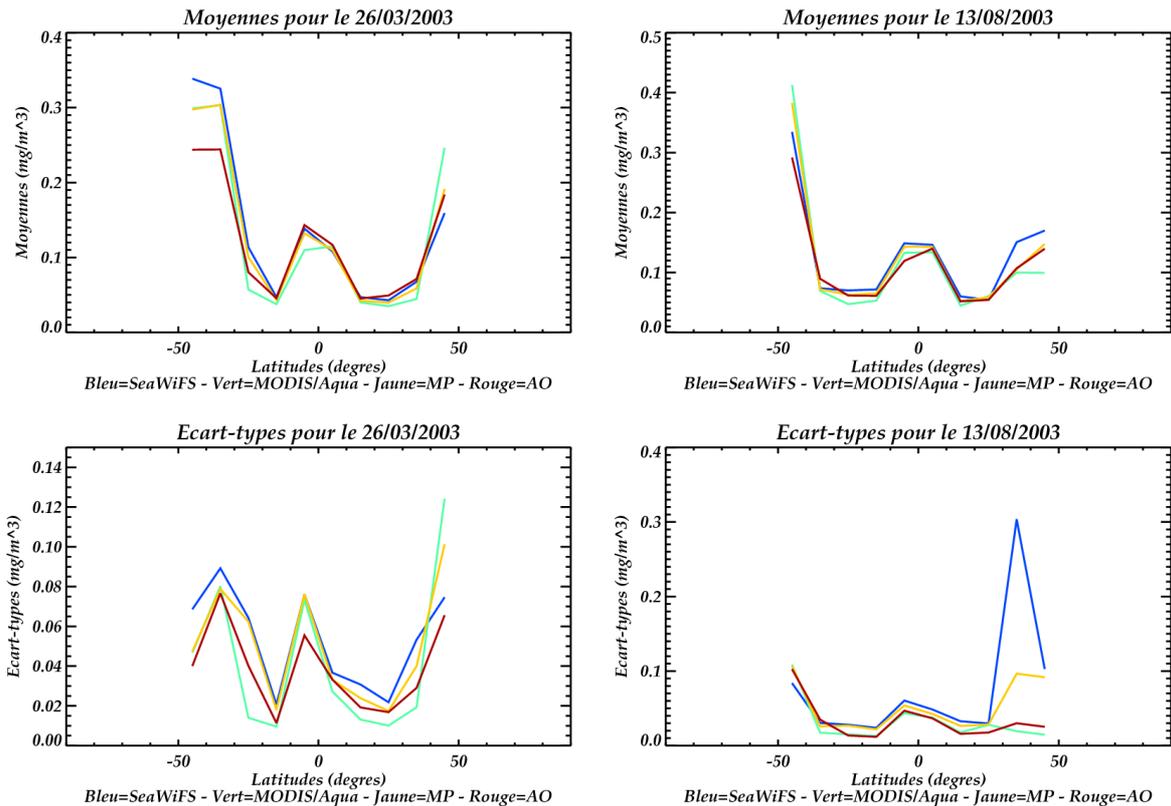


Figure 3.20 – Tendances latitudinales pour le 26 mars 2003 (à gauche) et le 13 août 2003 (à droite). En haut : les moyennes. En bas : les écart-types. La courbe correspondant à SeaWiFS est en bleu, celle relative à MODIS/Aqua est en vert, celle correspondant à la moyenne pondérée (algorithme LOG12) en orange et celle relative à l’analyse objective en rouge.

3.4 - Conclusion

Dans cette partie, nous avons essayé de quantifier au mieux les différences entre les produits combinés par moyenne pondérée et les produits combinés par analyse objective. Le meilleur moyen est de comparer les données combinées aux observations *in situ* et nous avons trouvé des erreurs et des biais quasiment identiques pour les deux jeux de données. Puisque les matchups sont vraiment très clairsemés et rares dans le temps et l’espace, nous avons comparé le comportement de la moyenne et de l’écart-type des données combinées sur des zones particulières. Nous avons vu que la moyenne sur la zone des produits combinés est toujours comprise entre celles des données originales. Pour l’écart-type, celui des données combinées par analyse objective est plus faible que celui des données combinées par moyenne pondérée, confirmant l’effet de lissage de cette première technique.

4 - Conclusion du chapitre

Dans ce chapitre, nous avons présenté et testé deux approches pour la combinaison des données couleur de l'eau pour des objectifs d'océanographie opérationnelle. Ces deux méthodes ont été appliquées sur l'océan global pour la combinaison des données de concentration en chlorophylle-*a* issues des capteurs SeaWiFS et MODIS/Aqua durant l'année 2003.

La première approche est la moyenne pondérée par l'erreur capteur. Bien que souvent considérée comme basique, elle améliore la couverture spatiale en ne prenant en compte que les valeurs satellitaires existantes. Elle est rapide. Sa principale limitation est la discontinuité dans la précision du produit combiné. S'il n'y a qu'une valeur mesurée, la précision de la valeur combinée sera celle du capteur en question. S'il y a deux estimations (i.e. une valeur mesurée pour chacun des capteurs), la précision de la donnée combinée sera améliorée. L'application de cette méthode en global sur l'année 2003 montre que la qualité des données combinées varie très peu entre l'application sur les données brutes et l'application sur le logarithme base 10 des données. Néanmoins, nous avons choisi de sélectionner une application sur le logarithme base 10 des données et un produit combiné sur une grille $1/12^\circ \times 1/12^\circ$ pour des raisons de rapidité d'exécution de l'algorithme.

La seconde approche est l'analyse objective. Etant donné un champ « first-guess » et des paramètres statistiques calculés directement à partir des données satellitaires, cette méthode permet d'interpoler des valeurs là où il n'y en a pas. En conséquence, la couverture spatiale des données combinées s'en trouve améliorée, par rapport à la moyenne pondérée. Son application aux données brutes donne des résultats pauvres voire erronés, en particulier en raison de la distribution lognormale et de la dynamique temporelle de la chlorophylle. L'analyse objective doit donc être appliquée au logarithme base 10 des données.

Apprécier la qualité des produits combinés est difficile. La meilleure façon de le faire est de comparer les données combinées aux mesures *in situ*. Nous avons vu ici que l'erreur RMS et le biais des données combinées par moyenne pondérée diffèrent très peu de ceux des données combinées par analyse objective. Mais, puisque les matchups sont peu nombreux, nous avons effectué des comparaisons de moyennes et d'écart-types dans des zones d'étude choisies par l'OBPG. La moyenne, sur chacune de ces zones, des produits combinés (par moyenne pondérée ou analyse objective) est toujours comprise entre celles des données originales. Pour l'écart-type, celui des données combinées par analyse objective est plus faible que celui des données combinées par moyenne pondérée, confirmant l'effet de lissage de cette première technique.

Chapitre 4

Combinaison de données de
chlorophylle :
Approche par ondelettes

CHAPITRE 4

COMBINAISON DES DONNEES DE CHLOROPHYLLE :

APPROCHE PAR ONDELETTES

Les ondelettes ont trouvé, de par la puissance de leur théorie, des applications dans de nombreux domaines : mathématiques (analyse, probabilités, fractales), traitement du signal (compression, astronomie, sismique), physique (mécanique quantique, turbulence). En effet, l'analyse en ondelettes permet l'étude temps-échelle de signaux dont la variabilité spatiale et / ou temporelle est importante. Nous avons donc envisagé une méthode de combinaison basée sur la transformée en ondelettes.

La théorie générale des ondelettes est tout d'abord rappelée, en particulier leur relation avec l'analyse multi-résolution qui est plus particulièrement adaptée à notre problème. La méthode utilisée pour estimer les données manquantes est ensuite présentée, ainsi que les deux approches envisagées. Puis, les résultats de leur application aux données de concentration en chlorophylle sont montrés. Enfin, l'approche retenue sera comparée aux deux méthodes décrites dans le chapitre précédent : la moyenne pondérée par l'erreur et l'analyse objective.

Sommaire

1 – GENERALITES SUR LA THEORIE DES ONDELETTES	145
1.1 – <i>Introduction par la musique</i>	145
1.2 – <i>Transformée en ondelettes</i>	146
1.3 – <i>Multi-résolution : cas des bases orthonormales</i>	149
1.4 – <i>Ondelettes et bancs de filtres</i>	154
1.5 – <i>Exemple de bases d'ondelettes orthonormales</i>	161
1.6 – <i>Conclusion</i>	170
2 – METHODOLOGIE.....	171
2.1 – <i>Introduction</i>	171
2.2 – <i>Estimation des données manquantes</i>	171
2.3 – <i>Fusion d'images par ondelette</i>	191
2.4 – <i>Application à la combinaison des données de chlorophylle</i>	193
3 – CONCLUSION DU CHAPITRE.....	195

1 - Généralités sur la théorie des ondelettes

1.1 - Introduction par la musique



Figure 4.1 – Les premières notes de « Bohemian Rhapsody », de Queen.

Musicienne depuis longtemps, j'aime introduire musicalement l'idée de la transformée en ondelettes. Un morceau de musique (comme les premières notes de « Bohemian Rhapsody », du groupe Queen, présentées Figure 4.1), représenté habituellement sur une portée musicale, est composé de notes correspondant à une fréquence particulière, dont la durée dépend de la nature de la note (blanche, noire, croche, double croche, etc.) et dont l'instant d'apparition se déduit de la position sur la portée. Une portée musicale correspond donc mathématiquement à un plan temps-fréquences.

L'analyse de Fourier est un outil de base en traitement du signal, indispensable dans de nombreux domaines, mais elle montre vite des limites justifiées dès lors que l'on sort du cadre rigoureux de sa définition : le domaine des signaux stationnaires d'énergie finie, c'est-à-dire les signaux dont la fréquence ne varie pas dans le temps et qui sont de carré sommable. Dans l'analyse de Fourier, tous les aspects temporels (début, fin et durée d'un évènement), bien que présents dans la phase, deviennent illisibles dans le spectre. En particulier, la transformée de Fourier permet « simplement » d'affirmer si un morceau de musique contient telle note et en quelle proportion. Par contre, elle ne permet pas de dire à quel moment du morceau cette note a été jouée. Par conséquent, le spectre de Fourier seul ne permet pas de dissocier deux partitions différentes ayant les mêmes notes.

Afin d'améliorer la localisation temporelle tout en préservant au mieux la localisation fréquentielle, Gabor [1946] a mis au point la première approche de localisation temps-fréquences. Sa technique consiste à découper le signal en différentes plages (ou fenêtres) de largeur fixe ; chaque plage est ensuite étudiée séparément des autres par analyse de

Fourier traditionnelle. Cette méthode est appelée transformée de Fourier à fenêtre glissante. Elle permet donc de retrouver à la fois les fréquences (les notes) et l'information temporelle (l'ordre dans lequel elles sont jouées). Cependant, la taille de la fenêtre étant fixe, cette représentation n'est pas adaptée aux signaux ayant des composantes de longueur très différentes de celle de la fenêtre analysante, de par le principe d'incertitude d'Heisenberg¹. Toujours dans l'exemple du morceau de musique, si la taille de la fenêtre est de la durée d'une noire, la transformée n'est pas adaptée à l'étude des croches et des blanches.

1.2 - Transformée en ondelettes

1.2.1 - Introduction

Pour l'étude haute résolution des signaux sismiques, Morlet [1981] propose une transformée où la taille de la fenêtre est variable, grâce à un paramètre d' « échelle ». Cette nouvelle approche est présentée dans cette partie.

Après avoir brièvement introduit le concept de famille d'ondelettes, nous nous penchons sur les deux types de transformée en ondelettes existants : les transformées en ondelettes continue et discrète. Dans cette partie, nous allons aborder le concept de transformée en ondelettes en commençant par la dimension 1, pour plus de facilité ; la dimension 2 sera introduite en fin de partie. A noter qu'entre la 1D et la 2D, l'idée est similaire, seules quelques formules changent.

1.2.2 - Famille d'ondelettes

Une famille d'ondelettes est une famille de fonctions générées par une fonction d'origine appelée « ondelette mère » et constituées de ses dilatées et translatées :

$$\left\{ \psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \right\}_{a \in \mathbb{R}_+^*, b \in \mathbb{R}}$$

a (positif et non nul) est le **paramètre de dilatation** (ou **paramètre d'échelle**). Il est relié à la fréquence. Plus a est grand, plus l'ondelette est étendue selon l'axe des temps. Par conséquent, les grandes valeurs de a sont associées aux basses fréquences. A l'inverse, les petites valeurs de a sont associées aux hautes fréquences. Le Tableau 4.1 montre la précision en temps et en fréquence en fonction du facteur d'échelle.

¹ Littéralement, ce principe signifie que, si on veut savoir si un signal « possède » la fréquence f_0 , on doit « observer » ce signal sur au moins une période, c'est-à-dire sur un intervalle de temps plus long que $\frac{1}{f_0}$.

Tableau 4.1 – Précision en temps et en fréquence en fonction du facteur d'échelle.

Facteur d'échelle	Fréquence	Précision en temps	Précision en fréquence
Faible	Haute	Elevée	Faible
Elevée	Faible	Faible	Elevée

Quant à b , il s'agit du **paramètre de translation**. Il est relié à la notion de position temporelle.

Une telle famille de fonctions constitue une **base d'ondelettes**.

1.2.3 - Transformée en ondelettes continue

La transformation en ondelettes continue consiste à **quantifier la ressemblance** du signal à analyser $f(t)$ avec l'ondelette $\psi_{a,b}(t)$. Cette grandeur est donc tout simplement le produit scalaire $\langle f, \psi_{a,b} \rangle$, opérateur de projection classique :

$$\forall a > 0, \forall b \in \mathbb{R}, \quad W_f(a,b) = \langle f, \psi_{a,b} \rangle = \int_{\mathbb{R}} f(t) \frac{1}{\sqrt{a}} \overline{\psi\left(\frac{t-b}{a}\right)} dt.$$

b est le point autour duquel l'analyse se fait.

a décide de la finesse de l'analyse. Pour un paramètre d'échelle assez grand, la représentation des coefficients d'ondelettes en fonction de b , le paramètre de translation, donne une représentation de la forme générale de la fonction, comme si l'on regardait la fonction de loin. Par contre, un facteur d'échelle faible correspond à une représentation des aspects et des détails singuliers, comme si l'on regardait la fonction de près. Cette propriété de microscope est très utile pour l'étude de la régularité d'une fonction [Mallat et Hwang, 1992], entre autres choses.

La transformée en ondelettes continue est utilisée pour l'« analyse » (détection de changement de fréquences, ...). Nous allons maintenant nous pencher sur la transformée en ondelettes discrète, orientée essentiellement vers la représentation des signaux qui est un axe qui nous intéresse plus particulièrement pour notre étude.

1.2.4 - Transformée en ondelettes discrète

Dans une transformée en ondelettes discrète, le principe est de prendre un ensemble discret d'échelles $\{a_n\}_{n=1 \dots n_{\max}}$ et, pour chaque échelle a_n , de prendre un ensemble discret de paramètres de translation $\{b_n^i\}_{i=1 \dots N}$, N étant la longueur du signal à analyser.

La façon la plus immédiate pour discrétiser la transformée en ondelettes consiste à échantillonner les échelles a et les paramètres de translation b de manière régulière :

$$a_n = na_0 \quad \text{avec } n = 1 \dots a_{\max}$$
$$b_k^i = i \text{ à l'échelle } k \quad \text{avec } i = 1 \dots N$$

Les propriétés de la transformée en ondelettes continues sont alors conservées, mais le coût de calcul est prohibitif : une telle discrétisation n'est donc, en pratique, que très rarement utilisée.

La discrétisation naturelle pour les échelles consiste à choisir $a = a_0^n$ avec $a_0 > 1$. On parcourt ainsi les fréquences du signal beaucoup plus rapidement, tout en restant précis sur les hautes fréquences (c'est-à-dire les petites échelles). Le choix le plus populaire est de prendre $a_0 = 2$: les échelles sont discrétisées sur des valeurs dyadiques¹. On obtient ainsi la transformée en ondelettes dyadique, qui s'implémente efficacement grâce à l'algorithme à trous [Mallat, 1998]. Cette méthode garde, à chaque échelle, le même nombre de points de discrétisation, et ce, bien que le support² de l'ondelette soit de plus en plus étalé lorsque les échelles croissent.

L'information donnée par un coefficient d'ondelettes correspond à un voisinage d'autant plus grand, dans le domaine spatial, que l'échelle est grande. Il n'est donc pas nécessaire d'utiliser, pour cette transformée, une discrétisation aussi précise aux échelles grossières qu'aux échelles fines. Utiliser des échelles dyadiques et une discrétisation en espace qui dépend de l'échelle rend la transformée en ondelettes beaucoup plus compacte et plus économique du point de vue de la mémoire et du temps de calcul nécessaires pour la calculer. Précisons cependant que seulement pour certaines ondelettes, bien choisies, l'échantillonnage dyadique fournit une véritable base algébrique (i.e. un ensemble complet et linéairement indépendant), permettant alors de représenter les signaux sans perte. Cette méthode s'implémente par l'algorithme pyramidal [Mallat, 1998].

1.2.5 - Conclusion

Nous avons vu les deux grandes catégories de transformées en ondelettes : la transformée en ondelettes continue et la transformée en ondelettes discrète.

Puisque l'objectif de ce travail est d'étudier des images à différentes échelles, c'est la transformée en ondelettes discrète qui se prête le mieux à notre problème. Dans la partie suivante, nous allons appréhender la notion d'analyse multi-résolution, qui utilise la transformée en ondelettes discrète.

¹ En puissances de 2.

² Partie non nulle de l'ondelette.

1.3 - Multi-résolution : cas des bases orthonormales

1.3.1 - Introduction

Nous avons vu que la transformée en ondelettes discrète convenait parfaitement à notre problème d'étude d'images à différentes échelles. De plus, notre intérêt se porte sur la reconstruction exacte du signal à partir de ses coefficients d'ondelettes : une base d'ondelettes orthogonales est alors nécessaire.

Dans cette partie, nous présentons le principe de l'analyse multi-résolution. Après avoir défini cette approche et en avoir donné quelques propriétés, nous introduisons les notions d'approximation / détails, fonctions échelle / ondelette et filtres miroirs conjugués.

1.3.2 - Définition de l'analyse multi-résolution

L'analyse multi-résolution est en quelque sorte la formulation mathématique du phénomène suivant : lorsqu'on regarde un paysage ou un objet, on observe plus ou moins de détails, suivant la distance à laquelle on se trouve, mais le paysage ou l'objet est le même. On peut dire que l'espace dans lequel il est représenté n'est pas le même, d'où cette présence ou absence de détails. Ces espaces ont des propriétés particulières, présentées ci-après.

L'approximation d'une fonction f à la résolution 2^{-j} ($j \in \mathbb{Z}$) est définie comme étant une grille d'échantillons fournissant les moyennes locales de f sur un voisinage d'une taille proportionnelle à 2^j . Une approximation multi-résolution est ainsi composée de grilles d'approximation imbriquées les unes dans les autres. Plus formellement, l'approximation d'une fonction à la résolution 2^{-j} est définie comme sa projection orthogonale sur un espace $V_j \subset L^2(\mathbb{R})$. L'espace V_j regroupe donc toutes les approximations possibles de f à la résolution 2^{-j} . La projection orthogonale de f est la fonction $f_j \in V_j$ qui minimise $\|f - f_j\|$.

Pour éviter toute confusion, notons qu'un paramètre d'échelle 2^j est l'inverse de la résolution 2^{-j} .

Une suite $(V_j)_{j \in \mathbb{Z}}$ de sous-espaces fermés de $L^2(\mathbb{R})$ est une approximation multi-résolution si les propriétés suivantes sont satisfaites [Mallat, 1989, 1998] :

☞ $\forall (j, k) \in \mathbb{Z}^2, f(t) \in V_j \Leftrightarrow f(t - 2^j k) \in V_j$: V_j est invariant par toute translation proportionnelle à l'échelle 2^j . Cet espace peut donc être assimilé à une grille uniforme d'intervalles espacés de 2^j et qui caractérise l'approximation du signal à la résolution 2^{-j} .

- ☞ $\forall j \in \mathbb{Z}, V_{j+1} \subset V_j$: une approximation à la résolution 2^{-j} contient toute l'information nécessaire pour calculer une approximation à une résolution plus grossière $2^{-(j+1)}$.
- ☞ $\forall j \in \mathbb{Z}, f(t) \in V_j \Leftrightarrow f\left(\frac{t}{2}\right) \in V_{j+1}$: dilater une fonction de V_j par 2 grossit les détails par 2 et cela définit l'approximation à la résolution plus grossière $2^{-(j+1)}$.
- ☞ $\lim_{j \rightarrow +\infty} V_j = \bigcap_{j=-\infty}^{+\infty} V_j = \{0\}$: plus la résolution 2^{-j} est petite (i.e. tend vers 0), plus on perd les détails de f et donc plus l'approximation tend vers 0.
- ☞ $\lim_{j \rightarrow -\infty} V_j = \overline{\bigcup_{j=-\infty}^{+\infty} V_j} = L^2(\mathbb{R})$: plus la résolution 2^{-j} est grande, plus l'approximation converge vers le signal original.

1.3.3 - Approximation et fonction d'échelle

L'approximation de f à la résolution 2^{-j} est définie comme étant le résultat de sa projection orthogonale sur V_j . On peut montrer qu'il existe une fonction ϕ telle que la famille

$\left\{ \phi_{j,n}(t) = \frac{1}{\sqrt{2^j}} \phi\left(\frac{t-2^j n}{2^j}\right) \right\}_{n \in \mathbb{Z}}$ des fonctions dilatées et translatées de ϕ est une base

orthonormale de V_j pour tout $j \in \mathbb{Z}$ [Mallat, 1998]. Une telle fonction ϕ est appelée **fonction d'échelle** ou fonction père.

La projection orthogonale de f sur V_j est obtenue par la relation :

$$P_{V_j}(f) = \sum_{n=-\infty}^{+\infty} \langle f, \phi_{j,n} \rangle \phi_{j,n}$$

Le produit scalaire $a_j[n] = \langle f, \phi_{j,n} \rangle$ fournit une approximation discrète à l'échelle 2^j . Il peut s'écrire sous forme d'un produit de convolution :

$$a_j[n] = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{2^j}} \phi\left(\frac{t-2^j n}{2^j}\right) dt = f * \bar{\phi}_j(2^j n)$$

où $\bar{\phi}_j(t) = 2^{-j/2} \phi(2^{-j}t)$. On peut montrer que l'approximation discrète $a_j[n]$ est le résultat d'un filtrage passe-bas de f échantillonné à 2^j [Mallat, 1998].

Remarque :

Il y a plusieurs possibilités pour normaliser la base $\{\phi_{j,n}\}_{n \in \mathbb{Z}}$. Dans le cas présenté plus haut, c'est la longueur de la fonction, c'est-à-dire l'intégrale du carré de la fonction, qui est normalisée à 1 : on a alors $\{\phi_{j,n}\}_{n \in \mathbb{Z}} = \{2^{-j/2} \phi(2^{-j}t - n)\}_{n \in \mathbb{Z}}$. On peut aussi faire en sorte que ce soit la fonction qui soit normalisée à 1 et ainsi $\{\phi_{j,n}\}_{n \in \mathbb{Z}} = \{2^{-j} \phi(2^{-j}t - n)\}_{n \in \mathbb{Z}}$.

1.3.4 - Filtrés miroirs conjugués

Une approximation multi-résolution est entièrement caractérisée par la fonction d'échelle ϕ qui génère la base orthogonale de chaque espace V_j .

Une analyse multi-résolution a la propriété $\forall j \in \mathbb{Z}, V_{j+1} \subset V_j$ (cf. 1.3.1) donc, en particulier, $2^{-1/2} \phi(2^{-1}t) \in V_1 \subset V_0$. Puisque $\{\phi(t-n)\}_{n \in \mathbb{Z}}$ est une base orthonormale de V_0 , on peut écrire la décomposition suivante :

$$\boxed{\frac{1}{\sqrt{2}} \phi\left(\frac{t}{2}\right) = \sum_{n=-\infty}^{+\infty} h[n] \phi(t-n)} \text{ avec } h[n] = \left\langle \frac{1}{\sqrt{2}} \phi\left(\frac{t}{2}\right), \phi(t-n) \right\rangle.$$

Cette équation est appelée **équation d'échelle**. La suite $h[n]$ est interprétée comme la réponse impulsionnelle d'un filtre discret. On peut montrer que ce filtre est un **filtre miroir conjugué** [Mallat, 1989].

Remarque 1 :

Un changement de variable simple ramène à l'autre équation d'échelle que l'on retrouve souvent dans la littérature :

$$\boxed{\frac{1}{\sqrt{2}} \phi(t) = \sum_{n=-\infty}^{+\infty} h[n] \phi(2t-n)}$$

Selon les cas, l'une des deux équations est plus adaptée que l'autre.

Remarque 2 :

On a : $\sum_{k=0}^{N-1} (h[k])^2 = 1$, i.e. $\sum_{k=0}^{N-1} h[k] = \sqrt{2}$ si c'est la longueur de la fonction, c'est-à-dire l'intégrale du carré de la fonction, qui est normalisée à 1. C'est ce que l'on nomme la « norme L2 ». Si c'est la fonction qui est normalisée à 1, on a $\sum_{k=0}^{N-1} h[k] = 1$; il s'agit alors de la « norme L1 ».

Remarque 3 :

Une fonction f est dite « à support compact » pour signifier qu'elle est nulle en dehors d'une partie compacte, i.e. $f(t)=0$ pour $t \in \mathbb{R} - S_c$ où S_c est un intervalle « bloc ». La fonction d'échelle ϕ est à support compact si, et seulement si, son filtre associé h est à support fini, c'est-à-dire tel que $h[n]=0$ pour $n \notin [0, N-1]$: on parle alors de filtre à Réponse Impulsionnelle Finie (RIF). Leurs supports sont alors identiques [Mallat, 1998].

Remarque 4 :

Il est aussi possible de déterminer la fonction d'échelle ϕ associée au filtre h . Il existe deux approches pour résoudre ce problème.

La première méthode est l'algorithme en cascade. Il s'agit d'itérer à partir d'une

fonction rectangulaire. On suppose que $\phi^{(0)}(t) = \begin{cases} 1 & \text{si } t \in [0,1] \\ 0 & \text{sinon} \end{cases}$. On a alors :

$$\phi^{(i+1)}(t) = \sqrt{2} \sum_{n=0}^{N-1} h[n] \phi^{(i)}(2t-n)$$

L'itération converge vers la fonction d'échelle : $\lim_{i \rightarrow +\infty} \phi^{(i)}(t) = \phi(t)$.

La seconde approche est récursive. Dans un premier temps, il faut déterminer les valeurs de $\phi(t)$ pour des valeurs entières de t . Pour cela, on utilise la relation :

$$\phi(n) = \sqrt{2} \sum_{k=0}^{N-1} h[k] \phi(2n-k)$$

Dans un second temps, on calcule les valeurs de $\phi(t)$ pour des valeurs demi-entières de t .

Pour cela, on utilise la relation : $\phi\left(\frac{n}{2}\right) = \sqrt{2} \sum_{k=0}^{N-1} h[k] \phi(n-k)$. Puis on calcule les valeurs de $\phi(t)$

pour des valeurs de t proportionnelles à $\frac{1}{4}$ et ainsi de suite. Ainsi, on obtient les

échantillons de la fonction d'échelle à tous les points répartis dyadiquement $t = \frac{n}{2^i}$.

1.3.5 - Espace des détails et ondelettes

Nous venons de voir que les bases $\{\phi_{j,n}\}_{n \in \mathbb{Z}} = \{2^{-j/2} \phi(2^{-j}t-n)\}_{n \in \mathbb{Z}}$ forment une base orthonormale de V_j pour tout $j \in \mathbb{Z}$. Si ces bases sont adaptées pour les problèmes d'approximation, elles ne possèdent pas de propriétés facilitant la détection de singularités dans un signal ; en revanche, les détails que l'on perd en passant d'une résolution $2^{-(j-1)}$ à

une résolution plus grossière et plus « lisse » 2^{-j} sont les composantes hautes fréquences du signal.

Soit W_j l'espace contenant ces détails. W_j est le complément orthogonal de V_j dans V_{j-1} : $V_{j-1} = V_j \oplus W_j$ et $V_j \perp W_j$. Par conséquent, la projection orthogonale de f sur V_{j-1} peut être décomposée comme la somme des projections orthogonales sur V_j et W_j :

$$P_{V_{j-1}}(f) = P_{V_j}(f) + P_{W_j}(f).$$

Le complément $P_{W_j}(f)$ fournit les détails de f qui apparaissent à l'échelle 2^{j-1} mais qui disparaissent à l'échelle plus grossière 2^j .

De même que pour la fonction échelle, on a $\forall j \in \mathbb{Z}, W_j \subset V_{j-1}$ donc, en particulier, $W_1 \subset V_0$. Si ψ est une fonction de W_0 , il existe une suite $\{g[n]\}_{n \in \mathbb{Z}}$ telle que :

$$\frac{1}{\sqrt{2}} \psi\left(\frac{t}{2}\right) = \sum_{n=-\infty}^{+\infty} g[n] \phi(t-n) \text{ avec } g[n] = \left\langle \frac{1}{\sqrt{2}} \psi\left(\frac{t}{2}\right), \phi(t-n) \right\rangle.$$

Cette équation est appelée **équation d'ondelette**.

En comprimant ou en dilatant ψ , on peut alors construire des bases orthonormales des autres espaces de détails :

$$\left\{ \psi_{j,n}(t) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{t-2^j n}{2^j}\right) \right\}_{n \in \mathbb{Z}} \text{ est une base de } W_j \text{ pour } j \in \mathbb{Z}.$$

Une telle fonction ψ est appelée **fonction d'ondelette** ou fonction mère.

Remarque 1 :

La somme des coefficients du filtre ondelette est nulle : $\sum_{k=0}^{N-1} g[k] = 0$.

Remarque 2 :

De même que pour la fonction d'échelle (cf. partie 1.3.3), il est possible de déterminer la fonction d'ondelette ψ à partir des coefficients du filtre associé g .

1.3.6 - Conclusion

Nous avons vu dans cette partie les propriétés de l'analyse multi-résolution qui permet d'approximer un signal à différentes résolutions et de déterminer les détails associés. L'approximation regroupe les basses fréquences du signal d'origine ; le filtre passe-bas utilisé est un filtre miroir quadratique calculé à partir d'une fonction d'échelle. Les détails sont

les hautes fréquences du signal à une résolution donnée ; le filtre passe-haut utilisé est calculé à partir d'une fonction d'ondelette.

Après la théorie, la pratique ! Dans la partie suivante, nous allons voir comment s'implémente l'analyse multi-résolution.

1.4 - Ondelettes et bancs de filtres

1.4.1 - Introduction

Dans cette partie, nous abordons le principe de la transformée en ondelettes orthogonale rapide, dont l'algorithme est dérivé des propriétés de l'analyse multi-résolution considérées dans la partie précédente. Puis nous déterminons les conditions sur les filtres dans l'objectif d'une reconstruction parfaite du signal à partir des coefficients d'ondelette. Le cas des signaux finis est ensuite étudié. Enfin, l'algorithme est étendu à la dimension 2.

1.4.2 - Algorithme pyramidal ou transformée en ondelettes orthogonale rapide

La transformée en ondelettes rapide décompose successivement chaque approximation $P_{V_j}(f)$ en une approximation plus grossière $P_{V_{j+1}}(f)$ plus les coefficients d'ondelettes (i.e. de détails) $P_{W_{j+1}}(f)$. Dans l'autre sens, la reconstruction à partir des coefficients d'ondelette retrouve chaque $P_{V_j}(f)$ à partir des $P_{V_{j+1}}(f)$ et des $P_{W_{j+1}}(f)$.

Puisque $\{\phi_{j,n}\}_{n \in \mathbb{Z}}$ et $\{\psi_{j,n}\}_{n \in \mathbb{Z}}$ sont des bases orthonormales de V_j et W_j , la projection sur ces espaces s'écrit :

$$a_j[n] = \langle f, \phi_{j,n} \rangle \text{ et } d_j[n] = \langle f, \psi_{j,n} \rangle.$$

Mallat [1998] a montré que ces coefficients peuvent être calculés à l'aide d'une cascade de convolutions discrète et de sous-échantillonnages :

☞ pour la décomposition, notant $\bar{x}[n] = x[-n]$:

$$a_{j+1}[p] = \sum_{n=-\infty}^{+\infty} h[n-2p] a_j[n] = a_j * \bar{h}[2p] \text{ et } d_{j+1}[p] = \sum_{n=-\infty}^{+\infty} g[n-2p] a_j[n] = a_j * \bar{g}[2p]$$

☞ pour la reconstruction, notant $\check{x}[n] = \begin{cases} x[p] & \text{si } n = 2p \\ 0 & \text{si } n = 2p + 1 \end{cases}$:

$$a_j[p] = \sum_{n=-\infty}^{+\infty} h[2p-n] a_{j+1}[n] + \sum_{n=-\infty}^{+\infty} g[2p-n] d_{j+1}[n] = \check{a}_{j+1} * h[p] + \check{d}_{j+1} * g[p]$$

a_{j+1} et d_{j+1} sont calculés en prenant tous les coefficients pairs résultant de la convolution de a_j avec \bar{h} et \bar{g} respectivement, comme le montre la Figure 4.2. Le filtre \bar{h} supprime les

hautes fréquences des a_j tandis que \bar{g} est un filtre passe-haut qui collecte les hautes fréquences restantes. La reconstruction est une interpolation qui met des zéros pour agrandir a_{j+1} et d_{j+1} et filtre les signaux ainsi obtenus (Figure 4.2).

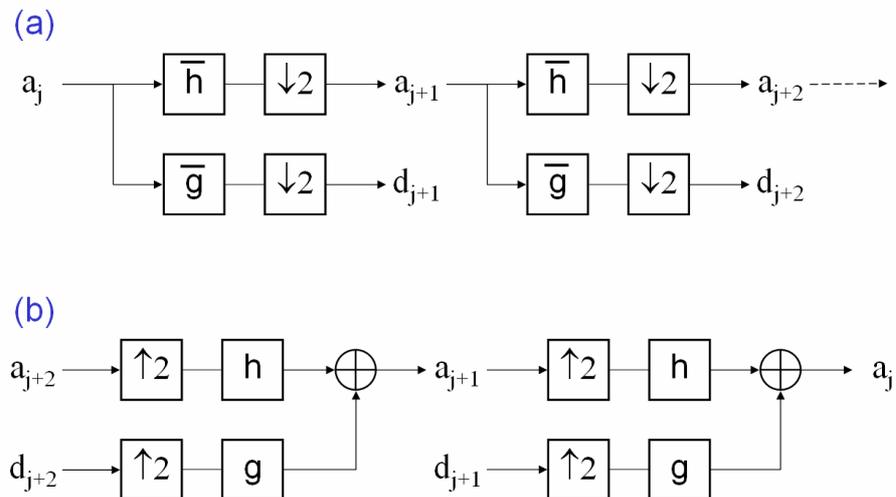


Figure 4.2 – (a) La transformée en ondelettes rapide se calcule avec une cascade de filtrages avec \bar{h} et \bar{g} suivis d'un sous-échantillonnage par un facteur 2. (b) La transformée en ondelettes rapide inverse reconstruit progressivement chaque a_j en insérant des zéros entre les échantillons de a_{j+1} et d_{j+1} , filtrant et ajoutant les sorties.

Le sous-échantillonnage par 2 fait que l'approximation a_{j+1} a deux fois moins de points que l'approximation a_j , d'où le nom d'algorithme pyramidal.

1.4.3 - Cas des signaux finis ou problèmes de bords

Introduction

Qu'il s'agisse de décomposer un signal ou bien de le reconstruire, nous sommes généralement amenés à filtrer des signaux de longueur finie. L'opération de filtrage consiste à calculer pour chaque échantillon une moyenne pondérée sur un nombre de voisins égal à la réponse impulsionnelle du filtre. Ainsi, partant d'un signal $x[n]$ de N échantillons, on obtient le signal $y[n]$ filtré par $A = [a_0 \ a_1 \ \dots \ a_{M-1}]$ de la manière suivante :

$$y[n] = \sum_{k=0}^{M-1} a_k \cdot x[n - k + \text{offset}]$$

M est la longueur de la réponse impulsionnelle du filtre choisi et *offset* un entier relatif définissant le centrage du filtre A . Si $\text{offset} = -M/2$, le filtre est dit **centré**.

On voit tout de suite que l'opération de filtrage pose un problème aux extrémités du signal, c'est-à-dire lorsque la somme $n - k + \text{offset}$ est soit inférieure à 0, soit supérieure à $N - 1$: les échantillons n'existent pas !

Différentes techniques très compliquées ont été envisagées pour pallier à cet inconvénient [Cohen *et al.*, 1993]. Ces méthodes de construction fonctionnent en théorie mais ne sont pas entièrement satisfaisantes d'un point de vue pratique. Aussi, il est souvent préférable d'utiliser des méthodes simples basées sur l'extension des signaux de part et d'autre de son intervalle de définition $[0, N - 1]$: c'est ce qu'on appelle le **padding**. Ces extensions sont réalisées à chaque itération de la transformée en ondelettes directe ou inverse.

Quelques méthodes sont brièvement précisées dans cette partie. Le lecteur pourra se référer à Strang et Nguyen [1996] pour des compléments d'information.

Zero-padding

Cette approche est naïve mais simple à mettre en œuvre : elle consiste tout simplement à supposer que tous les échantillons sont nuls en dehors de l'intervalle de définition du signal :

$$x[n < 0] = x[n > N - 1] = 0.$$

L'inconvénient de cette méthode est la discontinuité occasionnée sur les bords : si les valeurs du signal à ses extrémités sont très différentes de 0, l'opération introduit des discontinuités qui se propagent à chaque itération. On risque alors d'avoir une très forte réponse sur les bords, artificielle de surcroît, et très peu d'information dans les zones d'intérêt.

« Smooth padding » ou remplissage lisse d'ordre 0

Cette méthode stipule que le signal peut être reconstruit en dehors de son support original par une simple extrapolation constante. En pratique, on répète la 1^e valeur en amont du signal et la dernière en aval :

$$\begin{aligned} x[-1] &= x[-2] = \dots = x[0] \\ x[N] &= x[N+1] = \dots = x[N] \end{aligned}$$

Périodisation

On suppose que le signal, dans son ensemble, se reproduit identique à lui-même :

$$\begin{aligned} x[-1] &= x[N-1] & x[-2] &= x[N-2] & \dots \\ x[N] &= x[0] & x[N+1] &= x[1] & \dots \end{aligned}$$

L'avantage de cette méthode est que cette périodisation provoque, dans le domaine spectral, une discrétisation du spectre qui ne gêne pas la transformée en ondelettes. C'est la raison pour laquelle la périodisation est souvent utilisée pour traiter les effets de bords sur des signaux de courte durée. Par contre, si le signal est très différent à chaque extrémité, des discontinuités importantes peuvent apparaître.

Symétrisation

Cette méthode consiste à calculer le signal « miroir » par rapport aux bords :

$$\begin{aligned} x[-1] &= x[1] & x[-2] &= x[2] & \dots \\ x[N] &= x[N-2] & x[N+1] &= x[N-3] & \dots \end{aligned}$$

Cette opération introduit moins de discontinuités lorsque l'amplitude du signal varie sensiblement entre ses deux extrémités. Cependant, son effet sur le plan spectral peut être complexe. En particulier, cette méthode crée artificiellement des discontinuités dans la dérivée première au niveau des bords.

Conclusion

Pour traiter les effets de bords, l'extension du signal de part et d'autre de son intervalle de définition est généralement mise en œuvre. Sur les quatre méthodes présentées, le zero-padding, le smooth-padding d'ordre 0 et la symétrisation impliquent le calcul de quelques coefficients supplémentaires à chaque étape de la décomposition du signal pour obtenir une reconstruction parfaite. Par contre, ce n'est pas le cas de la périodisation, car la base d'ondelettes alors constituée est toujours orthogonale [Mallat, 1989]. Comme nous le verrons par la suite, cette caractéristique d'invariabilité du nombre des coefficients d'ondelettes est nécessaire pour la méthode développée : nous utiliserons donc l'extension du signal par périodisation pour la suite de notre étude.

1.4.4 - Formulation matricielle

Les formules définies précédemment pour la transformée en ondelettes rapide s'implémentent très facilement et très rapidement à l'aide de matrices.

Soient :

$X = [\dots x[-1] \ x[0] \ x[1] \ \dots]^T$, le vecteur colonne des échantillons du signal d'entrée,

$A = [\dots a[-1] \ a[0] \ a[1] \ \dots]^T$, le vecteur colonne des coefficients d'approximation,

$D = [\dots d[-1] \ d[0] \ d[1] \ \dots]^T$, le vecteur colonne des coefficients de détails.

Ainsi, la décomposition du signal X se traduit par : $\begin{bmatrix} A \\ D \end{bmatrix} = \begin{bmatrix} PB \\ PH \end{bmatrix} X$ où :

$$\text{PB} = \begin{bmatrix} \dots & \dots & \dots & \dots & \dots & \dots \\ h[0] & h[1] & \dots & \dots & \dots & \dots \\ \dots & h[-1] & h[0] & h[1] & \dots & \dots \\ \dots & \dots & \dots & h[-1] & h[0] & h[1] \\ \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix}$$

est la matrice de filtrage passe-bas

(convolution avec le filtre h) et de sous-échantillonnage par 2,

$$\text{PH} = \begin{bmatrix} \dots & \dots & \dots & \dots & \dots & \dots \\ g[0] & g[1] & \dots & \dots & \dots & \dots \\ \dots & g[-1] & g[0] & g[1] & \dots & \dots \\ \dots & \dots & \dots & g[-1] & g[0] & g[1] \\ \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix}$$

est la matrice de filtrage passe-haut

(convolution avec le filtre g) et de sous-échantillonnage par 2.

La relation de reconstruction se déduit simplement de la relation de décomposition :

$$X = \begin{bmatrix} \text{PB} \\ \text{PH} \end{bmatrix}^{-1} \begin{bmatrix} A \\ D \end{bmatrix}.$$

Comme les bases d'ondelettes choisies sont orthogonales, on a la simplification :

$$\begin{bmatrix} \text{PB} \\ \text{PH} \end{bmatrix}^{-1} = \begin{bmatrix} \text{PB} \\ \text{PH} \end{bmatrix}^T.$$

1.4.5 - Cas de la dimension 2

La construction d'ondelettes orthogonales à partir de l'analyse multi-résolution s'étend aisément à la dimension 2. Etant donnée une analyse multi-résolution $(V_j)_{j \in \mathbb{Z}}$ de $L^2(\mathbb{R})$, on définit une famille de sous-espaces $(v_j)_{j \in \mathbb{Z}}$ de $L^2(\mathbb{R}^2)$ comme étant le produit tensoriel : $v_j = V_j \otimes V_j$. Alors $(v_j)_{j \in \mathbb{Z}}$ est une multi-résolution de $L^2(\mathbb{R}^2)$ [Mallat, 1998]. Une analyse multi-résolution ainsi définie est qualifiée de **séparable**.

Si ϕ est la fonction d'échelle associée à l'analyse multi-résolution $(V_j)_{j \in \mathbb{Z}}$, O_j les espaces de détails et ψ l'ondelette mère associée, on peut construire la fonction d'échelle Φ associée à $(v_j)_{j \in \mathbb{Z}}$ par :

$$\forall (x,y) \in \mathbb{R}^2, \quad \Phi(x,y) = \Phi(x)\Phi(y).$$

Si on note θ_j le sous-espace supplémentaire de v_j dans v_{j-1} alors :

$$\theta_j = \theta_j^1 \oplus \theta_j^2 \oplus \theta_j^3 \text{ avec } \theta_j^1 = V_j \otimes O_j, \theta_j^2 = O_j \otimes V_j \text{ et } \theta_j^3 = O_j \otimes O_j.$$

On voit que chaque espace de détails est en fait composé de 3 autres espaces. Chacun de ces 3 espaces θ_j^k correspond à une ondelette mère ψ^k définie selon :

$$\begin{aligned}\Psi^1(x, y) &= \phi(x) \psi(y) \\ \Psi^2(x, y) &= \psi(x) \phi(y) . \\ \Psi^3(x, y) &= \psi(x) \psi(y)\end{aligned}$$

Ces ondelettes mères sont telles que $\Psi_{j,k,l}^i(x, y) = 2^{-j/2} \Psi^i(2^{-j}x - k, 2^{-j}y - l)$ et on peut montrer que $\{\Psi_{j,k,l}^1, \Psi_{j,k,l}^2, \Psi_{j,k,l}^3\}_{(k,l) \in \mathbb{Z}^2}$ est une base orthonormale de θ_j [Mallat, 1998].

Les 3 ondelettes Ψ^1 , Ψ^2 et Ψ^3 sont respectivement des ondelettes horizontale, verticale et diagonale. Elles servent donc à détecter les détails dans des directions privilégiées.

Pour chaque échelle 2^j , une image I se décompose selon la formulation :

$$a_j[n] = \langle I, \Phi_{j,n} \rangle \text{ et } d_j^k[n] = \langle I, \Psi_{j,n}^k \rangle \text{ avec } k = \{1,2,3\}.$$

Les formules de décomposition s'écrivent donc :

$$\begin{aligned}a_{j+1}[n] &= a_j * \overline{hh}[2n] \\ d_{j+1}^1[n] &= d_j^1 * \overline{hg}[2n] \\ d_{j+1}^2[n] &= d_j^2 * \overline{gh}[2n] \\ d_{j+1}^3[n] &= d_j^3 * \overline{gg}[2n]\end{aligned}$$

On peut montrer que les convolutions de dimension 2 séparables peuvent être factorisées en 2 convolutions de dimension 1, appliquées sur les lignes puis sur les colonnes d'une image. La Figure 4.3 montre que ces 4 convolutions 2D sont calculées avec 6 convolutions 1D. Les lignes de a_j sont tout d'abord convoluées avec \overline{h} et \overline{g} puis sous-échantillonnées par 2. Les colonnes de ces 2 images de sortie sont ensuite convoluées avec \overline{h} et \overline{g} respectivement puis sous-échantillonnées, donnant 4 images a_{j+1} , d_{j+1}^1 , d_{j+1}^2 et d_{j+1}^3 .

Posons \tilde{I} l'image deux fois plus grande que I obtenue à partir de cette dernière en insérant des lignes et des colonnes de 0 entre 2 lignes et 2 colonnes consécutives. Similairement au cas 1D, l'approximation a_j est retrouvée à partir de l'approximation plus grossière a_{j+1} et des coefficients d'ondelette d_{j+1}^k :

$$a_j[n] = \tilde{a}_{j+1} * hh[n] + \tilde{d}_{j+1}^1 * gh[n] + \tilde{d}_{j+1}^2 * hg[n] + \tilde{d}_{j+1}^3 * gg[n]$$

Ces 4 convolutions séparables peuvent aussi être factorisées en 6 convolutions 1D appliquées selon les colonnes puis les lignes, comme le montre la Figure 4.3.

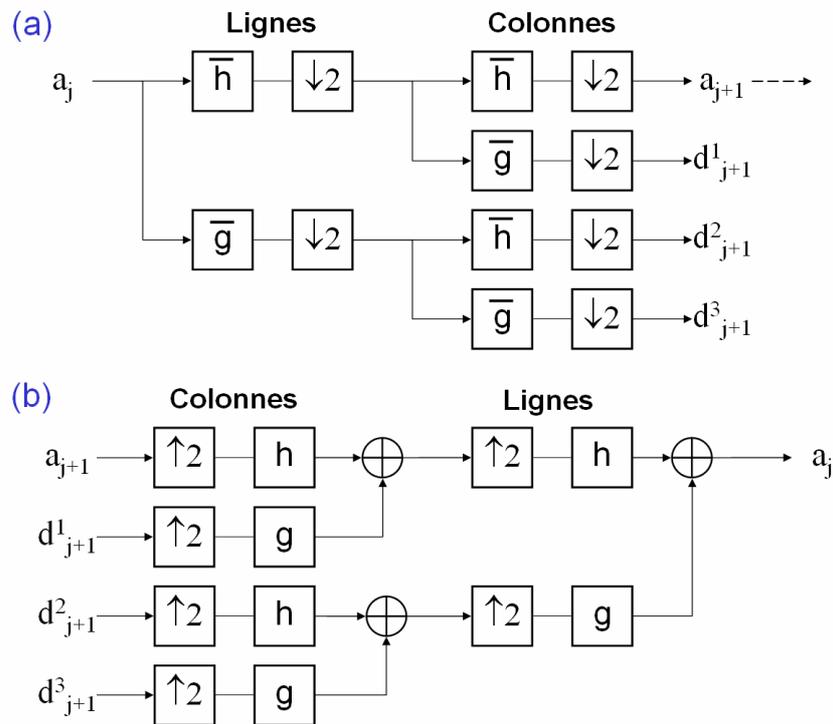


Figure 4.3 – (a) Décomposition multi-résolution de l'approximation de l'image à l'échelle j (a_j) en utilisant 6 groupes de convolution 1D et en sous-échantillonnant selon les lignes, puis les colonnes de l'image. (b) Reconstruction de a_j en insérant des zéros entre les lignes et les colonnes de a_{j+1} et d_{j+1}^k puis en filtrant les résultats ainsi obtenus.

Les coefficients d'ondelettes d'une image sont généralement représentés comme montré Figure 4.4.

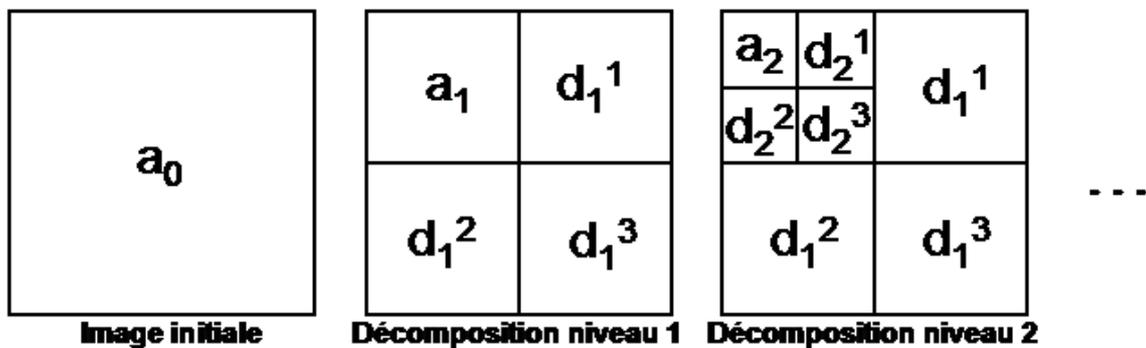


Figure 4.4 – Représentation de la décomposition multi-résolution pyramidale. « a_j » est l'approximation de l'image initiale « a_0 » à la résolution j . « d_j^1 », « d_j^2 » et « d_j^3 » sont respectivement les détails horizontaux, verticaux et diagonaux relatifs à la résolution j .

1.4.6 - Conclusion

La transformée en ondelettes orthogonale rapide consiste en un filtrage suivi d'un sous-échantillonnage par 2, appliquée en cascade. En dimension 1, un filtrage passe-bas permet de récupérer l'approximation du signal tandis qu'un filtrage passe-haut permet de récupérer les détails, c'est-à-dire toutes les différences entre l'approximation à la résolution supérieure et l'approximation à la résolution courante. En dimension 2, on applique 2 filtres successifs, sur les lignes puis sur les colonnes ; 2 filtres passe-bas successifs permettent d'obtenir l'approximation de l'image ; un filtre passe-haut sur les lignes suivi d'un filtre passe-bas sur les colonnes permettent d'obtenir les détails horizontaux ; un filtre passe-bas sur les lignes suivi d'un filtre passe-haut sur les colonnes permettent d'obtenir les détails verticaux ; enfin, 2 filtres passe-haut sur permettent d'obtenir les détails diagonaux.

De plus, sous certaines conditions entre les filtres orthogonaux passe-haut et passe-bas, la reconstruction du signal original à partir des coefficients d'approximation et de détails est parfaite.

Ensuite, les algorithmes de décomposition en ondelettes et de reconstruction sont simples à mettre en œuvre à l'aide de la formulation matricielle.

Enfin, dans le cas de signaux finis, il existe des moyens pour pallier à l'inconvénient des effets de bords.

Maintenant que nous avons étudié l'algorithme de transformée en ondelettes rapide, nous allons nous pencher sur les bases d'ondelettes existantes.

1.5 - Exemple de bases d'ondelettes orthonormales

1.5.1 - Introduction

Le choix de l'ondelette adaptée n'est pas aisé. Pour une analyse multi-résolution, les bases d'ondelettes orthonormales sont préférées, en particulier les ondelettes à support compact si on veut un calcul exact.

Dans cette partie, les différentes bases d'ondelettes testées sont présentées : l'ondelette de Haar, les ondelettes de Daubechies, les Symmlets, les Coiflets, l'ondelette de Shannon et les ondelettes de Battle-Lemarié.

1.5.2 - Choix de l'ondelette

Le choix de l'ondelette est délicat, mais il est important car il conditionne la qualité des résultats obtenus. Plusieurs caractéristiques peuvent guider ce choix.

Par exemple, on peut prendre une forme proche d'un motif que l'on souhaite mettre en évidence dans le signal, sans en connaître l'échelle. On peut aussi rechercher un bon compromis de résolution temps / fréquence.

On peut aussi s'attacher à ses propriétés mathématiques. En particulier, l'un des principaux atouts de l'analyse par ondelettes est sa capacité à produire des coefficients forts là où le signal est irrégulier et des coefficients très faibles dans les zones plus lisses. La petitesse des coefficients d'ondelettes sur les zones lisses est contrôlée par le nombre de moments nuls de l'ondelette. Une ondelette ψ a N moments nuls si :

$$\forall k \in \{0, \dots, N-1\}, \int_{\mathbb{R}} x^k \psi(x) dx = 0.$$

Le choix de l'ondelette est donc un compromis entre la taille de son support et son nombre de moments nuls ; plus son support est petit, moins nombreux seront les gros coefficients affectés par une irrégularité d'un signal ; d'un autre côté, prendre une ondelette avec beaucoup de moments nuls permet d'avoir des coefficients faibles sur les parties régulières du signal.

1.5.3 - L'ondelette de Haar

Définition

Historiquement, la base de Haar est la première base orthonormale d'ondelettes.

La fonction d'échelle est : $\phi(t) = \begin{cases} 1 & \text{si } t \in [0,1] \\ 0 & \text{sinon} \end{cases}$ (cf. Figure 4.5),

donc l'équation d'échelle est : $\phi(t) = \frac{1}{\sqrt{2}} \phi(2t) + \frac{1}{\sqrt{2}} \phi(2t-1)$.

On en déduit les coefficients du filtre passe-bas associé : $h[0] = \frac{1}{\sqrt{2}}$ et $h[1] = \frac{1}{\sqrt{2}}$.

Puis les coefficients du filtre passe-haut : $g[0] = \frac{1}{\sqrt{2}}$ et $g[1] = -\frac{1}{\sqrt{2}}$.

L'équation d'ondelette est donc : $\psi(t) = \frac{1}{\sqrt{2}} \phi(2t) - \frac{1}{\sqrt{2}} \phi(2t-1)$.

donc la fonction d'ondelette est donnée par : $\psi(t) = \begin{cases} 1 & \text{si } t \in \left[0, \frac{1}{2}\right] \\ -1 & \text{si } t \in \left[\frac{1}{2}, 1\right] \\ 0 & \text{sinon} \end{cases}$ (cf. Figure 4.5).

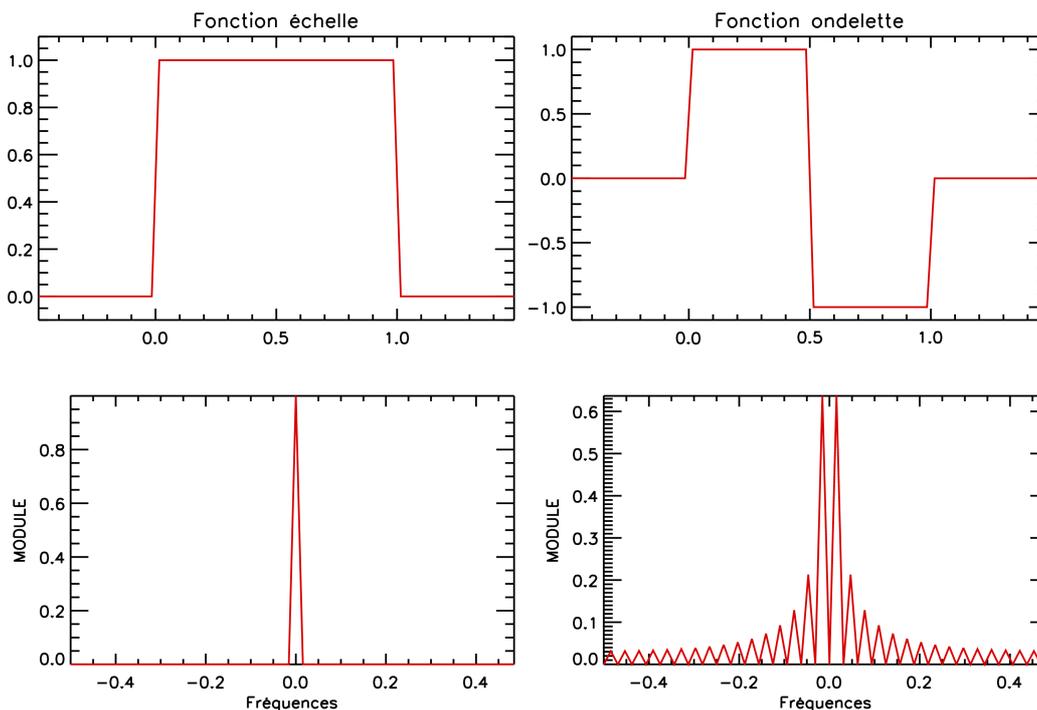


Figure 4.5 – A gauche, la fonction échelle de Haar (en haut), ainsi que le module de sa transformée de Fourier (en bas). A droite, la fonction ondelette de Haar (en haut), ainsi que le module de sa transformée de Fourier (en bas).

Avantages

Cette ondelette est très simple et facile à implémenter. De plus, elle est à support compact, donc le calcul de la transformée en ondelettes est exact.

Inconvénients

L'ondelette de Haar, quoique très simple, comporte un inconvénient majeur. Elle est discontinue et sa transformée de Fourier décroît très lentement : la localisation en fréquence est mauvaise (cf. Figure 4.5, en bas, à droite).

1.5.4 - Les ondelettes à support compact d'Ingrid Daubechies

Définition

Les ondelettes orthonormales à support compact (les filtres échelle et ondelettes associés sont donc à réponse impulsionnelle finie) sont définies à partir du théorème énoncé par Ingrid Daubechies [1988].

Ces ondelettes ont un support de taille minimale $2p - 1$ pour un nombre donné p de moments nuls. On peut montrer que ψ a p moments nuls si la transformée de Fourier \hat{h} de

h a un 0 d'ordre p pour la pulsation $\omega = \pi$ (ou la fréquence $f = \frac{\omega}{2\pi} = \frac{1}{2}$). Un tel polynôme trigonométrique s'écrit [Mallat, 1998] :

$$\hat{h}(\omega) = \sqrt{2} \left(\frac{1 + e^{-i\omega}}{2} \right)^p R(e^{-i\omega})$$

avec $(|R(\omega)|)^2 = R(e^{-i\omega})R(e^{i\omega}) = \sum_{k=0}^{p-1} \binom{p-1+k}{k} \left(\frac{1}{2} - \frac{1}{4}(e^{-i\omega} + e^{i\omega}) \right)^k$. $R(e^{-i\omega})$ est obtenu par

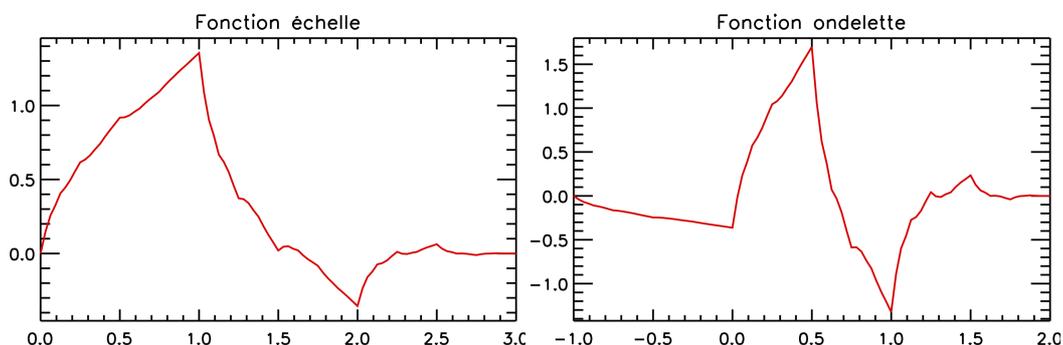
factorisation spectrale : $R(e^{-i\omega}) = \prod_{k=0}^{p-1} (e^{-i\omega} - r_k)$ où les r_k sont les racines du polynôme $R(e^{-i\omega})$,

c'est-à-dire les solutions de $(|R(\omega)|)^2 = R(e^{-i\omega})R(e^{i\omega}) = 0$. Ces racines vont par paires de nombres complexes conjugués. Par conséquent, il y a plusieurs factorisations possibles : la factorisation par minimum de phase (les racines sont choisies à l'intérieur du cercle unité), la factorisation par maximum de phase (les racines sont choisies à l'extérieur du cercle unité) ou une solution mixte. Les ondelettes de Daubechies sont construites en factorisant par minimum de phase.

Les ondelettes de Daubechies d'ordre 2 et 7 sont montrées Figure 4.6 et Figure 4.7 respectivement.

Avantages

Ces ondelettes sont particulièrement intéressantes dans la mesure où on peut choisir la régularité voulue en imposant un certain nombre de moments nuls : la régularité augmente avec p . De plus, étant à support compact, le calcul de la transformée en ondelette est exacte.



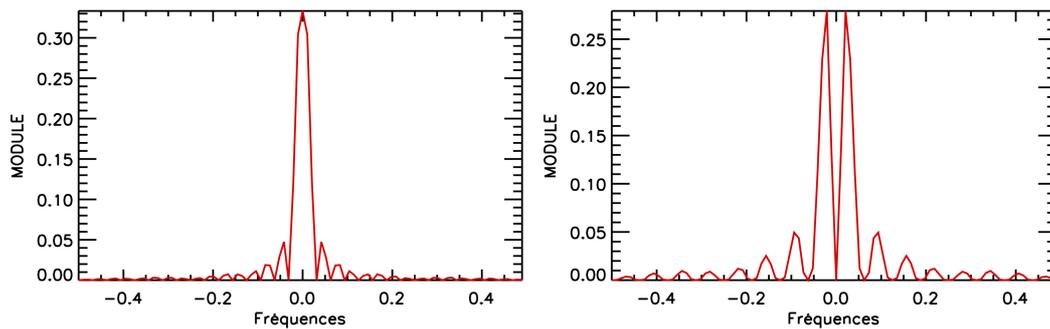


Figure 4.6 – A gauche, la fonction échelle de Daubechies d'ordre 2 (en haut), ainsi que le module de sa transformée de Fourier (en bas). A droite, même chose pour la fonction ondelette.

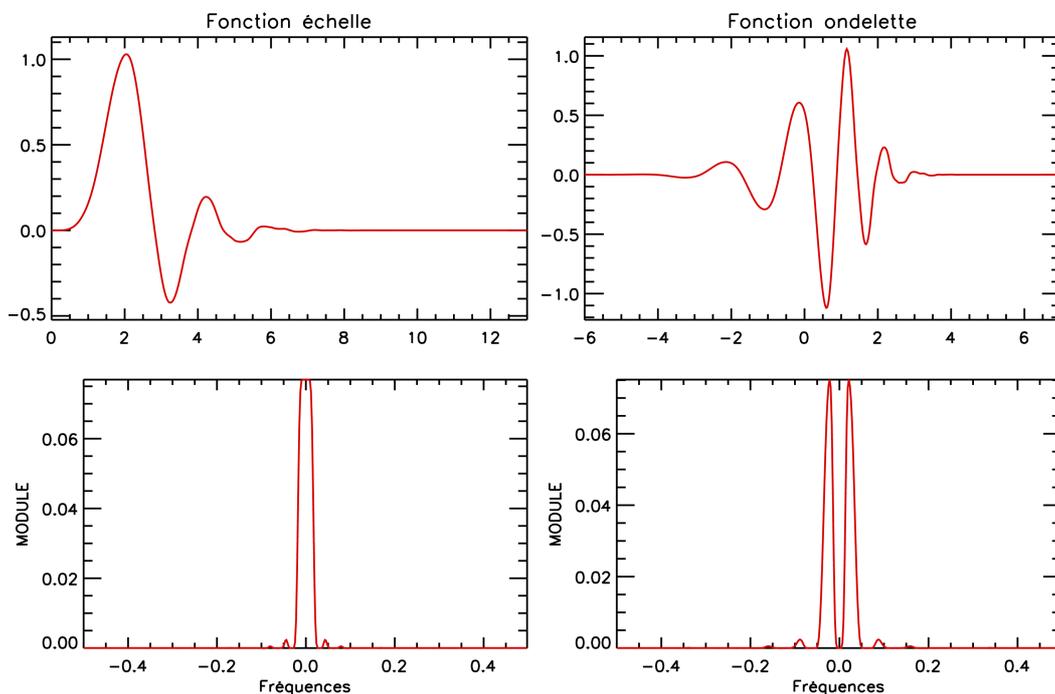


Figure 4.7 – A gauche, la fonction échelle de Daubechies d'ordre 7 (en haut), ainsi que le module de sa transformée de Fourier (en bas). A droite, même chose pour la fonction ondelette.

Inconvénients

Ces ondelettes ne sont pas symétriques, ce qui introduit un déphasage. Ceci peut être problématique dans certains problèmes, comme la détection de frontières. On peut obtenir les ondelettes les moins asymétriques en choisissant la factorisation par maximum de phase.

Remarque : il est à noter que l'ondelette de Daubechies d'ordre $p = 1$ est l'ondelette de Haar.

1.5.5 - Les Symmlets

Les ondelettes de Daubechies sont très asymétriques car elles sont construites en sélectionnant les racines en minimum de phase de $(|R(\omega)|)^2$. En effet, les filtres alors construits ont leur énergie concentrée près du point de départ de leur support, d'où l'asymétrie. C'est aussi le cas des ondelettes en dérivant.

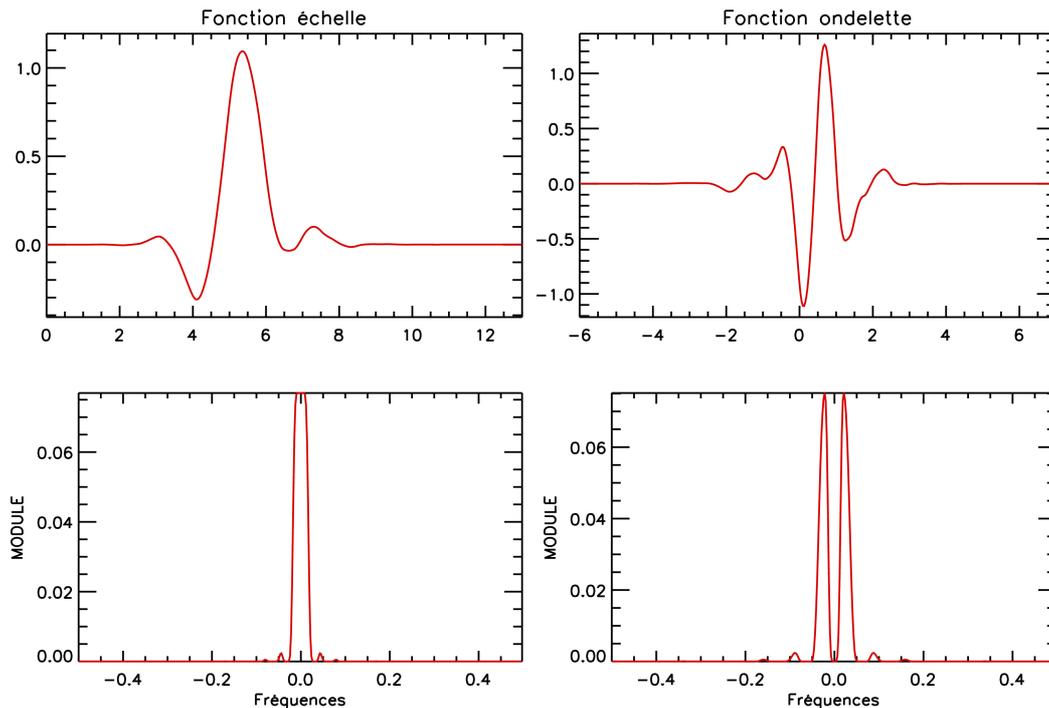


Figure 4.8 – A gauche, la fonction échelle Symmlet d'ordre 7 (en haut), ainsi que le module de sa transformée de Fourier (en bas). A droite, même chose pour la fonction ondelette.

Pour obtenir une ondelette symétrique ou anti-symétrique, le filtre h doit être lui aussi symétrique ou anti-symétrique par rapport au centre de son support, ce qui induit que sa transformée de Fourier $\hat{h}(\omega)$ a une phase complexe linéaire. Ainsi, les filtres Symmlets (pour SYMétrique) de Daubechies sont obtenus en optimisant le choix de la racine de $(|R(\omega)|)^2$ pour obtenir une phase quasi linéaire. La Figure 4.8 montre les fonctions échelle et ondelette de la Symmlet d'ordre 7.

1.5.6 - Les Coiflets

Pour une application en analyse numérique, Coifman a demandé à Ingrid Daubechies de construire une famille d'ondelettes ψ ayant p moments nuls et un support de taille minimal, mais la fonction échelle devant satisfaire :

$$\int_{-\infty}^{+\infty} \phi(t) dt = 1 \text{ et } \int_{-\infty}^{+\infty} t^k \phi(t) dt = 0 \text{ pour } k \in [1, p[.$$

Cette condition supplémentaire nécessite que la taille du support des Coiflets soit $3p - 1$ et non $2p - 1$ comme pour les ondelettes de Daubechies. La Figure 4.9 montre les fonctions échelle et ondelette de la Coiflet d'ordre 4.

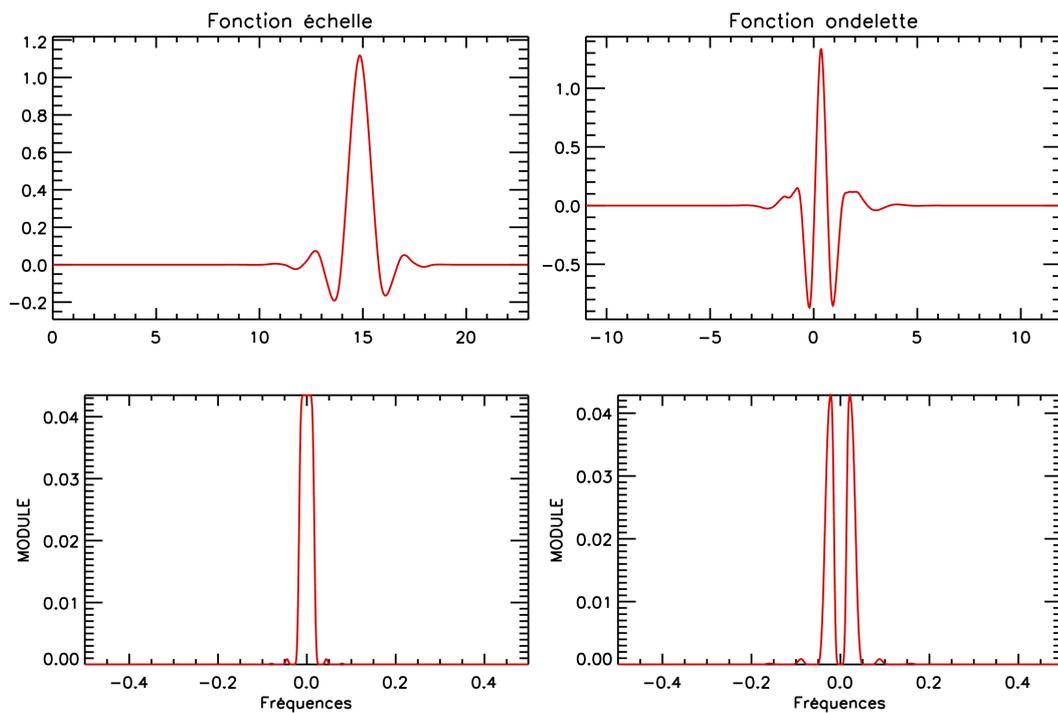


Figure 4.9 – A gauche, la fonction échelle Coiflet d'ordre 4 (en haut), ainsi que le module de sa transformée de Fourier (en bas). A droite, même chose pour la fonction ondelette.

1.5.7 - L'ondelette de Shannon

Définition

L'ondelette de Shannon peut être considérée comme l'inverse de l'ondelette de Haar puisque :

☞ la transformée de Fourier de sa fonction échelle s'écrit : $\hat{\phi}(\omega) = \begin{cases} 1 & \text{si } \omega \in [-\pi, \pi] \\ 0 & \text{sinon} \end{cases}$,

☞ la transformée de Fourier du filtre échelle associé s'écrit : $\hat{h}(\omega) = \begin{cases} \sqrt{2} & \text{si } \omega \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \\ 0 & \text{sinon} \end{cases}$.

On peut déterminer :



les coefficients du filtre échelle :

$$h[k] = \begin{cases} \frac{\sqrt{2}}{2} & \text{si } k = 0 \\ \sqrt{2} \frac{(-1)^{\frac{k-1}{2}}}{k\pi} & \text{si } k \text{ est impair,} \\ 0 & \text{sinon} \end{cases}$$



la transformée de Fourier de la fonction ondelette :

$$\hat{\psi}(\omega) = \begin{cases} e^{-i\omega/2} & \text{si } \omega \in [-2\pi, -\pi] \cup [\pi, 2\pi] \\ 0 & \text{sinon} \end{cases}$$



la fonction ondelette : $\psi(t) = \frac{\sin 2\pi(t-1/2)}{2\pi(t-1/2)} - \frac{\sin \pi(t-1/2)}{\pi(t-1/2)}$.

Inconvénients

Les fonctions échelle et ondelette décroissent très lentement dans le temps ; les coefficients du filtre d'échelle varient en $1/k$. La localisation temporelle est donc très pauvre. Les fonctions échelle et ondelette ne sont pas à support compact donc les filtres associés sont à réponse impulsionnelle infinie. Dans le cas de signaux finis, il faut garder les coefficients les plus significatifs : la Figure 4.10 montre les 41 coefficients significatifs des filtres échelle et ondelettes. Cela pose des problèmes de reconstruction parfaite.

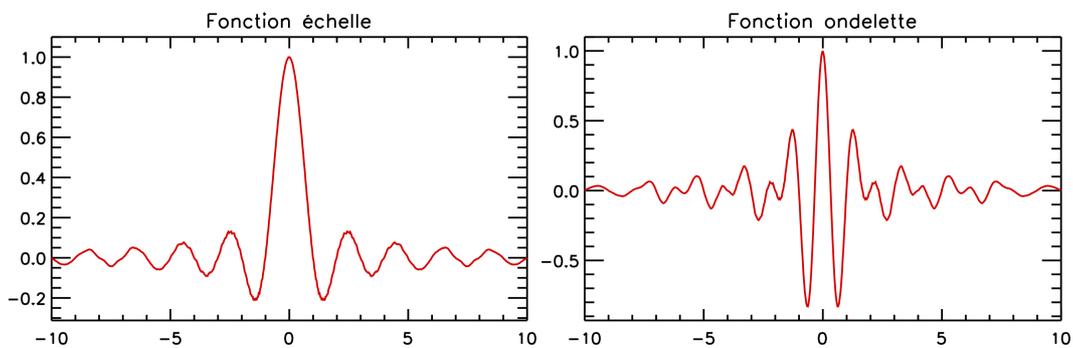


Figure 4.10 – La fonction échelle de Shannon (à gauche), la fonction ondelette associée (à droite).

1.5.8 - Les ondelettes splines de Battle-Lemarié

Définition

Les ondelettes splines polynomiales ont été introduites par Battle [1989] et Lemarié [1988] dans le cadre des approximations multi-résolution splines. Posant

$$S_n(\omega) = \sum_{k=-\infty}^{+\infty} \frac{1}{(\omega + 2k\pi)^n}, \text{ on a les expressions suivantes [Mallat, 1998] :}$$

la transformée de Fourier de la fonction échelle s'écrit : $\hat{\phi}(\omega) = \frac{e^{-i\varepsilon\omega/2}}{\omega^{m+1} \sqrt{S_{2m+2}(\omega)}}$ où m

est le degré de la fonction spline associée et $\varepsilon = \begin{cases} 1 & \text{si } m \text{ est pair} \\ 0 & \text{si } m \text{ est impair} \end{cases}$

la transformée de Fourier du filtre échelle s'écrit : $\hat{h}(\omega) = e^{-i\varepsilon\omega/2} \sqrt{\frac{S_{2m+2}(\omega)}{2^{2m+1} S_{2m+2}(2\omega)}}$ où

$\varepsilon = \begin{cases} 1 & \text{si } m \text{ est pair} \\ 0 & \text{si } m \text{ est impair} \end{cases}$

la transformée de Fourier de la fonction ondelette s'écrit :

$$\hat{\psi}(\omega) = \frac{e^{-i\omega/2}}{\omega^{m+1}} \sqrt{\frac{S_{2m+2}(\omega/2 + \pi)}{S_{2m+2}(\omega) S_{2m+2}(\omega/2)}}$$

La fonction échelle est symétrique par rapport à $\frac{1}{2}$ si m est impair et par rapport à 0 si m

est pair. La fonction ondelette est symétrique par rapport à $\frac{1}{2}$ si m est impair et anti-

symétrique par rapport à $\frac{1}{2}$ si m est pair.

La Figure 4.11 montre les fonctions d'échelle et d'ondelette de Battle-Lemarié d'ordre 3.

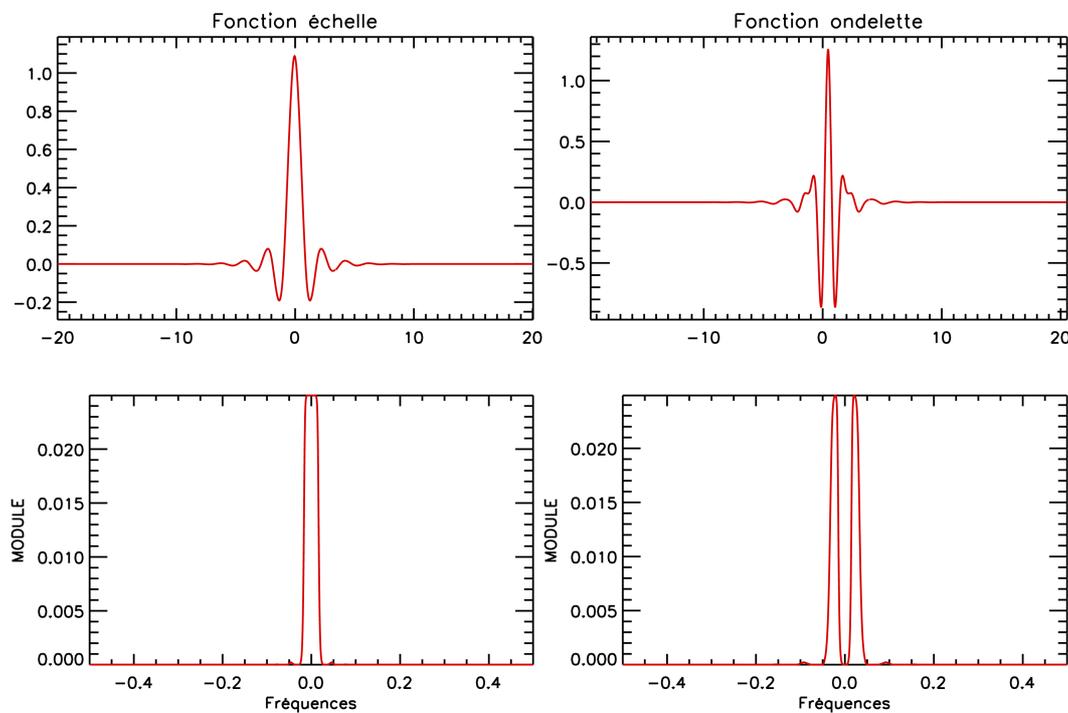


Figure 4.11 – A gauche, la fonction échelle de Battle-Lemarié d'ordre 3 (en haut), ainsi que le module de sa transformée de Fourier (en bas). A droite, même chose pour la fonction ondelette.

Avantages

Ces ondelettes sont à décroissance exponentielle. Ainsi, même si les filtres sont à réponse impulsionnelle infinie, ne garder que les coefficients centraux les plus importants pour traiter les signaux finis ne pose pas de problème quand à la reconstruction, qui est quasi parfaite. Par exemple, garder les 41 coefficients centraux pour l'ordre 3 est tout à fait raisonnable.

1.5.9 - Conclusion

Dans cette partie, nous avons vu plusieurs exemples d'ondelettes orthogonales. Certaines sont à support compact, comme Haar, Daubechies, les Symmlets et les Coiflets ; c'est-à-dire qu'elles sont non nulles sur un intervalle donné uniquement. Leurs filtres associés sont alors à réponse impulsionnelle finie et la reconstruction est parfaite. D'autres, tels l'ondelette de Shannon et les ondelettes splines de Battle-Lemarié, ne sont pas à support compact. Leurs filtres associés sont donc à réponse impulsionnelle infinie, mais, moyennant un choix judicieux du nombre de coefficients utilisés, peut amener à une reconstruction quasi-parfaite.

Tous ces types d'ondelettes seront testés par la suite, afin de déterminer le plus adapté à notre problème.

1.6 - Conclusion

Cette partie nous a permis de nous familiariser avec la théorie des ondelettes. Nous avons tout d'abord approché le principe de la transformée en ondelette, continue et discrète, ainsi que l'utilisation de la seconde dans le cas de l'analyse multi-résolution, c'est-à-dire la décomposition du signal en une approximation et les détails correspondant à une résolution donnée. Puis nous avons vu que les étapes de décomposition en ondelettes et de reconstruction s'implémentent aisément à l'aide de bancs de filtres : l'approximation du signal est obtenu par filtrage passe-bas de celui-ci, tandis que les détails sont obtenus par filtrage passe-haut du signal. Enfin, nous avons caractérisé les différentes bases d'ondelettes testées dans cette étude.

Nous allons maintenant nous intéresser à la méthodologie mise en place pour combiner plusieurs jeux de données irrégulièrement réparties sur une grille régulière.

2 - Méthodologie

2.1 - Introduction

Dans cette partie, nous nous intéressons à la combinaison des données couleur de l'eau à l'aide de la transformée en ondelettes.

Dans un premier temps, nous détaillons la méthode mise au point pour estimer les données manquantes dans les données de concentration en chlorophylle. Dans un deuxième temps, nous voyons que l'analyse pyramidale permet de fusionner des images complètes, de manière rapide et tout en gardant les détails et les structures propres à la donnée. Enfin, nous montrons comment nous avons choisi de combiner ces deux techniques pour combiner des images couleur de l'eau issues de plusieurs capteurs.

2.2 - Estimation des données manquantes

2.2.1 - Introduction

La méthode d'estimation des données manquantes a été réalisée en collaboration avec Antonio TURIEL de l'Institut de Ciències del Mar (Barcelone, Espagne). Cette méthode est basée sur ses travaux [Turiel et Parga, 2000b ; Turiel *et al.*, 2003] ainsi que sur ceux de Simoncelli et Buccigrossi [Buccigrossi et Simoncelli, 1997 ; Simoncelli et Buccigrossi, 1997 ; Simoncelli, 1999 ; Buccigrossi et Simoncelli, 1999]. Le principe est de développer un modèle de probabilité pour des images (des images photographiques naturelles, dans le cas des études de Simoncelli et Buccigrossi : la fameuse photographie de Lena, Einstein, les bateaux, le singe, etc. ; les cartes de concentration en chlorophylle dans notre cas) basé sur l'observation empirique de leurs statistiques dans le domaine de la transformée en ondelettes (décomposition multi-résolution : voir partie 1.3). En effet, il existe des relations statistiques conditionnelles entre les coefficients d'ondelette à différentes résolutions. Ainsi, il est possible de construire un modèle non-paramétrique sur la probabilité conditionnelle d'un coefficient connaissant les coefficients à une résolution plus grossière.

Dans cette partie, nous allons tout d'abord replacer la méthodologie envisagée dans son contexte, en montrant les différentes étapes de la réflexion menée pour aboutir à cette méthode. Puis nous abordons l'implémentation du prédicteur utilisé pour estimer les données manquantes. Des tests de configuration ondelette / type de données sont ensuite réalisés et l'application aux cartes de données de concentration en chlorophylle est alors montrée.

2.2.2 - Statistiques à l'origine de la méthode

Commençons par nous familiariser avec les appellations qui seront utilisées par la suite. On peut considérer l'analyse multi-résolution comme un arbre généalogique et les coefficients d'ondelette par orientation (détail horizontal, vertical ou diagonal) et par échelle (résolution) comme les membres d'une même « famille », comme décrit Figure 4.12. Ainsi, un coefficient fils correspond à une orientation et à une échelle données. Un coefficient fils a deux cousins, correspondant aux coefficients des deux autres orientations pour la même échelle et la même localisation spatiale. Un coefficient fils a un père, de même orientation mais relatif à la résolution juste plus grossière à la localisation spatiale correspondante : par conséquent, un coefficient père a 4 coefficients fils. Enfin, un coefficient fils a deux oncles, correspondant aux coefficients des deux autres orientations pour la même résolution que le père et la même localisation spatiale : les deux coefficients oncles sont donc les cousins du coefficient père...

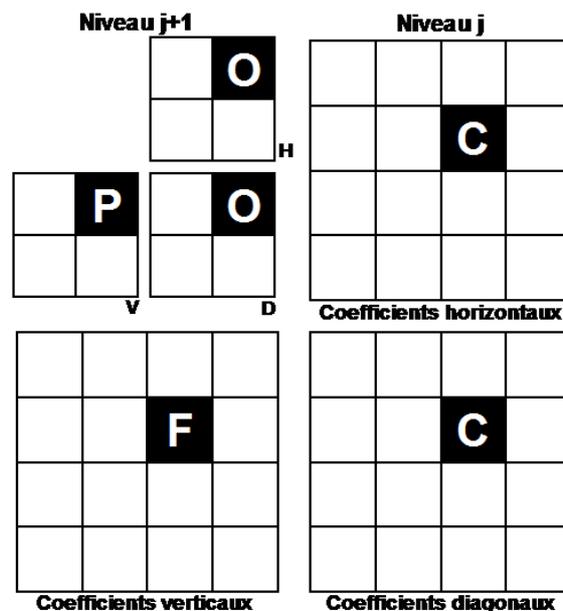


Figure 4.12 – Appellation des différents coefficients. « F » est le coefficient fils. Les coefficients « C » sont les 2 cousins du coefficient fils. « P » est le coefficient père associé à 4 coefficients fils. Les coefficients « O » sont les 2 oncles du coefficient fils.

En ce qui concerne les images utilisées pour les calculs statistiques effectués par la suite, le jeu de données consiste en 887 imageries complètes (i.e. sans donnée manquante) de taille 128 x 128 pixels (la plus grande taille possible pour pouvoir récupérer un maximum d'imageries complètes) obtenues à partir des données hebdomadaires (pratiquement autant de variabilité spatiale que les données journalières, mais moins de trous) de concentration en chlorophylle de MODIS/Aqua entre 2002 et 2006.

Les familles d'ondelettes testées sont celles décrites au paragraphe 1.5.

De plus, puisque la distribution des données de concentration en chlorophylle-a est lognormale [Campbell, 1995], les calculs ont été effectués à chaque fois sur les données naturelles puis sur leur logarithme base 10.

Ceci étant posé, intéressons-nous maintenant au cœur du problème. Nous avons vu dans la partie 1.3 que l'idée sous-entendue par la décomposition multi-résolution d'une image était d'analyser toutes les échelles de celle-ci, de la plus petite (i.e. résolution la plus fine) à la plus grande (i.e. résolution la plus pauvre) et de les comparer entre elles. Il existe une propriété très connue portant sur les images de divers systèmes réels : la **persistance** des détails. Qualitativement, cette propriété se traduit par le fait que les structures de l'image qui sont typiques et très localisées (par exemple, les fronts) tendent à avoir une énergie qui persiste à la même localisation spatiale à travers les différentes échelles, comme on peut le constater Figure 4.13 pour une imagerie MODIS/Aqua du 29 juin 2003.

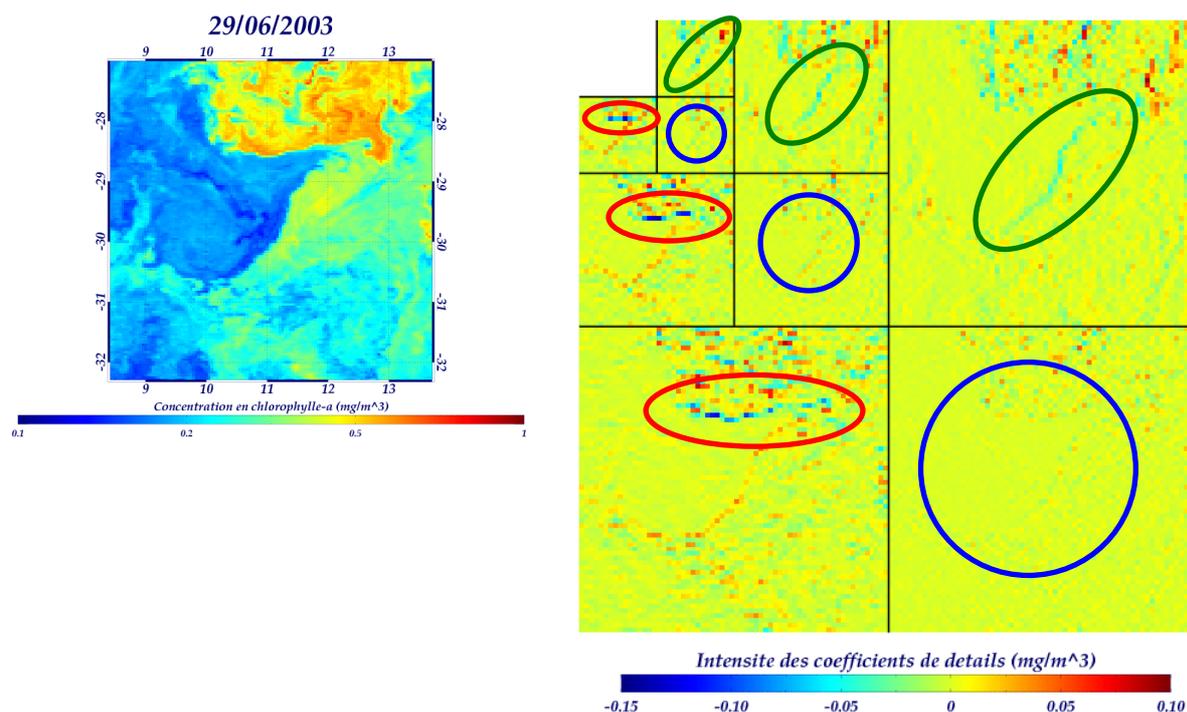


Figure 4.13 – A gauche : imagerie 128 x 128 pixels issue de MODIS/Aqua du 29 juin 2003. A droite : cascade des coefficients d'ondelettes résultants de l'application, sur 3 niveaux consécutifs, de l'algorithme pyramidal à l'imagerie montrée à gauche (ondelette utilisée : Haar).

Quantitativement, ce phénomène de persistance se traduit par une relation coefficient père / coefficient fils assez significative, soit mathématiquement :

$$I(\alpha_P, \alpha_F) > I_0$$

où α_F est un coefficient d'ondelette fils, α_P son coefficient d'ondelette père associé, $I(\alpha_P, \alpha_F)$ l'inférence (ou déduction) du coefficient d'ondelette père connaissant le coefficient d'ondelette fils et I_0 une valeur seuil pour l'inférence significative. Ce comportement sous-entend que les coefficients d'ondelette fils sont fonction, entre autres, de leur père associé :

$$\alpha_F = \text{fonction}(\alpha_P, \eta_1, \eta_2, \dots)$$

où les η_i sont des variables aléatoires indépendantes de α_P et α_F et de n'importe quelle autre variable $\eta_j, j \neq i$.

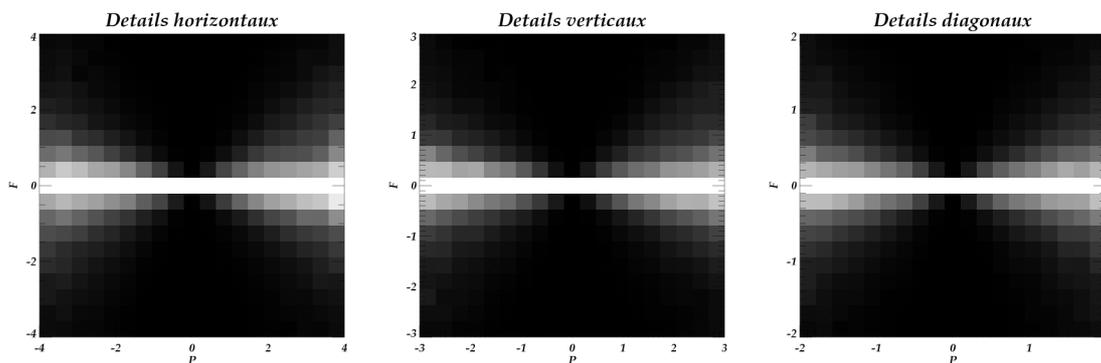


Figure 4.14 – Représentation de la probabilité conditionnelle de F (coefficient d'ondelette fils) sachant P (coefficient d'ondelette père associé au fils F) pour les détails horizontaux (à gauche), verticaux (au milieu) et diagonaux (à droite) : le contraste correspond à la probabilité (le blanc représentant la probabilité la plus forte), mise à part que chaque colonne a été normalisée indépendamment des autres, pour plus de visibilité. Ondelette utilisée : Battle-Lemarié d'ordre 3 avec 41 coefficients retenus.

Examinons les relations statistiques existant entre chaque coefficient fils et son père. La Figure 4.14 montre l'histogramme conditionnel ou l'espérance mathématique $E(\alpha_F|\alpha_P)$ du coefficient fils α_F conditionné par son coefficient père α_P associé, obtenu à partir de notre jeu d'images décrit plus haut. Les coefficients sont (approximativement) décorrélés au premier ordre, puisque la valeur attendue pour α_F est à peu près nulle pour chaque valeur de α_P . Cette caractéristique est visible pour toutes les familles d'ondelettes testées et qu'on s'intéresse aux valeurs naturelles ou à leur logarithme base 10.

Penchons-nous maintenant sur le moment d'ordre 2 : la variance $\text{var}(\alpha_F|\alpha_P)$ de α_F sachant α_P . Sur la Figure 4.15, on observe que, toujours sur notre jeu d'images de chlorophylle, cette variance dépend paraboliquement des coefficients d'ondelette père :

$$\text{var}(\alpha_F|\alpha_P) = A_\psi \alpha_P^2 + B_\psi \text{ avec } A_\psi B_\psi > 0.$$

Une analyse sur notre panel d'ondelettes décrites au paragraphe 1.5 nous a permis de vérifier que les coefficients A_ψ et B_ψ sont propres à chaque ondelette ψ .

On a aussi pu ainsi se rendre compte que certaines familles d'ondelettes ne conviennent pas aux données de concentration en chlorophylle car la courbe de variance conditionnelle correspondante ne présente pas de caractéristique parabolique : il s'agit des Coiflets, des ondelettes de Daubechies et de l'ondelette de Shannon.

Revenons à notre relation liant le coefficient fils α_F à son père α_P . Etant donné la forme de la courbe de la variance conditionnelle, on peut alors proposer le modèle suivant (le « . » symbolise une multiplication entre variables aléatoires) :

$$\alpha_F = \text{fonction}(\alpha_P, \eta_1, \eta_2, \dots) = \eta_1^\psi \cdot \alpha_P + \eta_2^\psi .$$

Ainsi, on a : $A_\psi = \text{var}(\eta_1^\psi)$ et $B_\psi = \text{var}(\eta_2^\psi)$.

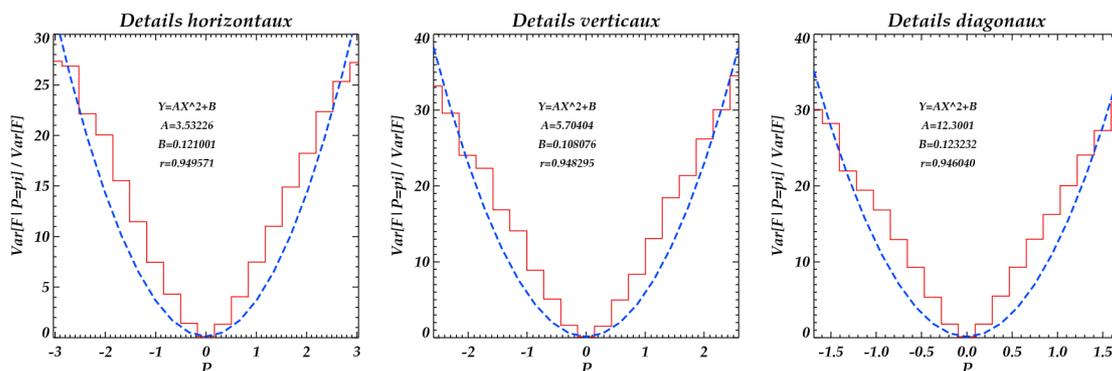


Figure 4.15 – Représentation de la variance conditionnelle de F (coefficient d'ondelette fils) sachant P (coefficient d'ondelette père associé au fils F) pour les détails horizontaux (à gauche), verticaux (au milieu) et diagonaux (à droite). La normalisation est faite par la variance des coefficients fils. Ondelette utilisée : Battle-Lemarié d'ordre 3 avec 41 coefficients retenus.

Tous ces éléments sont à mettre en relation avec la théorie des cascades multiplicatives de la turbulence. On pense que l'existence d'une cascade invariante d'échelle au sein de l'écoulement turbulent sous-jacent se voit aussi dans les propriétés de n'importe quel scalaire dominé par l'advection, dont la concentration en chlorophylle-a est un exemple. Une discussion en profondeur sur la signification de la cascade et de sa pertinence pour expliquer les données scalaires serait trop longue pour paraître dans ce manuscrit ; le lecteur pourra se tourner vers l'excellent ouvrage d'Uriel Frisch [1964] pour une discussion générale. L'existence d'une cascade multiplicative implique, dans le cas d'une représentation en ondelettes dyadiques, que $\alpha_F \doteq \eta \cdot \alpha_P \quad \forall \psi$, le symbole « \doteq » signifiant égalité au sens

des distributions [Turiel et Parga, 2000a,b]. On a une égalité stricte $\alpha_F = \eta \cdot \alpha_P$ pour l'ondelette ψ optimale [Turiel et Parga, 2000b].

Or on peut remarquer que, pour certaines ondelettes, les coefficients B_ψ que nous avons obtenus sont très proches de 0. Par conséquent, pour une bonne ondelette, on peut considérer $\eta_2 \approx 0$ avec une bonne approximation. On a alors :

$$\alpha_F = \eta \cdot \alpha_P \Rightarrow \eta = \frac{\alpha_F}{\alpha_P}$$

c'est-à-dire que, pour une ondelette proche de l'ondelette optimale, il est possible de récupérer les coefficients de cascade η à partir des coefficients d'ondelette père et fils. Au contraire, si l'ondelette n'est pas adaptée à la donnée étudiée, on ne pourra pas négliger η_2 et l'inférence père / fils sera donc mauvaise.

La structure de la relation entre les coefficients père et fils se dessine plus précisément quand on se place dans le domaine logarithmique. La relation précédente devient alors :

$$\log_2|\alpha_F| = \log_2|\eta| + \log_2|\alpha_P|.$$

La courbe représentative doit donc être une droite de pente 1. La Figure 4.16 montre l'histogramme conditionnel $E(\log_2|\alpha_F| | \log_2|\alpha_P|)$ pour notre jeu d'images et dans le cas de l'ondelette de Battle-Lemarié d'ordre 3 (41 coefficients retenus). La partie droite de cette distribution est concentrée autour d'une droite de pente unitaire, confirmant ainsi la théorie : ainsi, pour les valeurs de $\log_2|\alpha_P|$ comprises entre -7 et +2, la prévision conditionnelle $E(\alpha_F | \alpha_P)$ est approximativement proportionnelle à $|\alpha_P|$ [Buccigrossi et Simoncelli, 1999]. On peut aussi noter que les sections verticales (i.e. l'histogramme conditionnel pour une valeur fixée de $|\alpha_P|$) ont à peu près la même forme pour les différentes valeurs de $|\alpha_P|$. Enfin, la partie gauche de la distribution est concentrée sur une ligne horizontale, suggérant que $|\alpha_F|$ est indépendant de $|\alpha_P|$ entre -12 et -7. Il est possible que les coefficients de petite amplitude comportent majoritairement des erreurs de quantification et d'autres sources d'incertitudes. La forme caractéristique de la courbe représentative a été obtenue pour toutes les familles d'ondelettes testées. D'autre part, la partie constituée de la droite de pente unitaire est visiblement plus longue dans le cas d'une application aux données de concentration en chlorophylle en comparaison au logarithme base 10 de ces valeurs.

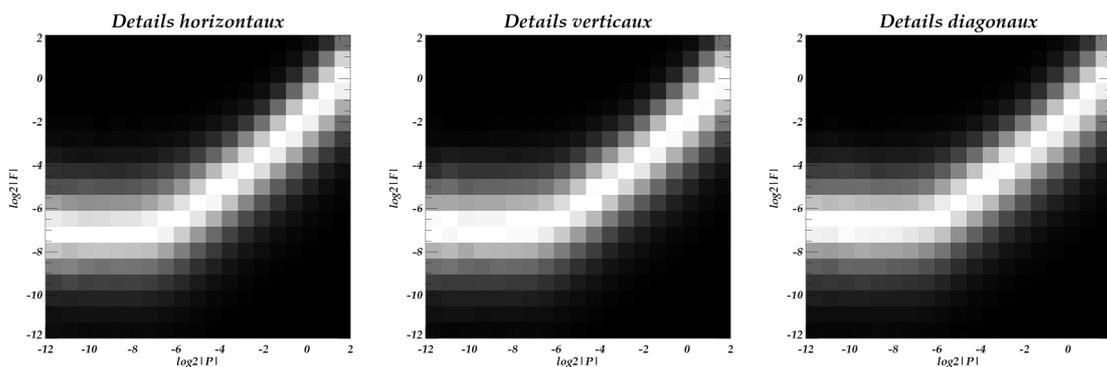


Figure 4.16 – Représentation de la probabilité conditionnelle de $\log_2|F|$ (logarithme base 2 ou logarithme naturel du coefficient d'ondelette fils) sachant $\log_2|P|$ (logarithme naturel du coefficient d'ondelette père associé au fils F) pour les détails horizontaux (à gauche), verticaux (au milieu) et diagonaux (à droite) : le contraste correspond à la probabilité (le blanc représentant la probabilité la plus forte), mise à part que chaque colonne a été normalisée indépendamment des autres, pour plus de visibilité. Ondelette utilisée : Battle-Lemarié d'ordre 3 avec 41 coefficients retenus.

En ce qui concerne la variance, le passage en logarithme base 10 permet d'obtenir la relation : $\text{var}(\log_2|\alpha_F| | \log_2|\alpha_P|) = \text{var}(\log_2|\eta|) = \text{cste}$. Cela signifie que η est également distribuée pour toute résolution et position. La courbe représentative de la variance conditionnelle empirique doit donc être approximativement une droite horizontale, comme montré Figure 4.17. Contrairement aux ondelettes de Battle-Lemarié et aux Symmlets qui présentent ces caractéristiques, ce n'est pas le cas pour les Coiflets, et aux ondelettes de Daubechies, de Haar et de Shannon.

Dans cette partie, nous avons vérifié que les statistiques nécessaires à notre problème étaient bien vérifiées sur nos données. Nous avons testé plusieurs familles d'ondelettes : ce sont les ondelettes de Battle-Lemarié et les Symmlets qui conviennent le mieux à nos données. De plus, nous nous sommes aperçus que l'application aux données de chlorophylle semblait mieux convenir à l'application aux logarithme base 10 de ces valeurs. Dans la partie suivante, nous allons nous intéresser à l'implémentation de l'estimateur.

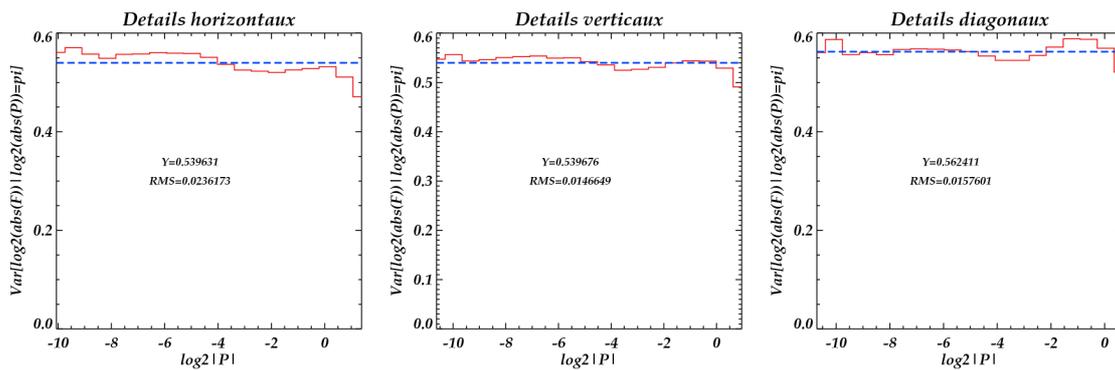


Figure 4.17 – Représentation de la variance conditionnelle de $\log_2|F|$ (coefficient d'ondelette fils) sachant $\log_2|P|$ (coefficient d'ondelette père associé au fils F) pour les détails horizontaux (à gauche), verticaux (au milieu) et diagonaux (à droite). Ondelette utilisée : Battle Lemarié d'ordre 3 avec 41 coefficients retenus.

2.2.3 - Principe de l'estimateur par ondelette ou calcul de l'inférence fils → père

L'objectif de notre estimateur est de déterminer le coefficient d'ondelette père le plus probable, connaissant ses 4 fils.

L'idée de l'estimateur est venue de la formule suivante, vue dans le paragraphe précédent :

$$\alpha_F = \eta \cdot \alpha_P \Rightarrow \log_2|\alpha_F| = \log_2|\eta| + \log_2|\alpha_P|$$

Comme il en va de même pour chacun des 4 coefficients d'ondelette fils associés à un coefficient d'ondelette père, on a le système suivant :

$$\begin{aligned} \log_2|\alpha_{F_1}| - \log_2|\alpha_P| &= \log_2|\eta_1| \\ \log_2|\alpha_{F_2}| - \log_2|\alpha_P| &= \log_2|\eta_2| \\ \log_2|\alpha_{F_3}| - \log_2|\alpha_P| &= \log_2|\eta_3| \\ \log_2|\alpha_{F_4}| - \log_2|\alpha_P| &= \log_2|\eta_4| \end{aligned}$$

Ce système comporte 5 inconnues : le coefficient d'ondelette père α_P (que l'on cherche à déterminer) et les η_i . Si les η_i ne peuvent être calculés directement, il est néanmoins possible de quantifier la densité de probabilité de $\log_2|\eta|$ à partir de tout notre jeu d'images. La Figure 4.18 montre les courbes obtenues pour l'ondelette de Battle-Lemarié d'ordre 3 (41 coefficients retenus) et la Symmlet d'ordre 7. On obtient ainsi une sorte d'« abaque » indispensable à nos calculs d'estimation.

Ondelette de Battle-Lemarié d'ordre 3

Symplet d'ordre 7

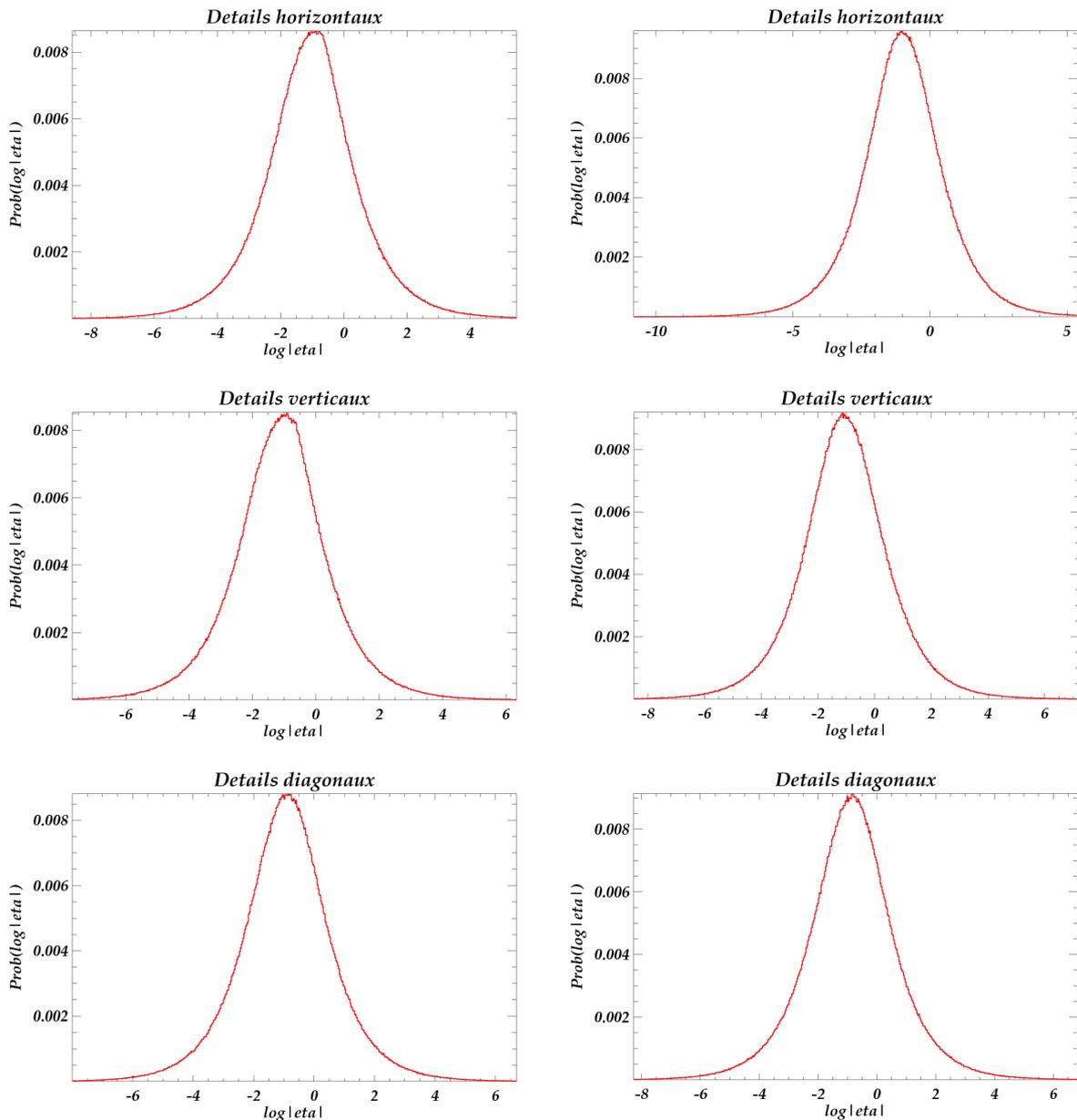


Figure 4.18 – Probabilité de $\log|\eta|=\log$ obtenue pour l'ondelette de Battle-Lemarié d'ordre 3 avec 41 coefficients retenus (à gauche) et la Symplet d'ordre 7 (à droite), et pour les coefficients de détails horizontaux (en haut), détails verticaux (au milieu) et détails diagonaux (en bas).

Intéressons-nous maintenant à l'expression : $\Delta(\theta) = \sum_{i=1}^4 \text{Prob}(\log_2|\alpha_{F_i}| - \theta) = \sum_{i=1}^4 \text{Prob}(\log_2|\eta_i|)$.

La valeur $\log_2|\alpha_p|$ que nous recherchons est tout simplement la valeur de θ la plus probable, i.e. la valeur θ_0 qui vérifie la relation : $\Delta(\theta_0) \geq \Delta(\theta) \quad \forall \theta$. En pratique, on fait décrire à θ un large panel de valeurs dont les bornes sont imposées par les extrema des 4

α_{F_i} ainsi que l'intervalle de présence de $\log_2|\eta|$ indiqué par l'abaque ; puis on calcule $\Delta(\theta)$ pour chaque valeur de θ grâce à l'abaque.

Connaissant θ_0 , on en déduit la valeur estimée du coefficient père : $|\alpha_p^{\text{est}}| = e^{\theta_0}$. Comme notre méthode ne permet de connaître que l'amplitude du coefficient père et pas son signe, nous avons choisi d'imposer, pour le coefficient père estimé, le signe du coefficient père original.

2.2.4 - Implémentation de l'algorithme

Dans cette partie, nous allons montrer les différentes étapes de notre algorithme au travers d'un exemple : une imagerie au milieu de laquelle nous avons posé artificiellement une large zone de valeurs manquantes (Figure 4.19). Pour cette étude que nous avons menée sur les imageries des données brutes (et non leur logarithme base 10), nous avons utilisé l'ondelette de Battle-Lemarié d'ordre 3 avec 41 coefficients retenus.

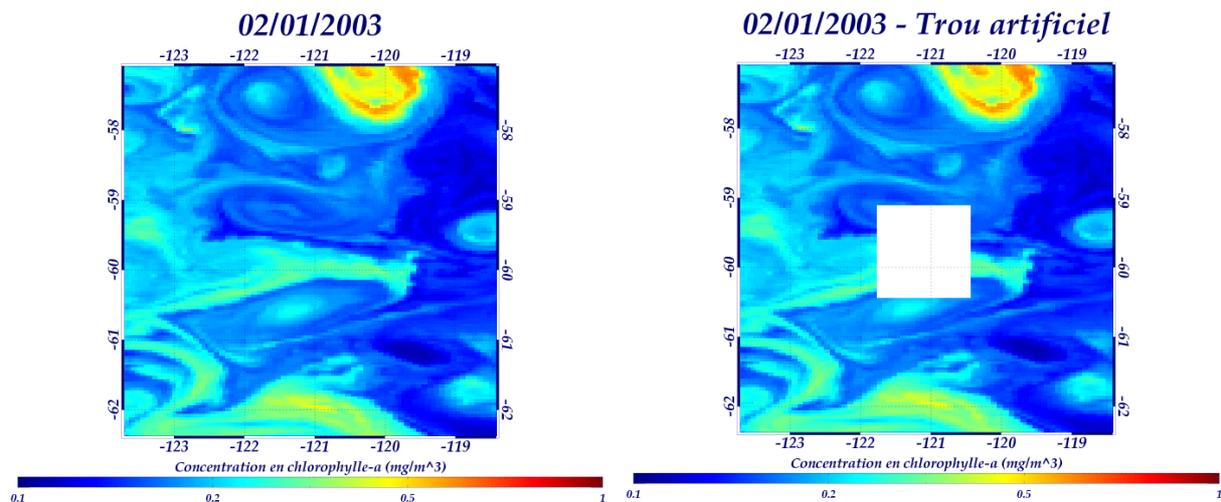


Figure 4.19 – A gauche : imagerie 128 x 128 pixels (i.e. environ 25 km x 25 km) de concentration en chlorophylle-*a* issue de MODIS/Aqua pour le 2 janvier 2003. A droite : même imagerie, avec la zone de données manquantes créée artificiellement pour notre exemple.

Notre algorithme de reconstruction comporte un prétraitement de l'image composé de deux étapes (Figure 4.20).

L'image originale est tout d'abord étendue symétriquement, comme le montre la Figure 4.21 : une fois horizontalement et une fois verticalement. Ceci est réalisé afin d'éviter les effets de bords durant l'analyse en ondelettes car l'utilisation de la périodisation (cf. partie 1.4.3 sur les problèmes de bords) est préférable sur une image la plus cyclique possible.

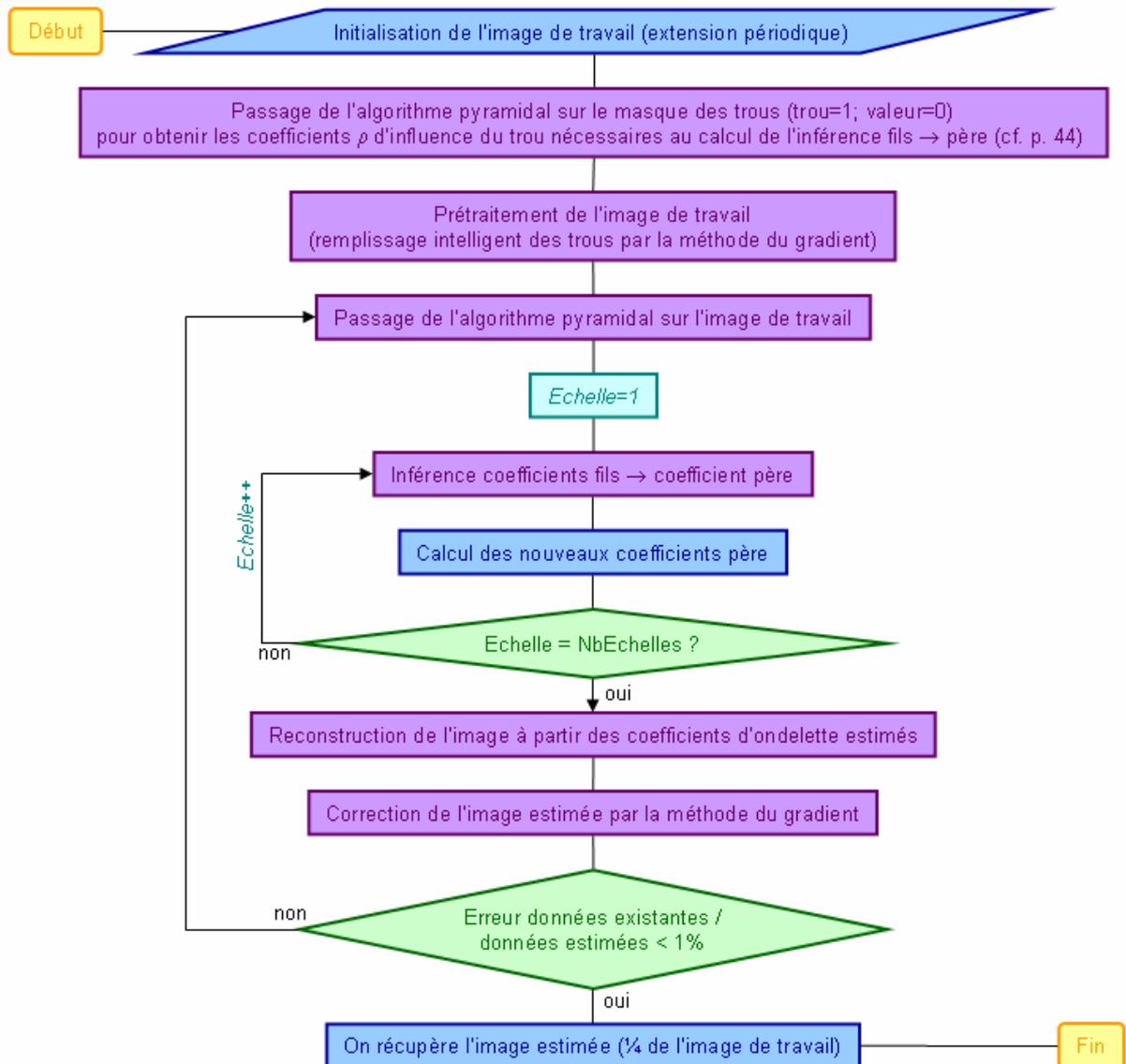


Figure 4.20 – Organigramme du logiciel de reconstruction d'une image.

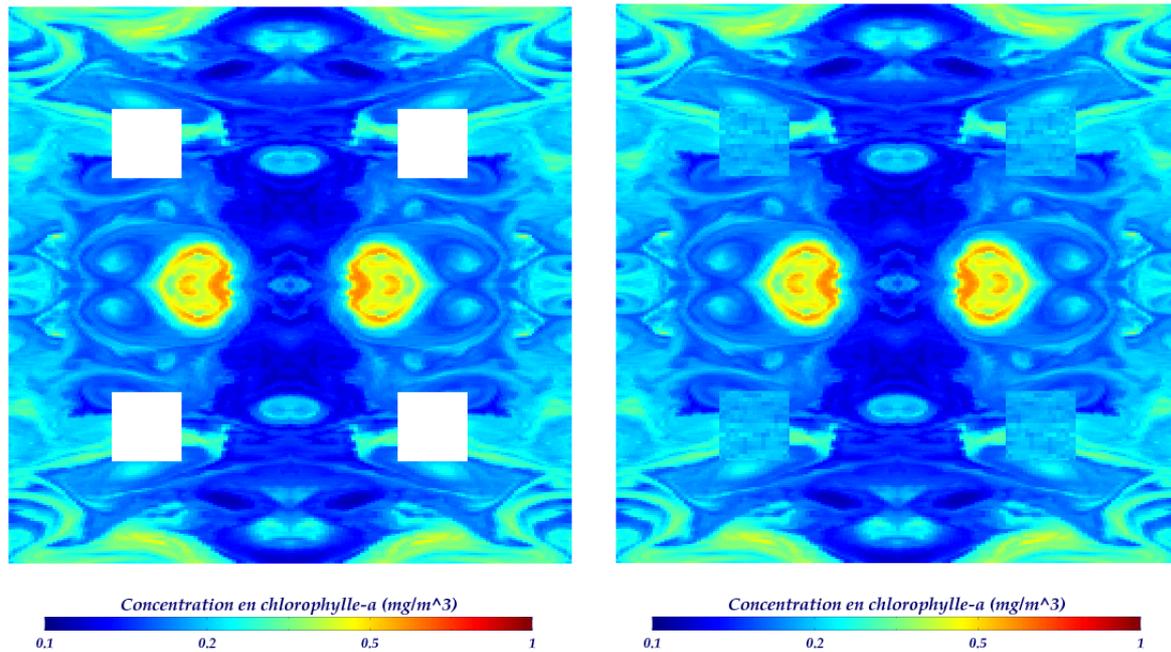


Figure 4.21 – A gauche : image de travail ; il s’agit de l’image que l’on souhaite reconstruire à laquelle on fait subir 2 symétrisations consécutives (horizontale puis verticale). A droite : image de travail dans laquelle les pixels manquants ont pris la valeur de la climatologie mensuelle correspondante.

La seconde étape de notre prétraitement consiste à donner artificiellement une valeur « sensée » aux données manquantes car l’algorithme pyramidal ne peut s’appliquer qu’à des images complètes. En effet, il est peu judicieux d’imposer dans les trous, sans autre traitement, une valeur comme la moyenne ou la climatologie par exemple, car les contours du trou seraient alors très fortement marqués et donc très visibles dans l’analyse en ondelettes. Pour cela, nous avons donc mis au point une méthode basée sur les travaux de Turiel et del Pozo [2002] que nous appellerons « méthode du gradient » par la suite. Cette méthode consiste justement à « gommer », en quelque sorte, ces contours trop nets laissés entre les valeurs imposées arbitrairement et les valeurs d’origine et, l’outil le plus adapté pour résoudre ce problème est le **gradient**. En pratique, le procédé se fait en deux temps. Dans un premier temps, le gradient du masque de trous de l’image (i.e. une image constituée de « 1 » là où l’image étudiée a des trous, « 0 » sinon) est calculé en utilisant la transformée de Fourier selon la formulation suivante :

$$\begin{cases} \text{grad}_x I = \text{TF}^{-1}\{\text{TF}\{I\}dx\} & \text{avec } dx = e^{i2\pi\frac{x}{N_x}} - 1 \text{ et } x \in [0, N_x - 1] \\ \text{grad}_y I = \text{TF}^{-1}\{\text{TF}\{I\}dy\} & \text{avec } dy = e^{i2\pi\frac{y}{N_y}} - 1 \text{ et } y \in [0, N_y - 1] \end{cases}$$

où « TF » désigne la transformée de Fourier (l’exposant « -1 » indique l’inversion de la transformée de Fourier), « grad_x » (resp. « grad_y ») désigne le gradient horizontal (resp.

vertical) de l'image I et N_x (resp. N_y) est le nombre de pixels horizontaux (resp. verticaux) de l'image. On obtient alors deux images (une pour chacune des deux directions, horizontale et verticale) présentant des « 1 » symbolisant des passages pixel \rightarrow trou dans la direction associée, des « -1 » symbolisant des passages trou \rightarrow pixel dans la direction associée et des « 0 » là où on n'observe pas de changement (cf. Figure 4.22 dans le cas de notre exemple).

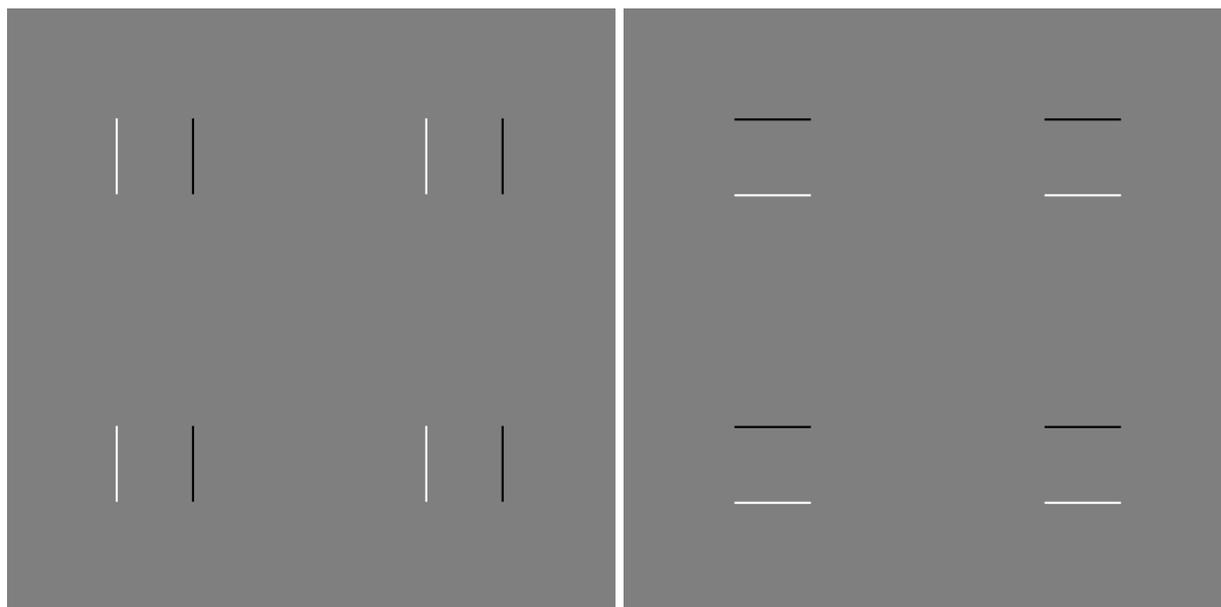


Figure 4.22 – Gradients horizontal (à gauche) et vertical (à droite) correspondant au masque de trous de l'image de travail montrée Figure 4.21. Le noir indique la valeur « -1 », le gris « 0 » et le blanc « 1 ».

Dans un deuxième temps, on calcule le gradient de l'image d'étude, i.e. l'image de travail (Figure 4.21) dans laquelle les valeurs de la climatologie mensuelle associée (janvier pour notre exemple) ont été imposée pour les pixels inconnus : on obtient, pour notre exemple, alors les images de gauche sur la Figure 4.23. Pour lisser les contours du trou, il suffit de mettre à « 0 » les gradients horizontal et vertical de l'image d'étude, correspondant aux gradients du masque de trous (Figure 4.22 ; cercles rouges, Figure 4.23) : dans notre exemple, on obtient donc les images de droite de la Figure 4.23.

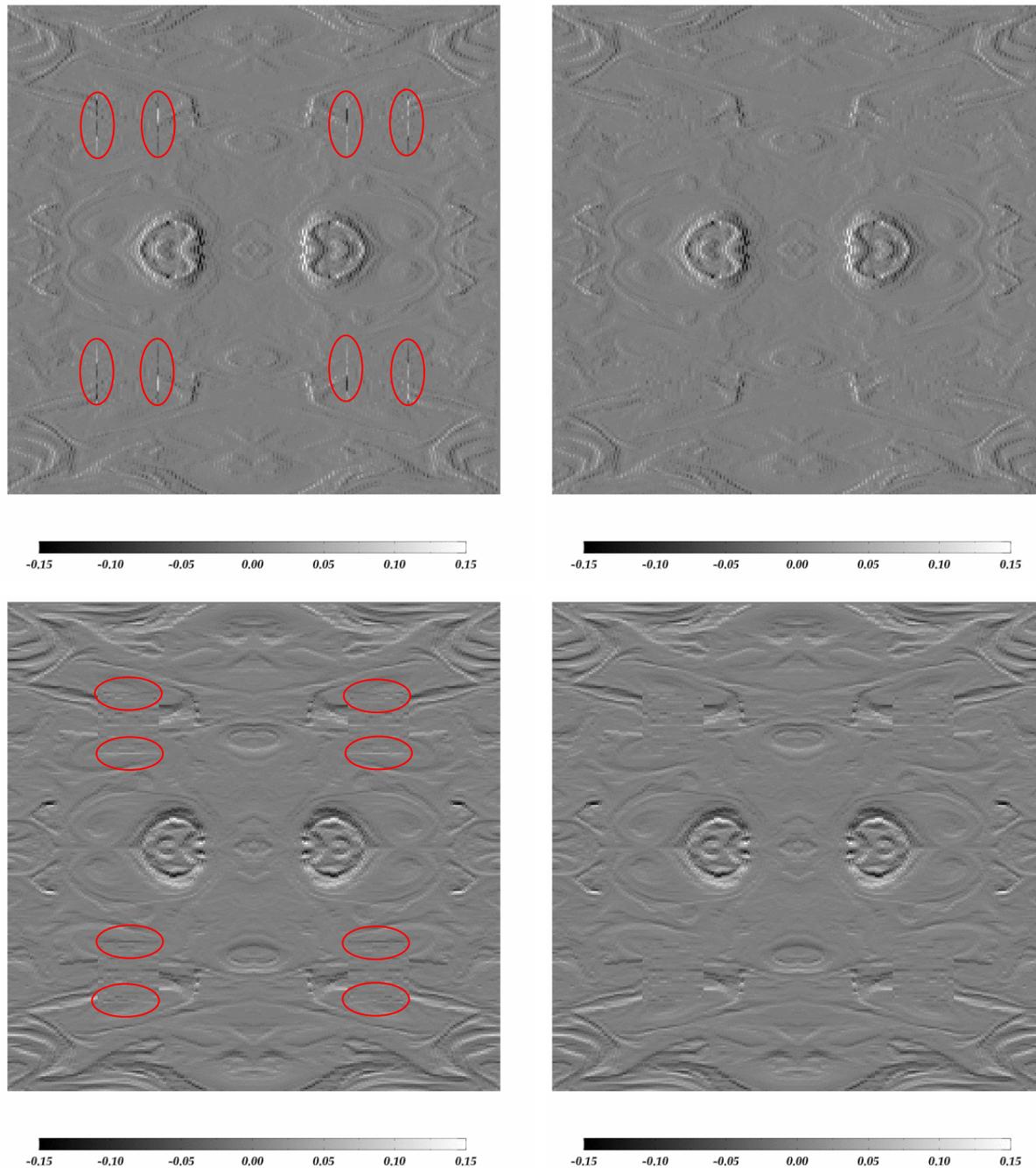


Figure 4.23 – Gradients horizontal (en haut) et vertical (en bas) de l'image d'origine (à gauche) et de l'image retouchée (à droite).

Ensuite, il suffit de reconstruire l'image à partir du gradient ainsi calculé, soit, toujours par la transformée de Fourier et en reprenant les notations vues plus haut :

$$I_{\text{est}} = \text{TF}^{-1} \left\{ \frac{\text{grad}_x I_{\text{est}} \cdot dx + \text{grad}_y I_{\text{est}} \cdot dy}{|dx|^2 + |dy|^2} \right\}.$$

Il ne faut pas oublier qu'à cause de la division de la formule ci-dessus, il faut annuler le coefficient correspondant à la fréquence 0 dans l'espace de Fourier. Or, cette fréquence représente l'intégrale de la fonction sur tout son domaine de définition. Par conséquent, après reconstruction, on obtient une image à moyenne nulle, c'est-à-dire que les valeurs de l'image sont centrées autour de 0 ; il faut donc rajouter à l'image reconstruite cette « valeur moyenne » pour qu'elle se rapproche le plus possible de l'image d'origine (i.e. avant reconstruction par la méthode du gradient). Cette valeur est déterminée par moindres carrés, de façon à minimiser l'erreur entre les pixels reconstruits et leur valeur à l'origine. Dans notre exemple, cette valeur moyenne vaut 0.2 mg/m^3 et l'image alors obtenue est celle de la Figure 4.24 (à gauche). L'estimation des pixels manquants par la méthode du gradient influe quelque peu (jusqu'à 15%) sur les pixels avoisinant le trou, comme on peut le voir Figure 4.24 (à droite) pour notre exemple. Néanmoins, on peut constater que l'erreur définie par le rapport entre la variance de la différence entre l'image reconstruite et l'image d'origine, et la variance de l'image originale est très faible et s'élève à 0.3% pour notre exemple.

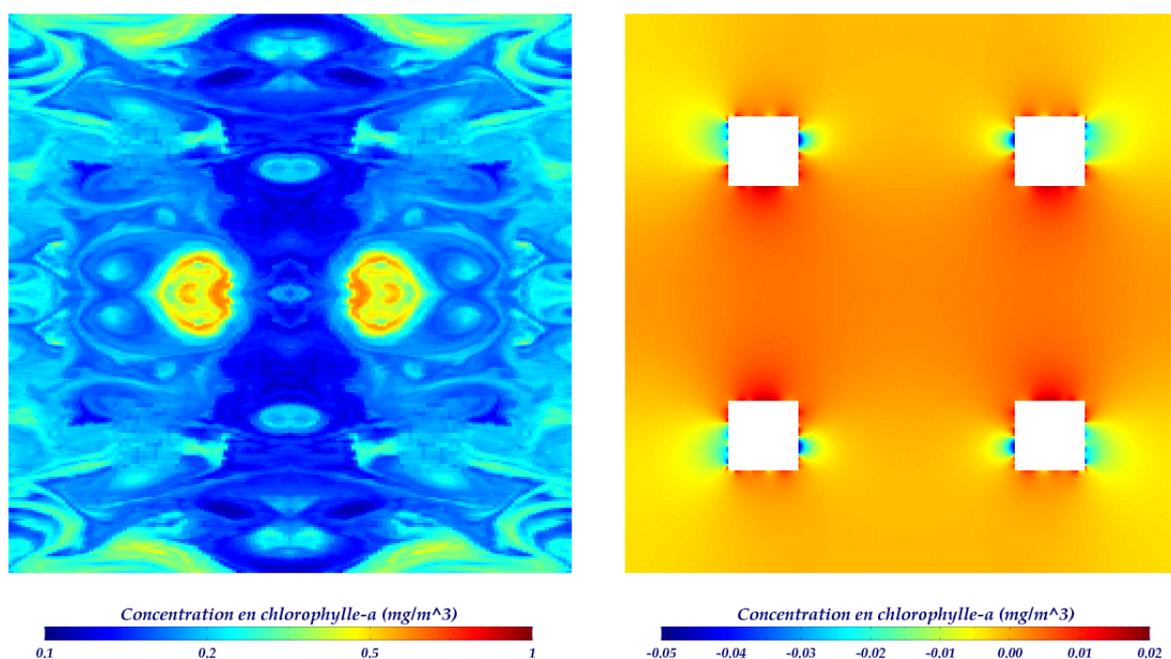


Figure 4.24 – A gauche : image de la Figure 4.21 dont les pixels manquants ont été estimés par la méthode des gradients. A droite : différence (en mg/m^3) entre l'image de gauche et celle de la Figure 4.21.

Maintenant que nous avons rempli « intelligemment » les pixels manquants de notre image de travail, nous pouvons nous intéresser à la phase de traitement proprement dit de notre algorithme de reconstruction. Celle-ci est constituée de 4 étapes.

La première consiste à appliquer l'**algorithme pyramidal** à notre image reconstruite, sur toutes les échelles possibles (au nombre de 8 dans notre exemple, car $2^8=256$, qui est la taille de notre image de travail). On obtient la pyramide des coefficients montrée Figure 4.26 (à gauche).

La seconde étape repose sur l'**inférence** coefficients d'ondelette fils \rightarrow coefficient d'ondelette père. Pour chaque passage d'une échelle j à l'échelle inférieure $j+1$, on va tout d'abord obtenir par inférence chaque coefficient père $\alpha_{j+1,\bar{k}}^{\text{inf}}$ (\bar{k} désignant la localisation spatiale du coefficient $\alpha_{j+1,\bar{k}}^{\text{inf}}$ pour l'échelle $j+1$) connaissant ses 4 coefficients fils $\alpha_{j,\bar{k}}^{\text{inf}}$ (cf. méthode détaillée au paragraphe 2.2.3). Le nouveau coefficient père $\alpha_{j+1,\bar{k}}^1$ qui remplace l'ancien $\alpha_{j+1,\bar{k}}^0$ dans la pyramide des coefficients d'ondelette est ensuite calculé comme suit :

$$\alpha_{j+1,\bar{k}}^1 = (1 - \rho_{j+1,\bar{k}}) \alpha_{j+1,\bar{k}}^0 + \rho_{j+1,\bar{k}} \alpha_{j+1,\bar{k}}^{\text{inf}}$$

où $\rho_{j+1,\bar{k}}$ désigne l'influence du trou sur le coefficient à l'échelle $j+1$ et à la position \bar{k} . En pratique, les coefficients $\rho_{j+1,\bar{k}}$ sont les coefficients d'approximation obtenus par application de l'algorithme pyramidal au masque des trous de l'image originale (i.e. image constituée de « 1 » là où l'image étudiée a des trous, « 0 » sinon). La Figure 4.25 montre ces coefficients dans le cas de notre exemple.

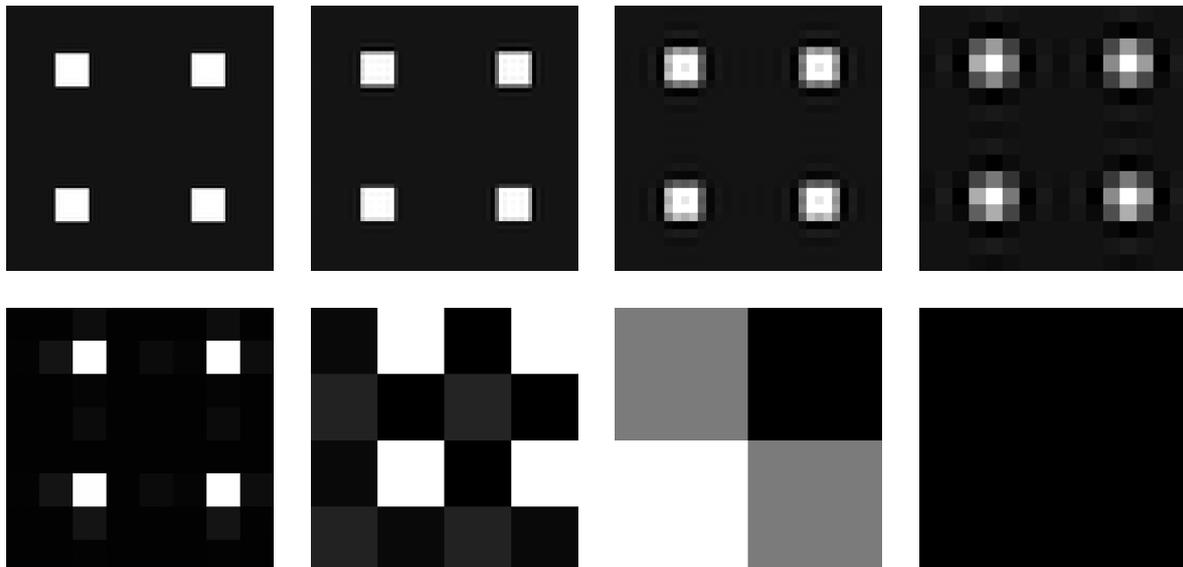


Figure 4.25 – Coefficients d'influence du trou sur le calcul du coefficient d'ondelette pour notre exemple. Les 8 échelles sont dans l'ordre, de haut en bas et de gauche à droite. Ondelette utilisée : Battle-Lemarié d'ordre 3 (41 coefficients retenus).

Ainsi, $\rho_{j+1,\bar{k}} = 0$ si le calcul de $\alpha_{j+1,\bar{k}}^n$ n'est influencé par aucun pixel manquant et $\rho_{j+1,\bar{k}} = 1$ si, au contraire, le calcul de $\alpha_{j+1,\bar{k}}^n$ n'est fait qu'à partir de pixels manquants. Et par conséquent, un coefficient $\alpha_{j+1,\bar{k}}^n$ qui n'est pas influencé par un pixel manquant reste inchangé.

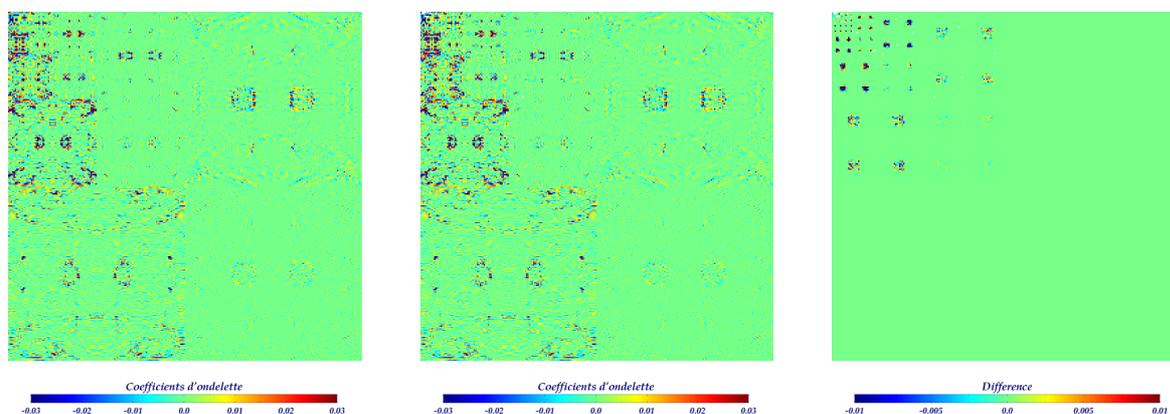


Figure 4.26 – A gauche : pyramide des coefficients d’ondelette originaux pour notre exemple. Au milieu : pyramide des coefficients d’ondelette après une modification. A droite : différence entre la seconde image et la première.

Cette estimation par inférence est faite par récurrence sur toutes les échelles de la pyramide. La Figure 4.26 montre la pyramide des coefficients d’ondelette pour notre exemple avant et après modification ; sur l’image de différence, on voit nettement les changements apportés par l’inférence pour les coefficients influencés par le trou.

La troisième étape de notre traitement consiste à reconstruire l’image par transformée en ondelette inverse de notre pyramide de coefficients ainsi estimés. Pour notre exemple, on obtient ainsi l’image de la Figure 4.27 (en haut à gauche).

On peut voir sur la Figure 4.27 (en haut à droite) que les valeurs de l’image alors obtenues sont écartées des valeurs de l’image d’origine. Pour palier à ce problème, la quatrième et dernière étape de notre traitement consiste en une correction de l’image reconstruite par la méthode du gradient. Pour cela, on va tout d’abord reconstruire par la méthode du gradient l’image différence entre l’image reconstruite (pour notre exemple : Figure 4.27, en haut à gauche) et l’image originale (pour notre exemple : Figure 4.21) : il s’agit en fait de l’image de la Figure 4.27 (en haut à droite) pour notre exemple. Puis on va déduire cette image différence reconstruite de l’image obtenue lors de la troisième étape de notre traitement. Pour notre exemple, on obtient alors l’image de la Figure 4.27 (en bas à gauche), qui présente nettement moins de différence avec la carte d’origine pour les pixels existants, comme le montre la Figure 4.27 (en bas à droite).

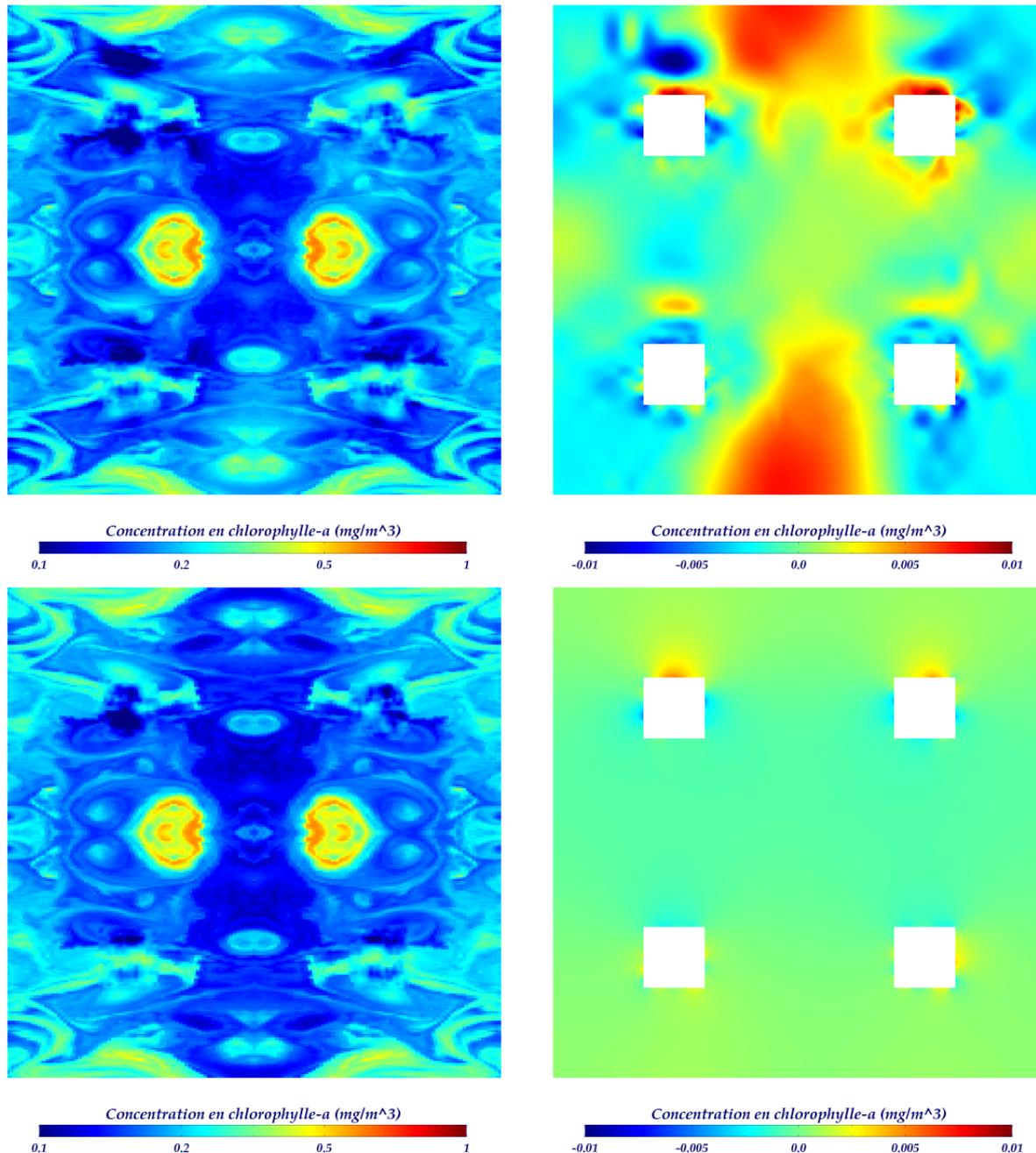


Figure 4.27 – En haut : image reconstruite avant correction par la méthode du gradient (à gauche) et la carte de différence par rapport à l’image d’origine montrée Figure 4.21 (à droite). En bas : même chose après correction.

Les 4 étapes qui constituent notre chaîne de traitement sont effectuées de manière itérative, jusqu’à ce que l’erreur définie par le rapport entre la variance de la différence entre l’image reconstruite et l’image d’origine, et la variance de l’image originale soit inférieure à 1%. Bien évidemment, si on constate une divergence (i.e. si l’erreur augmente au lieu de diminuer), on garde la reconstruction avec l’erreur la plus petite. Pour notre exemple, la carte reconstruite finalement obtenue est montrée Figure 4.28 (à gauche). L’erreur sur les valeurs

des pixels existants est de 0.74%, ce qui est très faible. A l'œil nu, on peut constater que l'estimation réalisée, en comparaison à l'image originale, est très satisfaisante. La Figure 4.28, à droite, montre le résultat obtenu quand on impose la valeur moyenne et non la climatologie mensuelle au tout début de notre analyse (i.e. avant la méthode du gradient) : on obtient une reconstruction nettement plus floue, montrant l'impact que peut avoir cette contrainte sur le résultat final.

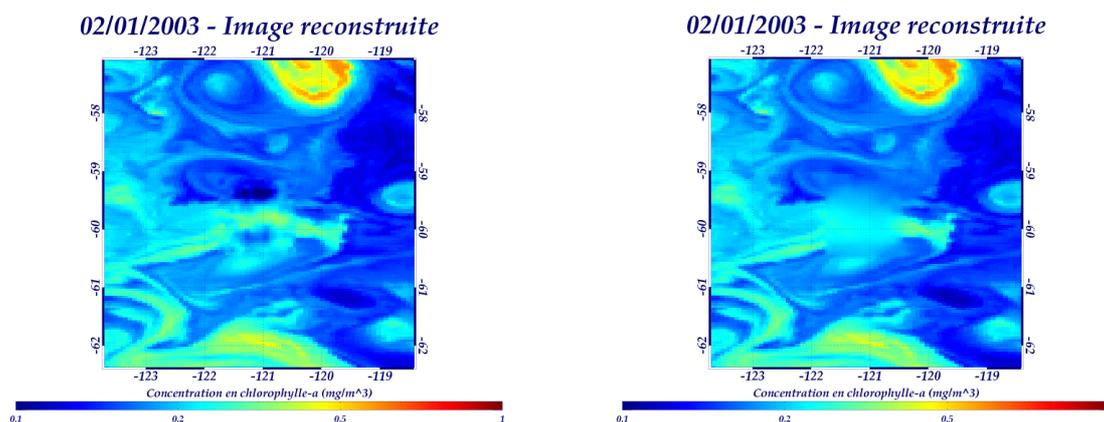


Figure 4.28 – A gauche : image reconstruite quand on impose, au début de l'algorithme, que les pixels manquants prennent la valeur de la climatologie mensuelle correspondante. A droite : image reconstruite quand on impose, au début de l'algorithme, que les pixels manquants prennent la valeur de la moyenne de l'image.

2.2.5 - Tests

Des tests ont été réalisés, dans l'objectif de déterminer l'ondelette et le type des données (concentration en chlorophylle-*a* ou son logarithme base 10) les plus adaptées à notre problème. Nous avons vu dans la partie 2.2.2 que seules les ondelettes de Battle-Lemarié et les Symmlets vérifiaient les statistiques théoriques pour nos données, qu'elles soient appliquées aux données de chlorophylle ou à leur logarithme base 10. Nous avons donc appliqué l'algorithme de reconstruction à l'ensemble de nos 887 imageries de tests dans lesquelles nous avons créé artificiellement un trou identique à celui de notre exemple Figure 4.19 (à droite), en utilisant 6 ondelettes différentes (Battle-Lemarié d'ordre 2, 3 et 4, indiqué « BattleLemarie_Ordre_NbCoeffs » par la suite ; Symmlet d'ordre 6, 7 et 8, indiqué « Symmlet_Ordre » par la suite) sur des données de chlorophylle (« DEC ») ou leur logarithme base 10 (« LOG »). Les images de l'Annexe A-3 montrent les résultats de la reconstruction de l'image de la Figure 4.19 (à droite) pour nos 12 combinaisons test.

Pour départager ces 12 cas, nous avons comparé l'erreur définie par le rapport entre la variance de la différence entre l'image reconstruite et l'image d'origine, et la variance de l'image originale, dans le trou et hors du trou, ainsi que la convergence de l'algorithme. On obtient les résultats montrés Figure 4.29.

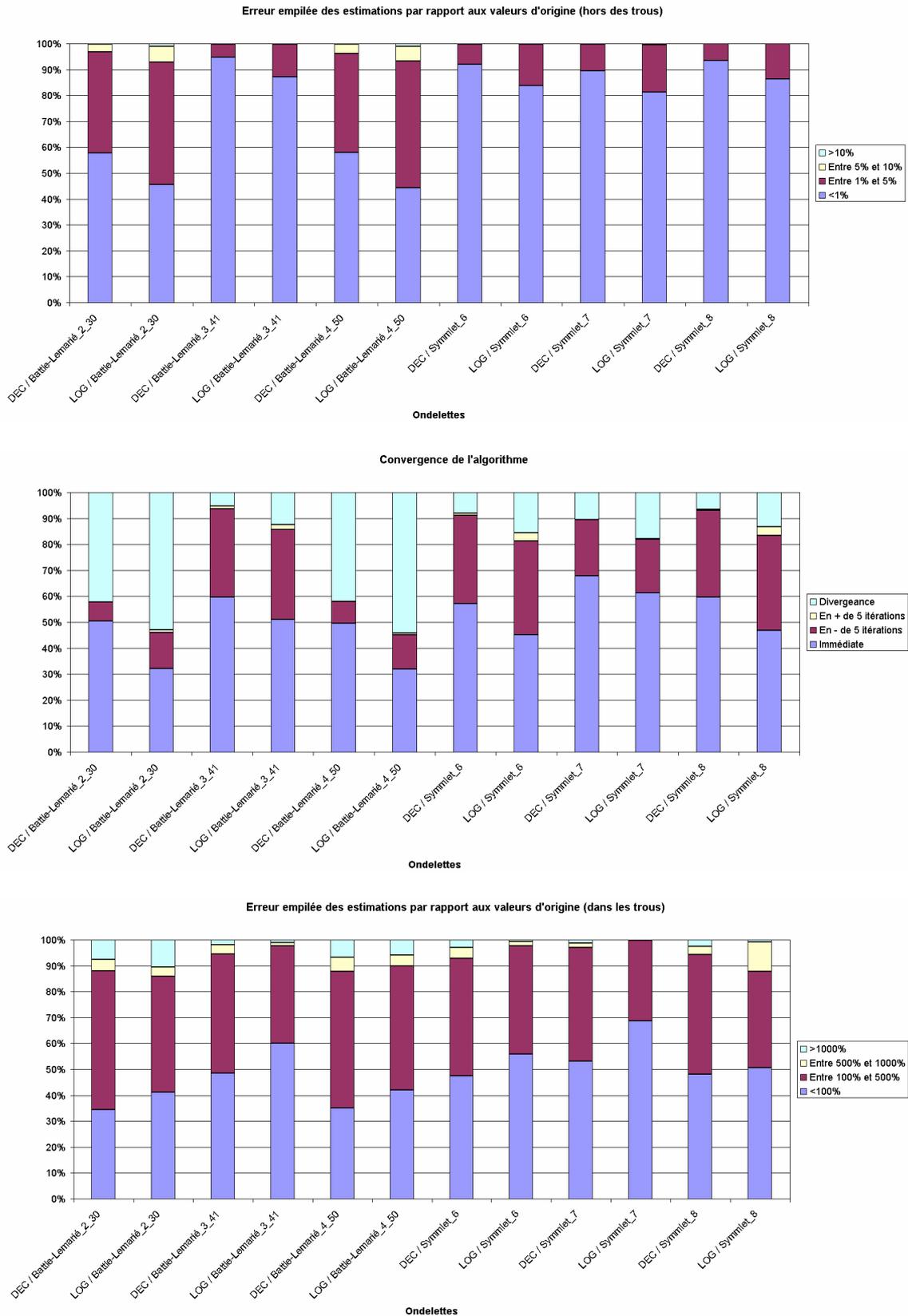


Figure 4.29 – Résultats de tests de reconstruction.

On peut tout d'abord constater que les ondelettes de Battle-Lemarié d'ordre 2 et 4 sont nettement moins bien adaptées à notre problème que les autres : les erreurs des estimations par rapport aux valeurs d'origine, que ce soit dans ou hors des trous, sont les plus élevées et l'algorithme diverge dans plus de 60% des cas, indiquant que ces ondelettes sont généralement instables pour notre étude. Si on s'intéresse aux 4 autres ondelettes, on peut remarquer que l'erreur entre les estimations et les valeurs d'origine hors du trou est supérieure dans le cas d'une application au logarithme base 10 de la chlorophylle par rapport à une application directe sur la chlorophylle (Figure 4.29, en haut). La convergence est aussi meilleure (Figure 4.29, au milieu).

On peut remarquer que la Symmlet d'ordre 7 appliquée au logarithme base 10 des données donne de bons résultats si on s'intéresse à la convergence et à l'erreur des estimations par rapport aux valeurs d'origine dans le trou. Néanmoins, si on regarde les 3 indicateurs, l'ondelette de Battle-Lemarié d'ordre 3 appliquée aux données de chlorophylle semble être la combinaison la plus adaptée. Néanmoins, pour éviter le risque d'avoir des concentrations en chlorophylle négatives, il est préférable de travailler sur le logarithme base 10 des données. La combinaison que nous avons décidé d'utiliser dans le cadre de notre problème est donc une **ondelette de Battle-Lemarié d'ordre 3 appliquée au logarithme base 10 de la chlorophylle**.

2.2.6 - Conclusion

Dans cette partie, nous avons vu la méthode que nous avons mise au point pour estimer les données manquantes dans une image de chlorophylle grâce à la transformée en ondelette. C'est une méthode en 2 étapes. La première consiste à imposer dans les trous la valeur de la climatologie mensuelle associée puis à « gommer » les contours ainsi artificiellement créés par la méthode des gradients. La seconde étape consiste à déterminer les coefficients d'ondelette père altérés par les trous en fonction de leurs coefficients d'ondelette fils correspondants. Des tests nous ont permis de déterminer l'ondelette la plus adaptée à notre problème : il s'agit de l'ondelette de Battle-Lemarié d'ordre 3 appliquée au logarithme base 10 des données de chlorophylle.

2.3 - Fusion d'images par ondelette

La méthode de fusion d'images par transformée en ondelette que nous allons appliquer ici est généralement utilisée pour combiner les images multispectrales avec des observations panchromatiques haute-résolution [Wald, 2002].

La première étape de cette méthode consiste à appliquer l'algorithme pyramidal sur chaque image à combiner : par exemple, les cartes SeaWiFS et MODIS/Aqua relatives au 26 mars 2003 si on souhaite récupérer une image combinée des deux capteurs pour ce jour-là. On a le choix de descendre la pyramide pour toutes les échelles, ou seulement jusqu'à un certain point. Quelques tests nous ont permis de voir qu'il semblait préférable de nous arrêter au premier niveau pour SeaWiFS et au second niveau pour MODIS/Aqua (i.e. même niveau de résolution).

Les coefficients d'ondelette sont ensuite combinés selon leur orientation (détails horizontaux, verticaux et diagonaux) et par échelle (cf. Figure 4.30). Nous avons choisi d'utiliser une moyenne pondérée par l'erreur capteur, tout en considérant les emplacements des pixels connus et inconnus pour chaque capteur. Dans le cas de la combinaison de SeaWiFS et MODIS/Aqua, il ne faut pas oublier que MODIS/Aqua a une résolution exactement deux fois plus fine que SeaWiFS : les coefficients de détails à l'échelle 1 des données combinées sont donc ceux de MODIS/Aqua uniquement.

La troisième et dernière étape consiste à retourner dans l'espace physique par transformée en ondelettes inverse. On obtient ainsi l'image combinée.

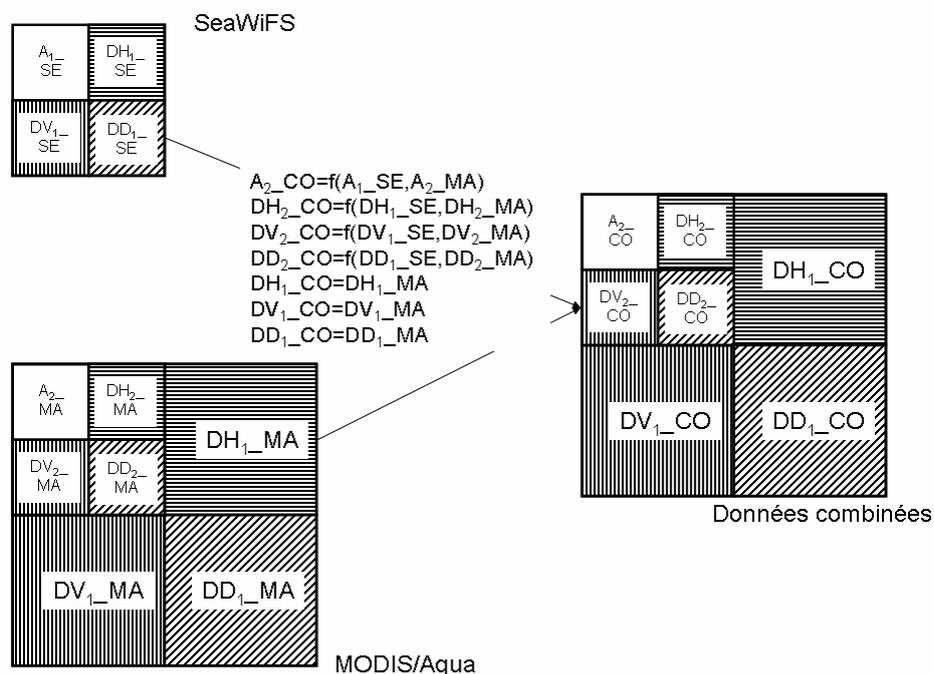


Figure 4.30 – Principe de la méthode de combinaison par ondelette.

Nous allons maintenant voir comment associer les méthodes de reconstruction et de fusion d'images pour combiner les cartes journalières de concentration en chlorophylle issues de plusieurs capteurs.

2.4 - Application à la combinaison des données de chlorophylle

La méthode envisagée consiste à reconstruire séparément chaque image de chlorophylle à combiner (i.e. relative à une journée et à un capteur) par la méthode montrée dans la partie 2.2, puis les combiner entre elles en combinant leurs coefficients d'ondelette par moyenne pondérée par l'erreur capteur (cf. partie 2.3). L'algorithme est présenté Figure 4.31.

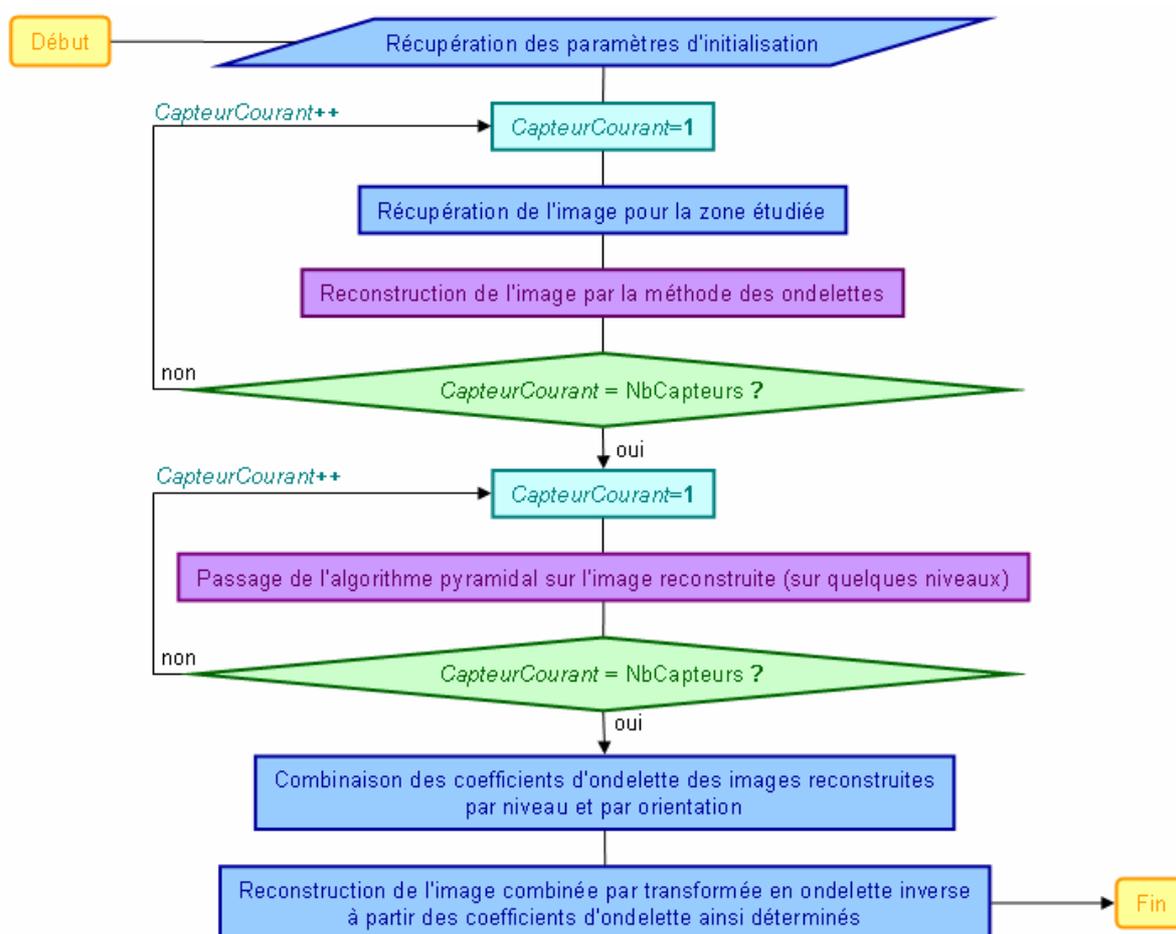


Figure 4.31 - Organigramme de l'algorithme de combinaison des données de concentration en chlorophylle-*a* par ondelettes.

La Figure 4.32 montre le résultat obtenu pour la combinaison des données issues de SeaWiFS (image 64 x 64 pixels) et MODIS/Aqua (image 128 x 128 pixels) le 26 mars 2003 dans la zone du Gulf Stream. Pour indication, le temps de calcul pour cette combinaison (image de 5° x 5° environ) est de 10 minutes environ pour un processeur 700 MHz ; c'est la

reconstruction qui est très gourmande, la combinaison en elle-même est extrêmement rapide (quelques secondes).

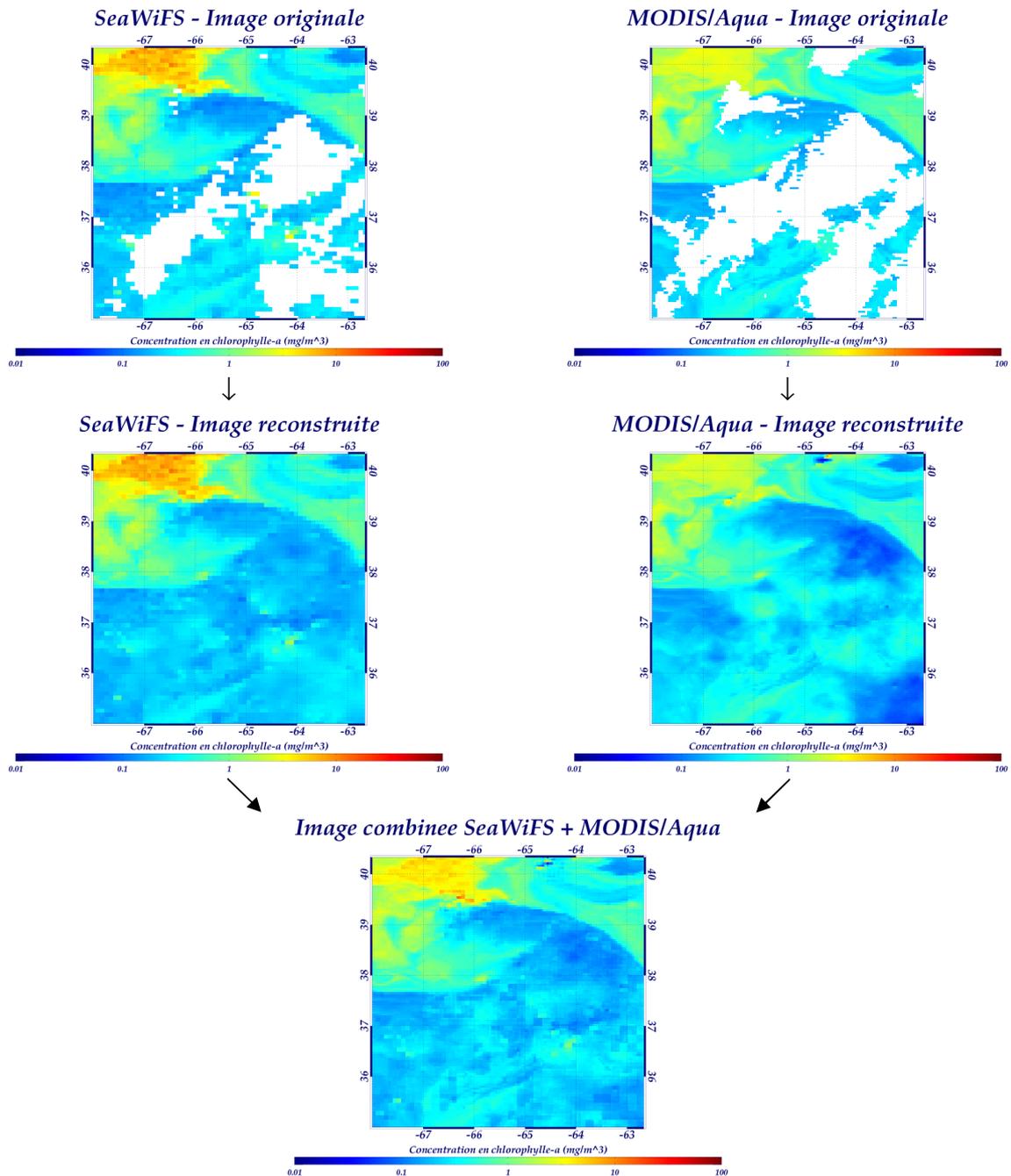


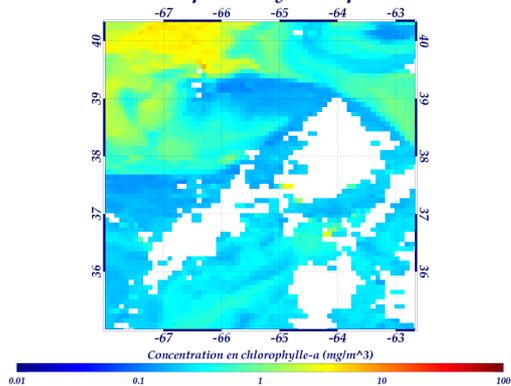
Figure 4.32 – Combinaison des données SeaWiFS (en haut à gauche) et MODIS/Aqua (en haut à droite) par ondelettes : en bas, résultat de l'application au logarithme base 10 des données de concentration en chlorophylle-*a* pour le 26 mars 2003. Les images au milieu sont le résultat de l'algorithme de reconstruction appliqué aux images originales (en haut).

3 - Conclusion du chapitre

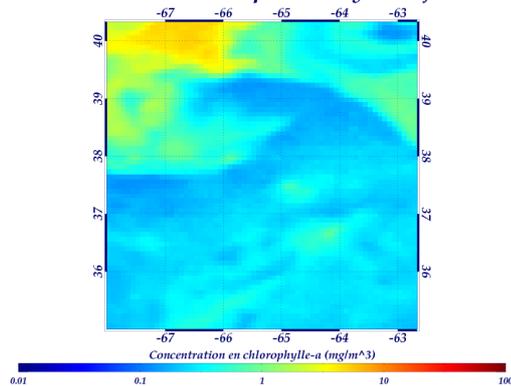
Dans ce chapitre, après avoir rappelé les notions fondamentales concernant les ondelettes, nous avons explicité la méthode mise en place pour la combinaison des données de chlorophylle. Elle s'articule autour de deux axes. La première étape est une estimation des données manquantes pour chaque image (i.e. relative à une journée et à un capteur). Nous avons tout d'abord introduit la théorie ayant mené au prédicteur mis en place avant de détailler la méthodologie en elle-même. Des tests ont ensuite été réalisés dans l'objectif de déterminer la configuration ondelette / type de donnée la plus adaptée à notre problème : il s'agit de l'ondelette de Battle-Lemarié d'ordre 3 appliquée au logarithme base 10 de chlorophylle. La deuxième étape de notre méthode de combinaison est la fusion des images reconstruites : elle consiste à combiner les coefficients d'ondelette des différentes images à combiner par niveau et par orientation, avant d'effectuer une transformée en ondelette inverse permettant de récupérer l'image des données combinées.

En guise de première application, nous avons effectué la combinaison de SeaWiFS et MODIS/Aqua pour le 26 mars 2003 dans la zone du Gulf Stream. La Figure 4.33 montre la comparaison entre les trois méthodes de combinaison de données dont il est question dans ce travail de thèse. D'une part, la moyenne pondérée ne tient compte que des données existantes, contrairement aux deux autres méthodes. D'autre part, l'image combinée par analyse objective est plus lissée (visuellement, les contours sont moins nets) que les images combinées obtenues par les deux autres algorithmes. La combinaison par ondelettes bénéficie donc des deux points forts : la couverture spatiale est améliorée au maximum tout en conservant les structures le plus possible. De plus, l'avantage supplémentaire de cette méthode est que l'on récupère une image combinée de la résolution la plus fine et non de la résolution la plus grossière comme c'est le cas pour les deux autres.

Donnees combinees par moyenne ponderee (LOG12)



Donnees combinees par analyse objective



Donnees combinees par ondelettes

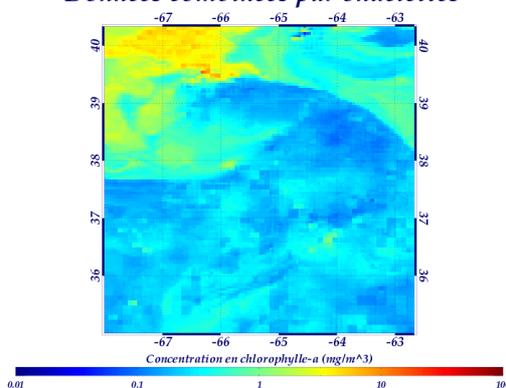


Figure 4.33 – Combinaison des données SeaWiFS et MODIS/Aqua par moyenne pondérée / algorithme LOG12 (en haut à gauche), analyse objective (en haut à droite) et ondelettes (en bas).

Chapitre 5

Application océanographique :
les Ondes Dominantes de
l'Océan Austral

CHAPITRE 5

APPLICATION OcéANOGRAPHIQUE :

LES ONDES DOMINANTES DE L'OcéAN AUSTRAL

Cette étude a pour objectif de déterminer les ondes dominantes de l'Océan Austral (35°S – 50°S) à partir des données de concentration en chlorophylle et de comparer les résultats obtenus avec les données altimétriques (Pottier *et al.* [2004]), dans le but de déterminer un lien éventuel entre physique et biologie. Ce travail a été réalisé en collaboration avec Jean-Pierre CERON du Centre Météorologique de Toulouse, qui a développé une méthode basée sur les travaux d'Hayashi [1977 et 1979] utilisée pour cette étude [Céron *et al.*, 2000]. Celle-ci permet de caractériser les propriétés stationnaire et / ou propagative des ondes zonales et / ou méridionales présentes dans les données analysées.

Après une description de la zone d'intérêt et des objectifs de cette étude, la méthode d'analyse est détaillée. L'application aux données altimétriques et de concentration en chlorophylle est ensuite traitée, suivie d'une discussion concernant les résultats obtenus.

Sommaire

1 – DESCRIPTIF DE LA ZONE D'ETUDE ET PROBLEMATIQUE.....	201
2 – METHODES D'ANALYSE	205
3 – APPLICATION AUX DONNEES ALTIMETRIQUES	209
3.1 – <i>Introduction</i>	209
3.2 – <i>Les données utilisées</i>	210
3.3 – <i>Résultats</i>	210
3.4 – <i>Discussion et conclusion</i>	216
4 – APPLICATION AUX DONNEES DE CHLOROPHYLLE	219
4.1 – <i>Introduction</i>	219
4.2 – <i>Application aux données mensuelles issues de SeaWiFS</i>	219
4.3 – <i>Application aux données hebdomadaires issues de la combinaison de SeaWiFS et MODIS/Aqua</i>	225
4.4 – <i>Discussion et conclusion</i>	231
5 – CONCLUSION DU CHAPITRE.....	237

1 - Descriptif de la zone d'étude et problématique

La zone d'étude est l'**Océan Austral**. Cet anneau océanique autour du continent Antarctique va de 66°70'S à 40°S environ, soit une largeur de 26° à 30° de latitude. Il présente un resserrement notable en deux régions : au Sud de l'Amérique et au Sud de l'Australie. Alors que le continent africain ne s'étend que jusqu'à 35°S, le continent australien (voir 4, Figure 5.1) et sa prolongation, la Tasmanie (**5**), vont jusqu'à 43°S, et le continent américain (**2**) s'avance jusqu'à 56°S. Or, dans cette région, au Sud du continent américain, le continent Antarctique (**1**), dont les limites connues sont en moyenne assez voisines du cercle polaire (66°33'S), présente une avancée vers le nord : la péninsule Antarctique (**10**), qui atteint 63°S.

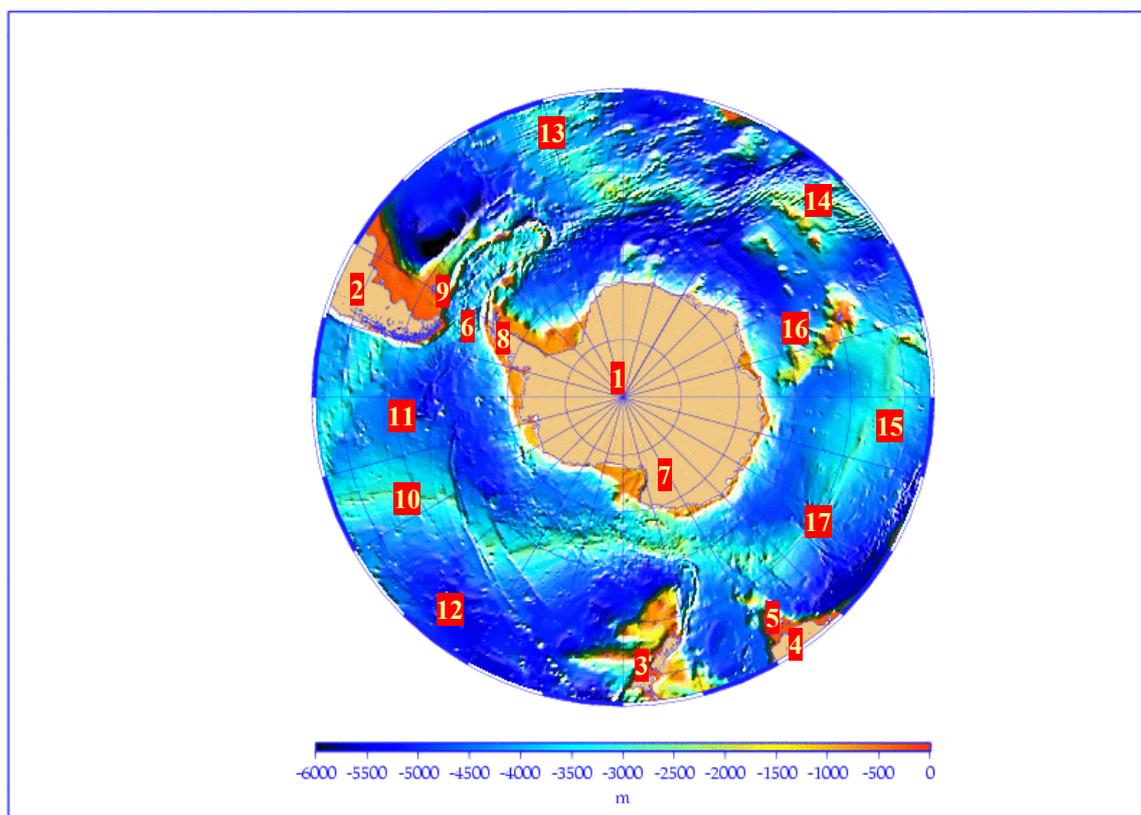


Figure 5.1 – La bathymétrie dans l'Océan Austral. 1 : continent Antarctique ; 2 : Amérique du Sud ; 3 : Nouvelle-Zélande ; 4 : Sud de l'Australie ; 5 : Tasmanie ; 6 : passage de Drake ; 7 : Terre Adélie ; 8 : péninsule Antarctique ; 9 : îles Falkland ; 10 : dorsale pacifique du S-E ; 11 : bassin pacifique du S-O ; 12 : bassin pacifique du S-O ; 13 : dorsale médiane atlantique ; 14 : dorsale indienne ; 15 : bassin indien ; 16 : plateau des Kerguelen ; 17 : dorsale antarctique indienne.

Dans le secteur américain, le Passage de Drake (6) ne fait que 420 milles nautiques¹ de largeur environ. Le resserrement au Sud de l'Australie (4) est beaucoup moins marqué ; la distance entre le cap S-E de la Tasmanie (5) et la Terre Adélie (7) est d'environ 1400 milles nautiques. La distance entre la pointe extrême Sud de l'Afrique et le continent Antarctique, sur le méridien 20°E, est de 2150 milles nautiques. Ainsi définie entre 40°S et 70°S, la superficie de l'Océan Austral est d'environ 77 millions de kilomètres carrés, soit 22% de la superficie totale de l'océan mondial. La Figure 5.1 montre la bathymétrie² de la zone d'étude. L'anneau océanique autour du continent Antarctique est subdivisé en trois bassins principaux, profonds de 4000 à 6500 mètres, séparés les uns des autres par des dorsales (Figure 5.1).

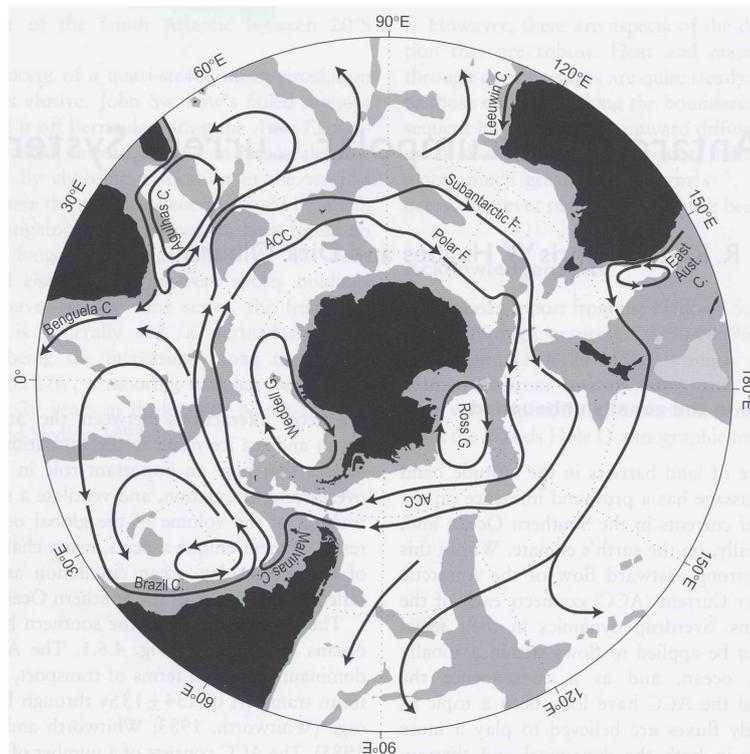


Figure 5.2 – Principaux courants des océans au sud de 20°S. Les profondeurs supérieures à 3500 m sont en grisé. Les abréviations signifient : F pour front, C pour courant et G pour gyre. [Rintoul *et al.*, 2001]

L'Océan Austral est le seul domaine océanique encerclant le globe [Rintoul *et al.*, 2001] (Figure 5.2). Le **Courant Circumpolaire Antarctique** (ACC : Antarctic Circumpolar Current), principal courant sur cette bande de latitude, s'y propage vers l'est : le transport

¹ 1 mille nautique = 1852 m.

² Représentation cartographique des reliefs sous-marins.

moyen est de $134 \pm 13 \text{ Sv}^1$ à travers le Passage de Drake [Whitworth, 1983 ; Whitworth et Peterson, 1985]. L'ACC est composé de plusieurs fronts circumpolaires, dont le rôle est de séparer les masses d'eau ; les fronts subantarctique et polaire sont les deux fronts principaux. Plus au sud de l'ACC, on peut trouver des gyres cycloniques dans les mers de Weddell et de Ross. Il existe aussi un courant se propageant vers l'ouest le long de la côte du continent antarctique.

L'ACC est le lien permettant les échanges de masses d'eau, de chaleur, de sel et de traceurs à toutes les profondeurs entre les principaux bassins océaniques mondiaux (Pacifique, Atlantique et Indien). Comme ces échanges représentent un contrôle important sur le climat global moyen, l'Océan Austral joue probablement un rôle important dans la transmission des anomalies climatiques tout autour du globe. Une variabilité interannuelle a souvent été observée à des latitudes sud proches du pôle et l'observation de l'extension des glaces des mers suggère des caractéristiques propagatives vers l'est autour de l'Océan Austral.

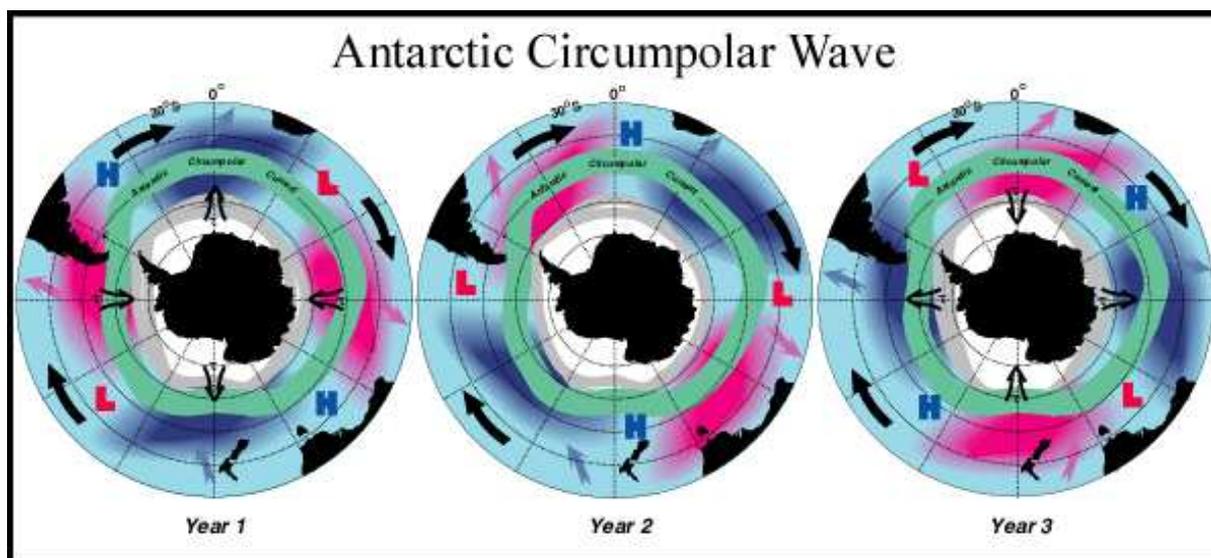


Figure 5.3 – Evolution de l'onde circumpolaire [White, 2003].

Dans les données de température de surface de la mer (SST : Sea Surface Temperature), de pression atmosphérique de surface de mer (SLP : Sea Level Pressure), de coups de vent méridionaux et d'extension de la glace de mer, White et Peterson [1996] ont repéré des anomalies se propageant autour du continent Antarctique avec une période de 4 à 5 ans et une vitesse d'environ 8 cm/s. Ces anomalies simultanées, associées au **nombre**

¹ $1 \text{ Sv} = 10^6 \text{ m}^3/\text{s}$.

d'onde zonal 2, ont été désignées sous le nom d'**Onde Circumpolaire Antarctique** (ACW : Antarctic Circumpolar Wave ; Figure 5.3). A la même époque, étudiant des données de hauteur de mer, Jacobs et Mitchell [1996] ont constaté un mode dominant caractérisé par une période de 4 ans environ et une longueur d'onde longitudinale de 180°. Utilisant 27 ans de données SST et SLP, Carril et Navarra [2001] ont montré la présence de deux nombres d'onde zonaux, 2 et 3, dans la structure de l'ACW durant les deux dernières décennies du siècle dernier, en se basant sur des données historiques de SST. La composante majeure de leur ACW était une structure de nombre d'onde 2 forcée par le pattern pacifique sud et partiellement liée à l'Oscillation Australe El Niño (ENSO : El Niño-Southern Oscillation). Récemment, Venegas [2003] a confirmé qu'une grande partie de la variance de la partie interannuelle de l'ACW pourrait être expliquée par une combinaison linéaire de deux signaux avec des caractéristiques spatio-temporelles différentes. Le premier, de période 5 ans, a un nombre d'onde zonal dominant 2 (ACW2) à travers l'Océan Austral. Il est indirectement forcé par le phénomène tropical ENSO avec une signature la plus forte dans le secteur Pacifique Est. Le second signal a un nombre d'onde zonal 3 avec une période de 3.3 ans.

Venegas [2003] a aussi noté que la structure SLP de l'ACW2 ressemble au mode dominant EOF (Empirical Orthogonal Function) de la variabilité atmosphérique que Thompson et Wallace [2000] ont dénommé SAM (Southern Annular Mode). Ce mode atmosphérique a été décrit en premier lieu par Hartmann et Lo [1998] et Watterson [2000] comme une vacillation des coups de vent d'ouest zonaux au niveau des latitudes moyennes dans l'Hémisphère Sud. Ce mode est associé à un déplacement Nord-Sud du jet aux moyennes latitudes qui se dirige vers l'ouest. Il présente une structure symétrique zonale autour du Pôle Sud, de signes opposés dans la région de la calotte polaire et autour d'un cercle zonal centré près de 45°S. L'importance du SAM pour la variabilité interannuelle dans l'Hémisphère Sud est toujours discutée. Hall et Visbeck [2002] suggèrent que « quand les observations adéquates pour toute la variabilité de l'Océan Austral existent, une grande partie de cette variabilité est, en fin de compte, calquée non sur l'ACW ou ENSO, mais sur le SAM ». Dans un article répondant à cette assertion, White [2003] conduit deux analyses indépendantes sur un jeu de 20 ans de données d'anomalies de SLP issues des réanalyses NCEP / NCAR, et montre qu'au contraire, les signaux ACW et ENSO dominant le SAM au niveau atmosphérique à des échelles temporelles interannuelles.

Mieux comprendre le fonctionnement de la pompe biologique et de la variabilité dans l'Océan Austral est à l'origine de cette étude. Des facteurs limitant le développement du phytoplancton (concentration en nutriments, quantité de lumière disponible) sont modulés

par la circulation océanique. Il est donc important de bien connaître les processus physiques qui vont affecter le développement du phytoplancton. Le but de cette étude est de mieux comprendre les interactions existant entre la physique et la biologie. Plus précisément, nous nous intéresserons à l'influence des ondes comme l'Onde Circumpolaire Antarctique sur l'activité biologique dans l'Océan Austral considéré comme une région de puits de CO₂ atmosphérique, encore mal quantifié. Dans cette optique, notre étude consiste à analyser l'évolution spatio-temporelle des structures liées aux ondes (en particulier l'Onde Circumpolaire Antarctique) à partir de deux types de données océaniques satellitaires :

- ❏ les données altimétriques (capteurs Topex/Poseidon, ERS, Jason et EnviSat) qui donnent la topographie dynamique des océans donc une information sur les courants,
- ❏ les données de concentration en chlorophylle dans les eaux de surface de l'océan (capteur SeaWiFS et produits combinés SeaWiFS-MODIS/Aqua).

Maintenant que nous avons vu les caractéristiques principales de notre zone d'étude et la problématique qui s'y rattache, nous allons nous intéresser aux méthodes d'analyse mises en place pour répondre à nos questions.

2 - Méthodes d'analyse

En océanographie ainsi qu'en météorologie, l'analyse spatio-temporelle des champs 2D tient une place importante pour l'étude des phénomènes physiques. Cependant, de nombreuses méthodes, telles que l'analyse spectrale spatio-temporelle, s'appliquent de manière globale, et non locale. Or, dans ces disciplines scientifiques, les processus physiques importants (mésoséchelle, onde, etc.) ont des caractéristiques souvent variables dans l'espace et le temps. Ces méthodes globales ne sont donc pas adaptées pour analyser ces phénomènes en raison de l'effet de moyenne sur l'échantillon analysé.

En météorologie, l'analyse spatio-temporelle des champs a connu de nombreuses applications, notamment grâce aux travaux d'Hayashi [1977, 1979]. Basés sur la cohérence spectrale entre ondes progrades et rétrogrades, ces calculs permettent de séparer la partie stationnaire de la partie propagative des phénomènes de nature ondulatoire. Le problème soulevé dans ce type de traitement est la nature changeante des phénomènes, tant du point de vue de leur structure spatiale que de leur caractéristique temporelle. Cette technique a été appliquée avec succès en météorologie par Jean-Pierre CERON. Elle a permis de faire ressortir les caractéristiques « instantanées » des signaux ondulatoires et ainsi de permettre

d'avoir une perception plus dynamique, et donc plus réaliste, de leur évolution [Céron *et al.*, 2000].

Deux méthodes ont été utilisées pour notre étude (cf. Figure 5.4) : l'analyse d'Hayashi [Hayashi, 1977, 1979] et une méthode nouvellement développée combinant une analyse de Gabor aux techniques d'Hayashi, l'analyse de Céron [Céron *et al.*, 2000]. Elles s'appliquent à des jeux de données 3D : longitude-latitude-temps. Ce qui va suivre consiste en un simple résumé de ces méthodes ; pour plus de détails, le lecteur pourra se référer aux articles originaux cités.

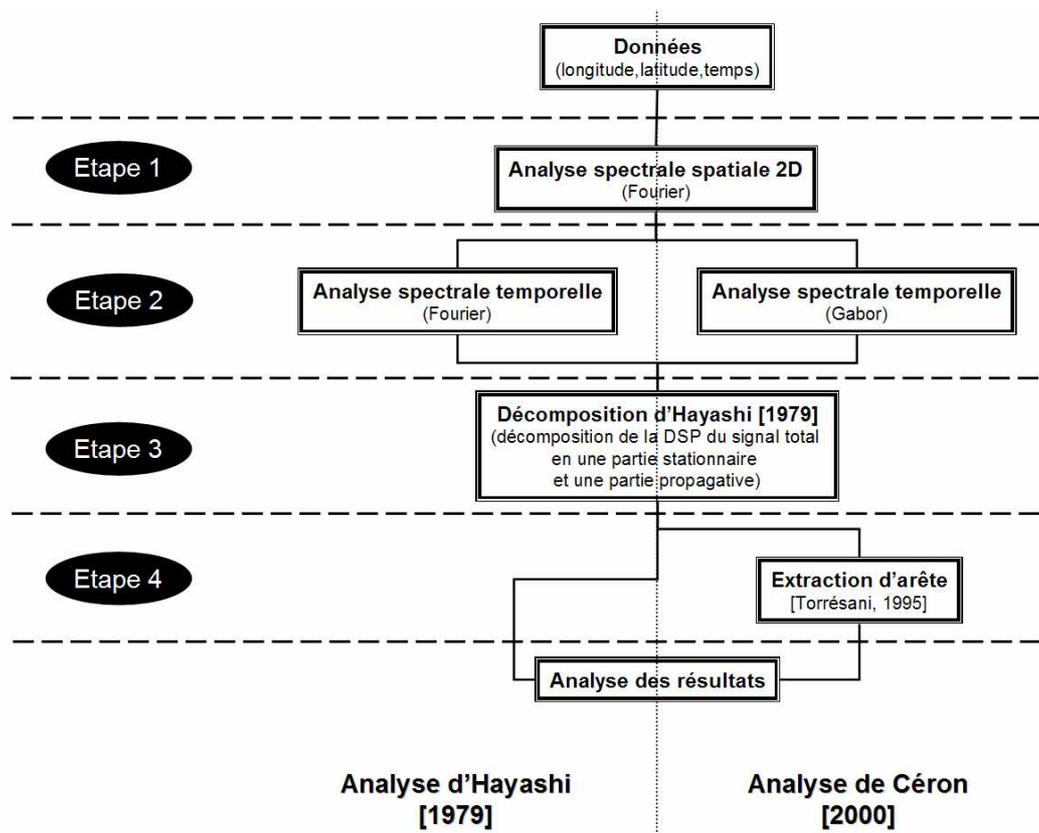


Figure 5.4 – Les étapes des analyses d'Hayashi (à gauche) et de Céron (à droite).

Une tendance bilinéaire est tout d'abord supprimée des données, avant de réaliser la première étape, commune aux deux méthodes : une analyse de Fourier 2D pour chaque carte longitude-latitude, c'est-à-dire autant de transformées de Fourier spatiales que d'observations disponibles. A l'issue de cette étape, on obtient donc, pour chaque observation, un jeu de coefficients spectraux spatiaux dépendant des nombres d'onde zonaux et méridionaux, i.e. le nombre de périodes qu'a un signal ondulatoire sur toute sa longueur selon les longitudes et les latitudes respectivement. On se place maintenant dans

le domaine spectral. Une sélection des principaux couples nombre d'onde zonal / nombre d'onde méridien est alors réalisée, par comparaison du module de chaque coefficient spectral avec un seuil déterminé à partir du niveau de bruit blanc : en effet, un couple est gardé à partir du moment où le module de l'un de ses coefficients spectraux spatiaux a été au moins une fois, sur toutes les observations, supérieur à ce seuil. Les étapes suivantes dépendent de la méthode utilisée.

Dans l'analyse d'Hayashi (cf. Figure 5.4, à gauche), la seconde étape consiste en une analyse spectrale temporelle de Fourier appliquée au signal constitué des variations temporelles des coefficients spectraux spatiaux, pour chaque couple nombre d'onde zonal / nombre d'onde méridien retenu. Enfin, le spectre de puissance spatio-temporel ainsi obtenu est divisé en deux parties, une stationnaire et une propagative [Hayashi, 1979]. Cette méthode permet ainsi de déterminer les variations de la Densité Spectrale de Puissance (DSP) en fonction des périodes pour le signal entier, la partie stationnaire et la partie propagative. La principale limitation de cette méthode réside dans son manque de localisation temporelle : en effet, l'analyse de Fourier permet de connaître les différentes fréquences excitées dans un signal, mais pas de savoir à quels instants ces fréquences ont été émises. Cette perte de localité n'est pas un inconvénient pour analyser des signaux dont la structure n'évolue pas ou peu (signaux dits « statistiquement stationnaires »), mais devient un problème pour l'étude de signaux non stationnaires, dont le contenu spectral varie au cours du temps (transitoires, ruptures, modulations,...), ce qui est le cas de nos ondes océaniques.

C'est pourquoi nous avons aussi utilisé une seconde méthode qui remplace l'analyse de Fourier temporelle des coefficients spectraux spatiaux (2^e étape, Figure 5.4) par une analyse temps-fréquence. L'idée de base est d'introduire dans l'analyse de Fourier usuelle une notion de localité temporelle en remplaçant la fonction analysée par le produit de celle-ci par une fenêtre convenablement choisie au préalable possédant de bonnes propriétés de localisation, puis en calculant la transformée de Fourier du produit ainsi formé. On renouvelle alors l'opération avec des copies translatées de la fenêtre, ce qui conduit à une analyse locale en tout point. Ce type d'analyse, appelée transformée de Gabor ou transformée de Fourier à fenêtre glissante, fournit ainsi la localisation des périodes dans le temps. La fenêtre analysante choisie est une gaussienne et les valeurs de la fenêtre correspondent aux poids alloués aux observations pour le calcul : par conséquent, chacun des spectres spatio-temporels locaux déterminés par analyse de Gabor est lié aux caractéristiques des

observations situées au centre de la fenêtre. La décomposition d'Hayashi permet ensuite d'acquérir les parties stationnaires et propagatives de tous les spectres de puissance spatio-temporels. A cette étape de l'analyse, nous sommes donc en présence des variations en temps et en période de la DSP pour le signal entier et les parties stationnaire et propagative : il s'agit de la modulation en amplitude de notre signal relatif à un couple nombre d'onde zonal / nombre d'onde méridien. Or, les signaux océaniques sont aussi modulés en fréquence. L'analyse de Céron intègre donc une quatrième étape supplémentaire, consistant en une extraction de l'arête de la transformée de Gabor, permettant justement d'obtenir les fréquences instantanées du signal. L'algorithme est basé sur les travaux de Torrèsani [1995] et repose sur les propriétés stationnaires des coefficients de phase de la transformée de Gabor.

Deux méthodes ont été utilisées. L'analyse d'Hayashi permet de visualiser les variations de la DSP en fonction des périodes pour le signal entier et les parties stationnaire et propagative. L'analyse de Céron détermine la modulation en amplitude et en fréquence du signal étudié, en visualisant respectivement les variations en temps et en période de la DSP pour le signal entier et les parties stationnaire et propagative, ainsi que les fréquences instantanées avec l'extraction de l'arête de la transformée de Gabor.

Ces deux méthodes sont complémentaires. En effet, l'analyse de Gabor présente l'inconvénient majeur de la taille de la fenêtre analysante, puisque le principe d'incertitude d'Heisenberg limite la taille de cette fenêtre si on ne veut pas perdre trop de l'information temporelle. Dans notre étude, nous avons estimé qu'une gaborette analysante maximum de taille $T/2$ (où T est la longueur temporelle du signal d'origine) était un bon compromis. Malheureusement, le résultat d'une analyse de Gabor utilisant une telle fenêtre offre une quasi-certitude sur l'intervalle des périodes comprises entre $-T/4$ et $+T/4$, mais est soumis à controverse sur les intervalles $[-T/2 ; -T/4]$ et $[+T/4 ; +T/2]$ (problème des extrémités de la fenêtre d'analyse), comme le montre la Figure 5.5. L'analyse d'Hayashi, utilisant la transformée de Fourier, permet de vérifier l'exactitude des résultats obtenus sur ces deux intervalles. En effet, la fenêtre analysante a alors la taille du signal d'origine (T) et le résultat obtenu est quasi-certain sur l'intervalle $[-T/2 ; +T/2]$ et soumis à controverse sur $[-T ; -T/2]$ et $[+T/2 ; +T]$. Bien sûr, une telle analyse ne signale que les périodes des pics de DSP, pas les instants auxquels ils apparaissent. Néanmoins, cela permet d'avoir une idée des périodes importantes du signal étudié.

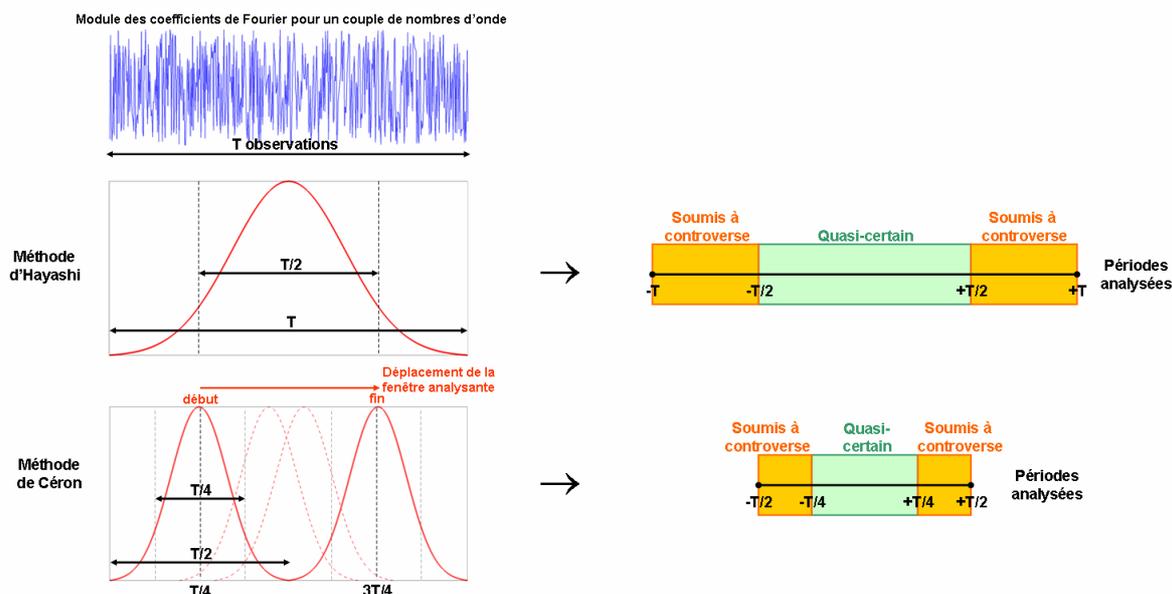


Figure 5.5 – Fenêtre analysante et méthodologie : schéma explicatif. Dans l'analyse d'Hayashi, on applique une transformée de Fourier au signal formé par la variation temporelle des coefficients spectraux spatiaux relatifs à un couple de nombres d'onde ; la fenêtre analysante est de la longueur du signal (T) et l'analyse est quasi-certaine sur l'intervalle des périodes comprises entre $-T/2$ et $T/2$. Quant à l'analyse de Céron, c'est une transformée de Gabor qu'on applique ; la taille de la fenêtre analysante est de $T/2$ au maximum (principe d'incertitude d'Heisenberg) et l'analyse est quasi-certaine sur l'intervalle des périodes comprises entre $-T/4$ et $T/4$.

Les deux parties suivantes montrent les résultats de l'application de ces deux méthodes aux données altimétriques, aux données mensuelles de chlorophylle de SeaWiFS et aux données hebdomadaires de chlorophylle issues des produits combinés SeaWiFS + MODIS/Aqua.

3 – Application aux données altimétriques

3.1 - Introduction

L'objectif de cette partie est d'analyser les signaux dominants dans l'Océan Austral en appliquant les méthodes présentées précédemment aux données physiques d'anomalies de hauteur de mer. Cette étude est un complément à celle menée dans l'article Pottier *et al.* [2004], fourni en Annexe B-1, où il était question d'une étude similaire mais sur un jeu de données plus court (période 1992 – 2003) et une zone s'étalant sur 10° de plus au sud.

Nous présentons, dans un premier temps, les données utilisées. Puis nous étudions les résultats obtenus que nous discutons dans une dernière partie.

3.2 - Les données utilisées

Les données analysées sont des cartes d'anomalies de hauteur de mer (SLA : Sea Level Anomalies ; unité : cm) issues de la combinaison de TOPEX-Poséidon, ERS-1/2, Jason-1 et Envisat [Ducet *et al.*, 2000]. Pour plus de détails sur la mesure ou les missions satellitaires altimétriques considérées, le lecteur pourra se référer à l'Annexe A-3.

Ces données ont une résolution spatiale de $1/4^\circ$ et une résolution temporelle de **7 jours**. La durée temporelle de la série à analyser est d'un peu moins de 14 ans, plus exactement du 14 octobre 1992 au 30 mai 2006, soit **711 observations**. Une observation sur l'ensemble de la zone d'étude, le continent Antarctique mis à part, c'est-à-dire de 35°S à 50°S , amène une étude en latitude sur **61 points** avec une résolution de **27.8 km**. Une observation sur 360° en longitude amène une étude sur **1440 points** avec une résolution variant entre **22.8 km** (pour 35°S) et **17.9 km** (pour 50°S).

Cette série présente des valeurs manquantes localisées sur les continents et les glaces. La méthode utilisée pour les estimer sans altérer le spectre du signal est basée sur l'algorithme LMS 1D (Least Mean Square).

3.3 - Résultats

Le Tableau 5.1 montre les maxima des DSP du signal entier et des parties stationnaire et propagative pour le signal analysé sur les bandes 35°S - 50°S , 35°S - 45°S et 45°S - 50°S . Seuls les 10 couples de nombres d'onde les plus significatifs sur la bande totale sont indiqués. L'importance de chaque couple sur le domaine d'étude est donnée par la valeur du maximum de DSP du signal entier. On constate que le couple 3/0 émerge nettement, quelle que soit la bande d'intérêt ; il est quand même nettement plus important sur la bande 45°S - 50°S . Quant au couple 2/0, nous avons vu dans Pottier *et al.* [2004] qu'il était lui aussi émergent sur la bande 35°S - 60°S , mais qu'il n'était significatif qu'à partir de 45°S : nos résultats le confirment (Tableau 5.1) et, puisque notre bande d'étude totale est moins étendue de 10° vers le sud, ce couple ressort beaucoup moins quand on analyse le domaine global. Comme l'indique leur nombre d'onde méridional égal à 0, les ondes correspondant à ces deux couples n'ont pas de propagation méridionale. Ces deux modes (dénommés ZWN2 et ZWN3 par la suite) sont aussi caractérisés par un maximum de DSP de la partie propagative supérieur à celui de la partie stationnaire : ces modes sont donc deux ondes propagatives qui se déplacent zonalement.

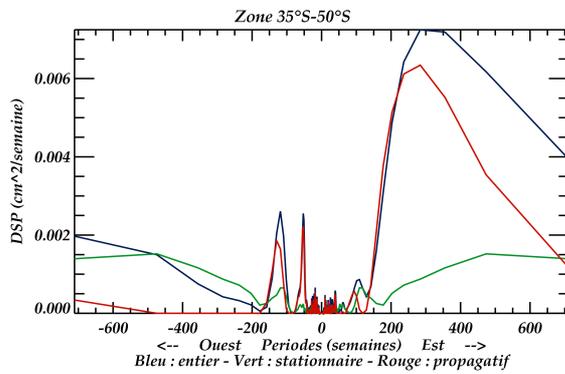
Tableau 5.1 – Maxima des DSP (cm²/semaine) du signal entier et des parties stationnaire et propagative pour les dix paires les plus significatives obtenues par l'analyse d'Hayashi sur toute la zone 35°S-50°S. Les résultats obtenus sur les bandes 35°S-45°S et 45°S-50°S sont aussi indiqués.

Couples de nombres d'onde (zonal / méridional)	Bande 35°S-50°S			Bande 35°S-45°S			Bande 45°S-50°S		
	Total	Sta.	Prop.	Total	Sta.	Prop.	Total	Sta.	Prop.
2/0	0.0072	0.0015	0.0063	0.0079	0.0020	0.0060	0.0273	0.0063	0.0224
2/1	0.0060	0.0023	0.0038	x	x	x	x	x	x
2/-1	0.0045	0.0011	0.0034	x	x	x	x	x	x
3/0	0.0156	0.0049	0.0146	0.0147	0.0054	0.0096	0.0309	0.0053	0.0278
4/0	0.0067	0.0014	0.0051	0.0092	0.0030	0.0060	0.0063	0.0020	0.0052
5/0	0.0087	0.0032	0.0051	0.0098	0.0044	0.0051	0.0082	0.0027	0.0052
6/0	0.0037	0.0016	0.0021	0.0038	0.0017	0.0021	0.0079	0.0029	0.0048
6/1	0.0023	0.0003	0.0019	x	x	x	x	x	x
8/0	0.0021	0.0005	0.0015	x	x	x	0.0065	0.0018	0.0046
10/0	0.0019	0.0004	0.0016	0.0039	0.0007	0.0032	x	x	x

La Figure 5.6 montre les résultats obtenus par l'analyse d'Hayashi, i.e. les variations de la DSP en fonction des périodes, tandis que la Figure 5.7 expose ceux obtenus par l'analyse de Céron, i.e. les variations en temps et en période de la DSP. Notons que cette dernière figure indique aussi les périodes instantanées (cf. lignes noires en gras) fournies par l'extraction de l'arête de la transformée de Gabor. Les périodes positives (resp. négatives) signifient une propagation vers l'est (resp. vers l'ouest).

Les résultats pour chaque couple de nombres d'onde s'analysent comme suit : tout d'abord, le signal entier est étudié, dans le but d'établir toutes ses caractéristiques ; puis, les composantes stationnaire et propagative sont comparées, dans l'objectif de confirmer les premières conclusions. Puisque l'analyse d'Hayashi est appliquée sur toute la série temporelle, les figures qui lui sont relatives montrent plus d'informations pour les périodes supérieures à 6.8 ans (711/2 semaines \approx 355 semaines \approx 6.8 ans) que celles supérieures à 3.4 ans (711/4 \approx 177 semaines \approx 3.4 ans) pouvant être détectées par l'analyse de Céron.

Couple 2/0



Couple 3/0

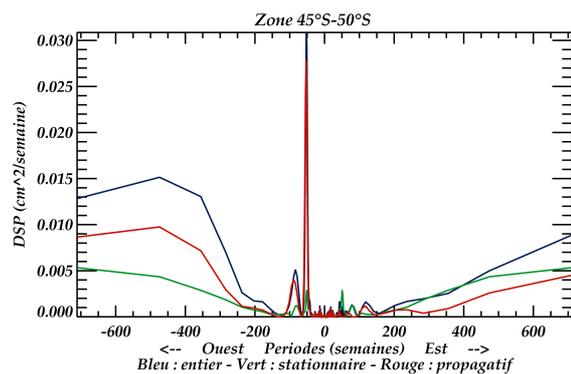
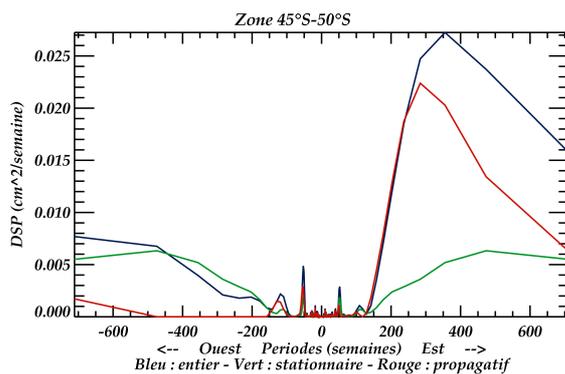
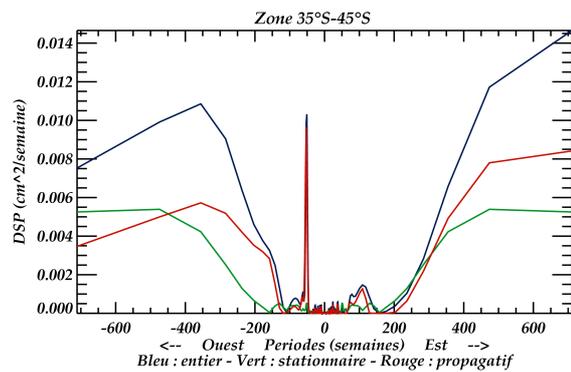
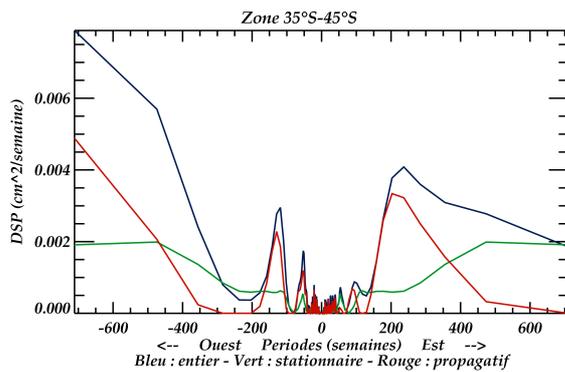
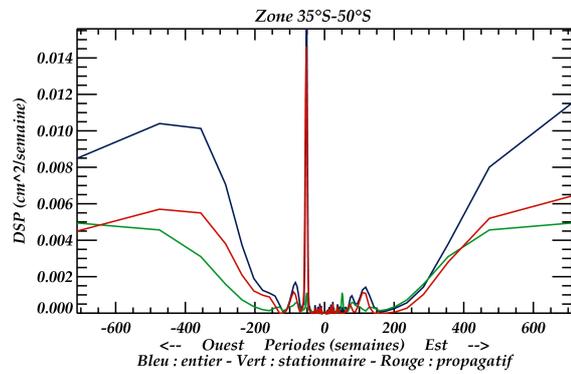


Figure 5.6 – Variations de la DSP selon les périodes fournies par l'analyse d'Hayashi sur les domaines 35°S-50°S (en haut), 35°S-45°S (au milieu) et 45°S-50°S (en bas) pour le signal entier (bleu), la partie stationnaire (vert) et la partie propagative (rouge) associée aux couples 2/0 (à gauche) et 3/0 (à droite).

Commençons par analyser le mode ZWN2. Les résultats de l'analyse d'Hayashi (Figure 5.6, à gauche) montrent que la majeure partie de l'énergie du signal entier se situe au niveau des périodes positives, ce qui est confirmé par les parties stationnaire et propagative. En effet, la partie propagative dénote clairement une propagation vers l'est avec une période d'environ 284 semaines (~5.4 ans), alors que la partie stationnaire est relativement faible. Cette propagation vers l'est est cohérente avec les résultats obtenus suivant l'analyse de Céron (Figure 5.7, à gauche). Les valeurs maximales de DSP du signal entier correspondant à un mode propagatif vers l'est sont observées entre janvier 1998 et juillet 1999, avec un maximum en septembre 1998 (DSP du signal entier : Figure 5.7, en haut à gauche). Les périodes instantanées associées (177 semaines \approx 3.4 ans et 237 semaines \approx 4.5 ans) sont plus faibles que celles fournies par l'analyse d'Hayashi. Dans tous les cas, cela doit être considéré avec précaution, puisque la période étudiée est supérieure à la demi-largeur de la fenêtre analysante (177 semaines). Ces résultats confirment ceux obtenus dans Pottier *et al.* [2004].

Intéressons-nous maintenant au mode ZWN3. Les résultats de l'analyse d'Hayashi (sur toutes les bandes de latitude : Figure 5.6, à droite) pointent une très nette propagation vers l'ouest, avec une période d'environ 53 semaines (~1 an). Cette propagation vers l'ouest est confirmée par les variations temporelles de la DSP du signal entier sur toutes les bandes de latitudes (Figure 5.8, en haut), qui montre les résultats obtenus avec l'analyse de Céron en utilisant une fenêtre analysante moins large (78 semaines au lieu de 177), afin d'augmenter la longueur temporelle de l'échantillon observé tout en gardant une résolution fréquentielle raisonnable. Le mode ZWN3 est partiellement filtré quand le cycle annuel est retiré du jeu de données SLA (non montré ici). Néanmoins, ce signal peut être interprété avec une grande confiance comme un mode propagatif vers l'ouest réel, comme en témoigne la continuité de l'arête qui dénote une phase structurée autour de la période 53 semaines. Quelle que soit la bande de latitude, l'évolution temporelle de la DSP du signal entier relatif au mode ZWN3 est très intéressante (Figure 5.8, en haut) : elle comporte deux maxima évidents (jusqu'à et à partir de septembre 1998) avec respectivement deux (en mars 1995 et avril 1996) et trois (en mars 2000, mars 2002 et janvier 2004) noyaux chacun. Notons aussi que la DSP présente un minimum aux alentours de septembre 1998, correspondant au maximum de DSP du couple 2/0. Ces résultats confirment, là encore, ceux obtenus dans Pottier *et al.* [2004]. Deux noyaux stationnaires sont aussi à remarquer, le premier en mars 1996 et le second en février 2002. On peut constater que ces noyaux ont des intensités différentes en fonction de la bande de latitude d'intérêt.

Couple 2/0

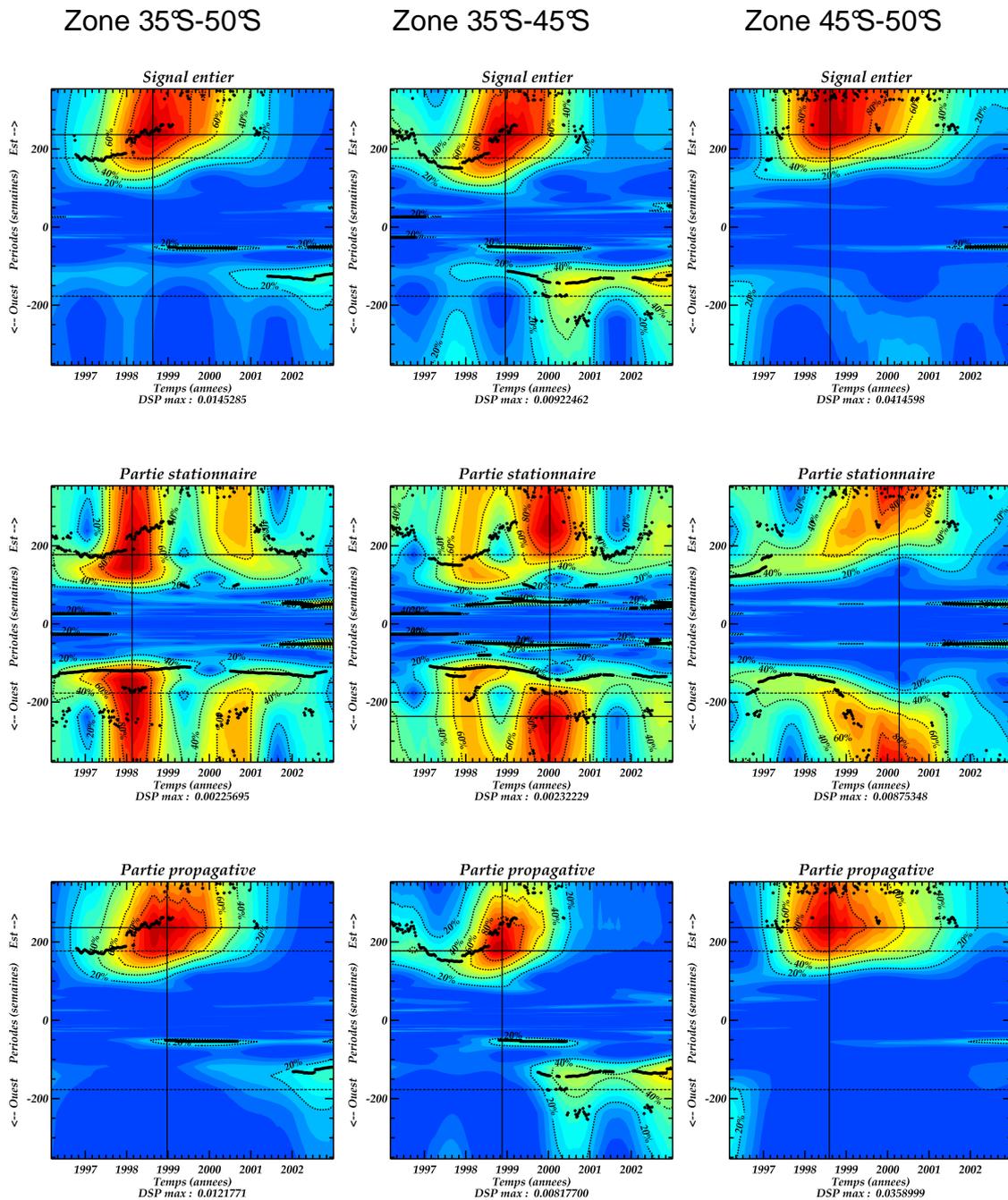


Figure 5.7 – Variations en temps et en période de la DSP du signal entier (en haut) et des parties stationnaire (au milieu) et propagative (en bas) associée au couple 2/0 sur la zone 35°S-50°S (à gauche), 35°S-45°S (au milieu) et 45°S-50°S (à droite). La DSP est représentée sous forme de pourcentage de la valeur maximale (bleu : 0% - rouge : 100%) ; notons que cette valeur maximale, indiquée sous chaque axe des abscisses (« PSD max ») est différente pour chaque graphe. L'arête (i.e. les périodes instantanées) est représentée par les lignes noires en gras. La demi-largeur de la fenêtre analysante, représentée par les pointillés, est de 177 semaines (3.4 ans). L'intersection des deux lignes fines continues indique le maximum.

Couple 3/0

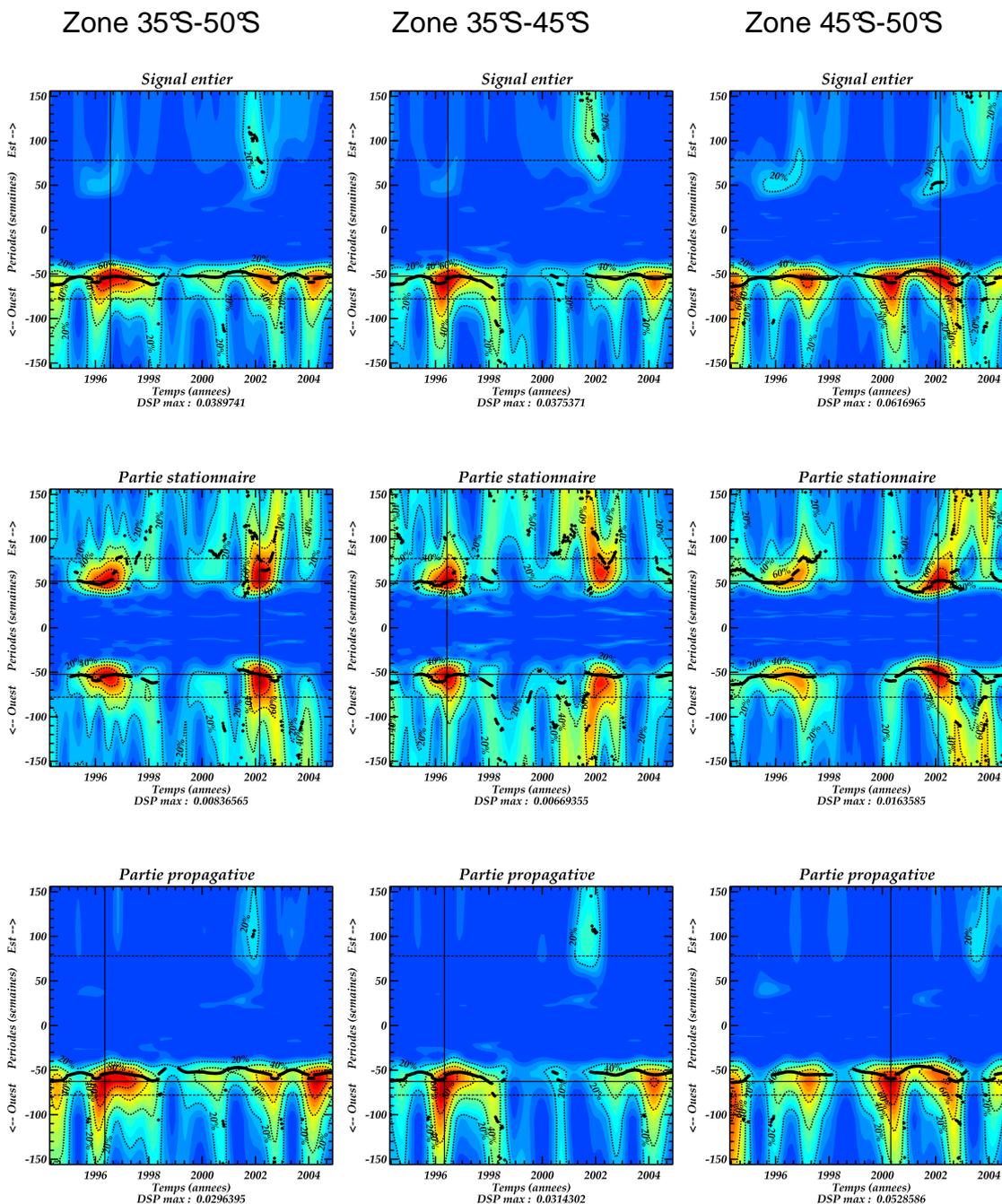


Figure 5.8 – Variations en temps et en période de la DSP du signal entier (en haut) et des parties stationnaire (au milieu) et propagative (en bas) associée au couple 3/0 sur la zone 35°S-50°S (à gauche), 35°S-45°S (au milieu) et 45°S-50°S (à droite). La DSP est représentée sous forme de pourcentage de la valeur maximale (bleu : 0% - rouge : 100%) ; notons que cette valeur maximale, indiquée sous chaque axe des abscisses (« PSD max ») est différente pour chaque graphe. L'arête (i.e. les périodes instantanées) est représentée par les lignes noires en gras. La demi-largeur de la fenêtre analysante, représentée par les pointillés, est de 78 semaines (1.5 an). L'intersection des deux lignes fines continues indique le maximum.

3.4 - Discussion et conclusion

Les analyses d'Hayashi et de Céron sont utilisées comme deux méthodes complémentaires pour détecter et caractériser des modes à la fois stationnaires et propagatifs dans un jeu de données de SLA sur la période 1992 – 2006 dans l'Océan Austral. Le principal avantage de l'analyse de Céron repose sur le fait qu'elle fournit la localisation des périodes dans le temps ; mais l'inconvénient est que seuls les modes détectés avec précision grâce à cette méthode seront ceux dont la période est inférieure à la moitié de la fenêtre la plus large pouvant être utilisée (177 semaines). Pour les périodes comprises entre 177 et 355 semaines, les résultats sont sujet à discussion et doivent être confirmés à l'aide de l'analyse d'Hayashi. Il n'y a aucune ambiguïté en ce qui concerne les résultats du mode ZWN3 propagatif vers l'ouest avec une période proche de 1 an (53 semaines). En revanche, l'analyse d'Hayashi est essentielle pour confirmer la propagation vers l'est associée au mode 2/0 pour lequel la période (supérieure à 5 ans) excède le domaine de Gabor (3.4 ans).

Confirmant Pottier *et al.* [2004], le mode ZWN2 montre une très nette propagation vers l'est. Ce mode a une période d'environ 5.5 ans (284 semaines, d'après l'analyse d'Hayashi) et arbore un maximum de DSP en septembre 1998. On peut ainsi associer avec confiance notre couple 2/0 à l'ACW, confirmant de ce fait des études précédentes sur des jeux de données océaniques et atmosphériques (White et Peterson, 1996 ; Jacobs et Mitchell, 1996 ; Carril et Navarra, 2001 ; Cai et Baines, 2001 ; Venegas, 2003). Dans les données de SLP, White *et al.* [2002] ont montré que le trajet de l'ACW suit le front subtropical près de 45°S dans les secteurs de l'Atlantique est, l'Indien, et le Pacifique ouest et central (aux latitudes de l'Océan Austral) avant de plonger vers le Sud dans le Pacifique est pour passer à travers le passage de Drake. Le maximum prononcé dans la région 45°S-50°S tend à confirmer cette particularité. De plus, nos résultats corroborent le scénario de Peterson et White [1998] selon lequel l'une des sources de l'ACW serait reliée au phénomène ENSO. L'étude conjointe de l'indice SOI (Southern Oscillation Index) fourni par le centre de prédiction du climat NOAA (National Oceanic and Atmospheric Administration ; <http://www.cpc.noaa.gov/data/indices/>) et de la DSP de la partie propagative du couple 2/0 extraite pour la période d'intérêt indique que leurs évolutions temporelles pourraient être liées (Figure 5.9). La DSP commence à augmenter quand l'indice SOI devient négatif, marquant l'évènement El Niño de 1997-1998. La valeur minimale du SOI en mars 1998

(-3.5) au moment du pic de l'ENSO précède ensuite le pic de DSP (septembre 1998) de quelques mois (5-6 mois). Enfin, la DSP diminue, durant l'évènement La Niña qui suit.

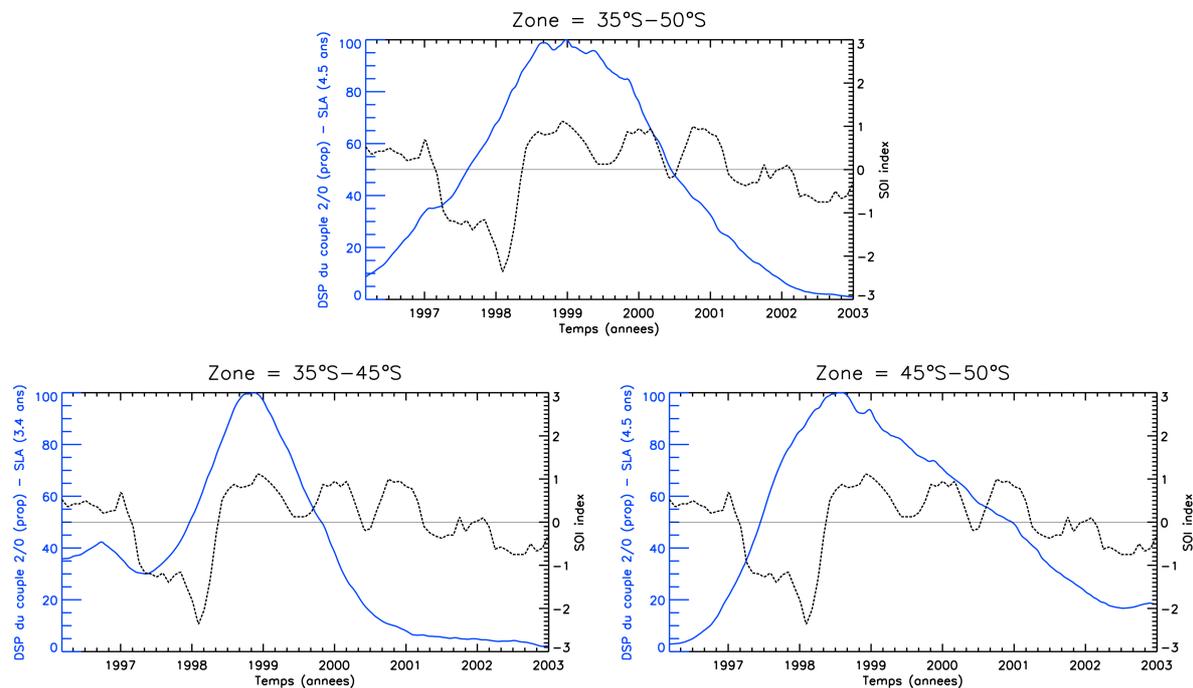


Figure 5.9 – En bleu : variations temporelles de la DSP de la partie propagative du mode ZWN2 sur les régions 35°S-50°S (en haut), 35°S-45°S (en bas, à gauche) et 45°S-50°S (en bas, à droite). Les périodes considérées correspondent à celles le long des arêtes (Figure 5.7). En pointillés noirs : variations temporelles de l'indice SOI du NOAA Climate Prediction Center [<http://www.cpc.noaa.gov/data/indices/>], lissé sur 3 mois.

Considérons maintenant le couplet 3/0. Les analyses d'Hayashi et de Céron révèlent une structure ZWN3 dans le signal SLA avec une propagation vers l'ouest et une période de 1 an. Ce résultat est cohérent avec ceux de Jacobs et Mitchell [1996], qui ont trouvé, dans leur décomposition de Fourier, un pic de période 1 an dans leur spectre de hauteur de mer, caractérisé par un nombre d'onde 3 et une propagation vers l'ouest. Une composante ACW ZWN3 avec une phase uniforme sur l'Antarctique a aussi été trouvée par Cai et Baines [2001], qui ont suggéré un lien avec le SAM atmosphérique. Plus récemment, White [2003], utilisant une décomposition en valeurs singulières sur les données de SLP, a identifié un signal de période proche de 1.1 an, se comportant comme le SAM mais mélangé avec une structure de variabilité de type ZWN3. Hughes *et al.* [2003] ont montré que la variabilité de la pression de subsurface mesurée par les marégraphes en Antarctique est liée à un forçage atmosphérique ressemblant au SAM. Nous avons tracé les variations de l'indice SAM de Marshall [2003] avec celles de la DSP de la partie propagative du mode ZWN3 sur la région 35°S-50°S relative à la période 53 semaines (Figure 5.10). Le mode propagatif vers l'ouest associé au couplet 3/0 est maximum quand les valeurs du SAM sont négatives et diminue

très fortement quand le SAM atteint de hautes valeurs en 1998. Suivant le raisonnement de Hall et Visbeck [2002], qui ont conclu qu'un ACC plus intense est étroitement associé à un SAM positif, on peut penser que notre mode ZWN3 propagatif vers l'ouest peut, en raison de l'effet Doppler, diminuer et même devenir mineur par rapport au mode stationnaire quand le transport de l'ACC augmente. La répartition entre les composantes stationnaires et propagatives de notre mode ZWN3 peut donc être liée à un couplage fort entre l'ACC et l'atmosphère.

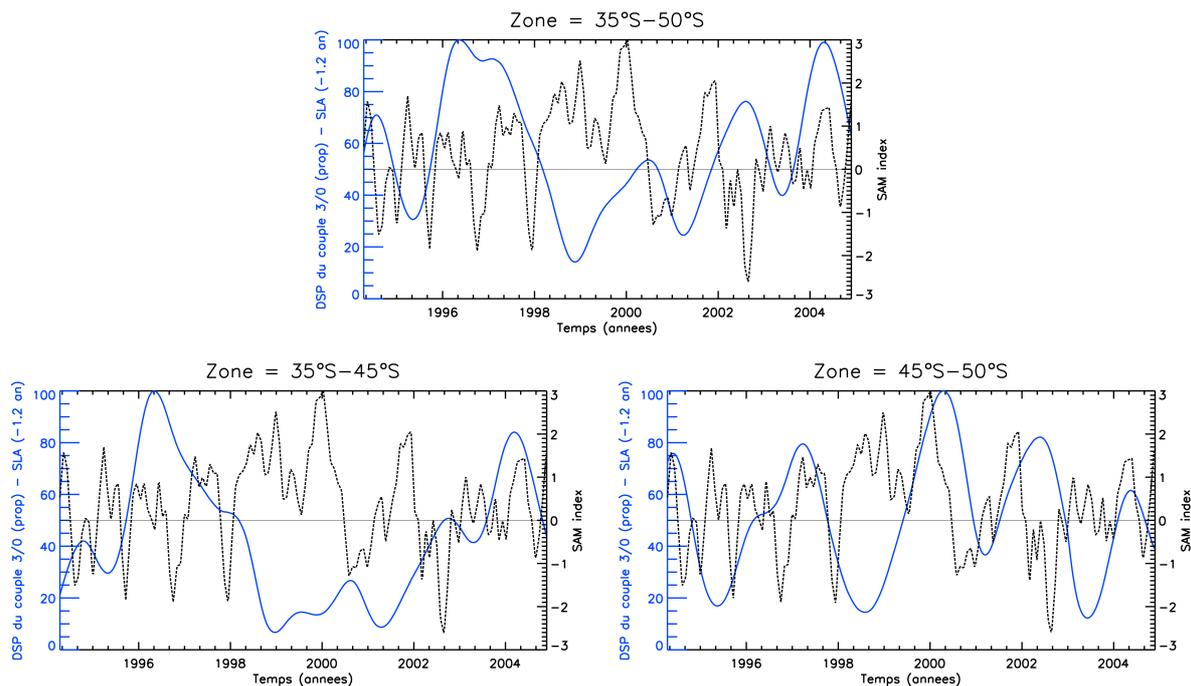


Figure 5.10 – En bleu : variations temporelles de la DSP de la partie propagative du mode ZWN3 sur les régions 35°S-50°S (en haut), 35°S-45°S (en bas, à gauche) et 45°S-50°S (en bas, à droite). Les périodes considérées correspondent à celles des arêtes de la Figure 5.8. En pointillés noirs : variations temporelles de l'indice SAM de Marshall [2003], lissé sur 3 mois.

En conclusion, nous avons confirmé les résultats obtenus dans une étude précédente, Pottier *et al.* [2004], et caractérisé 2 ondes dominantes dans les données SLA au niveau de l'Océan Austral : l'ACW, associée au couple 2/0, qui se propage vers l'est avec une période d'environ 4-5 ans et une onde associée au couple 3/0 qui se propage vers l'ouest avec une période proche de 1 an. Dans la partie suivante, nous allons appliquer les deux méthodes d'analyse aux données mensuelles de concentration en chlorophylle de SeaWiFS et comparer les résultats obtenus.

4 - Application aux données de chlorophylle

4.1 - Introduction

L'objectif de cette partie est d'analyser les signaux dominants dans l'Océan Austral en appliquant les méthodes présentées précédemment aux données biologiques de concentration en chlorophylle et de comparer les résultats obtenus à ceux des données altimétriques.

Nous présentons, dans un premier temps, les résultats de l'application aux données mensuelles issues du capteur SeaWiFS, sur la période du 1^{er} septembre 1997 au 31 mai 2006. Dans un second temps, nous commentons les résultats de l'application de la méthode aux données hebdomadaires (moyenne des données journalières sur une semaine) issues de la combinaison des données SeaWiFS et MODIS/Aqua pour la période du 1^{er} juillet 2002 au 31 mai 2006. Enfin, dans une dernière partie, nous comparons et discutons les résultats obtenus.

4.2 - Application aux données mensuelles issues de SeaWiFS

4.2.1 - Les données utilisées

Les données analysées sont des cartes de concentration en chlorophylle-a issues du capteur SeaWiFS (cf. Chapitre 1-3.2.3). Nous avons utilisé des cartes mensuelles pour que les cartes soient les plus remplies possibles. En effet, les cartes aux résolutions temporelles inférieures présentent trop de données manquantes (dues essentiellement aux nuages).

Ces données ont une résolution spatiale de **1/12°** et une résolution temporelle de **1 mois**. La durée temporelle de la série à analyser est d'un peu moins de 9 ans, plus exactement du 1^{er} septembre 1997 au 31 mai 2006, soit **105 observations**. Une observation sur l'ensemble de la zone d'étude, le continent Antarctique mis à part, c'est-à-dire de 35°S à 50°S, amène une étude en latitude sur **180 points** avec une résolution de **9.2 km**. Une observation sur 360° en longitude amène une étude sur **4320 points** avec une résolution variant entre **7.6 km** (pour 35°S) et **6 km** (pour 50°S).

Cette série présente non seulement des valeurs manquantes localisées sur les continents et les glaces (comme les données altimétriques), mais aussi quelques données manquantes dues aux nuages et autres phénomènes empêchant la mesure. Là encore, la méthode utilisée pour remplir ces trous sans altérer le spectre du signal est basée sur l'algorithme prédictif LMS 1D (Least Mean Square). Les données trop côtières, présentant

une concentration en chlorophylle très élevée, ont préalablement été supprimées pour ne pas altérer l'algorithme.

4.2.2 - Résultats

Le Tableau 5.2 montre les maxima des DSP du signal entier et des parties stationnaire et propagative pour les données mensuelles de chlorophylle-a SeaWiFS sur les bandes 35°S-50°S, 35°S-45°S et 45°S-50°S. On constate que le couple 3/0, déjà noté en altimétrie dans la partie précédente, émerge nettement sur le domaine entier, particulièrement sur la bande 35°S-45°S. Toujours comme pour l'altimétrie, l'onde correspondant à ce couple est une onde propagative qui se déplace zonalement.

Tableau 5.2 – Maxima des DSP (cm²/semaine) du signal entier et des parties stationnaire et propagative pour les dix paires les plus significatives obtenues par l'analyse d'Hayashi sur toute la zone 35°S-50°S. Les résultats obtenus sur les bandes 35°S-45°S et 45°S-50°S sont aussi indiqués.

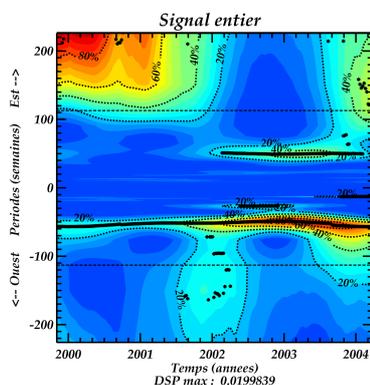
Couples de nombres d'onde (zonal / méridional)	Bande 35°S-50°S			Bande 35°S-45°S			Bande 45°S-50°S		
	Total	Sta.	Prop.	Total	Sta.	Prop.	Total	Sta.	Prop.
2/0	2.9e-6	1.1e-6	1.8e-6	2.8e-6	0.9e-6	2.0e-6	11.5e-6	3.8e-6	7.7e-6
2/1	2.4e-6	1.1e-6	1.4e-6	1.8e-6	0.8e-6	1.0e-6	0.4e-6	0.1e-6	2.6e-6
2/-1	3.4e-6	0.7e-6	2.8e-6	1.9e-6	0.7e-6	1.3e-6	0.5e-6	0.2e-6	0.3e-6
3/0	6.1e-6	0.6e-6	5.5e-6	9.0e-6	2.3e-6	6.8e-6	4.9e-6	1.7e-6	3.3e-6
3/1	3.8e-6	0.7e-6	2.9e-6	1.6e-6	0.4e-6	1.1e-6	0.6e-6	0.2e-6	0.4e-6
3/-1	1.7e-6	0.5e-6	1.1e-6	1.8e-6	0.2e-6	1.4e-6	0.2e-6	0.1e-6	0.2e-6
4/0	0.9e-6	0.2e-6	0.4e-6	1.3e-6	0.3e-6	0.9e-6	0.9e-6	0.3e-6	0.6e-6
5/0	0.7e-6	0.1e-6	0.5e-6	1.0e-6	0.3e-6	0.6e-6	3.6e-6	0.7e-6	2.8e-6
5/-1	2.6e-6	0.1e-6	2.4e-6	2.2e-6	0.3e-6	1.8e-6	0.4e-6	0.1e-6	0.3e-6
6/0	0.8e-6	0.2e-6	0.6e-6	0.3e-6	0.1e-6	0.3e-6	2.5e-6	0.6e-6	1.9e-6

La Figure 5.12 montre les résultats relatifs au couple 3/0 obtenus par l'analyse d'Hayashi sur les données hebdomadaires de hauteur de mer (à gauche) et les données mensuelles de chlorophylle-a SeaWiFS (à droite), i.e. les variations de la DSP en fonction des périodes, tandis que les Figures 5.13 et 5.14 exposent ceux obtenus par l'analyse de Céron sur les données hebdomadaires de hauteur de mer et les données mensuelles de chlorophylle-a SeaWiFS respectivement, i.e. les variations en temps et en période de la DSP. L'analyse d'Hayashi permet l'étude de périodes allant jusqu'à 105/2 mois \approx 53 mois \approx 4.4 ans, tandis que l'analyse de Céron analyse des périodes jusqu'à 105/4 \approx 26 mois \approx 2.2 ans.

Pour le mode ZWN2, les résultats des analyses d'Hayashi et de Céron appliquées aux anomalies de hauteur de mer et aux données de chlorophylle sur la période comprise entre septembre 1997 et mai 2006 (i.e. la même période que notre archive SeaWiFS) (cf. Annexe

A-5.1) montrent que la période à 4-5 ans n'est pas du tout visible sur les données de chlorophylle, alors que l'on commence à la deviner sur les données altimétriques, particulièrement sur la bande 45°S-50°S, comme on peut le voir sur la Figure 5.11.

Anomalies de hauteur de mer



Chlorophylle

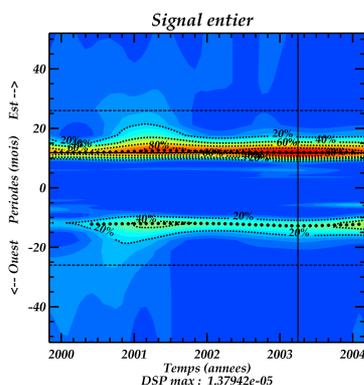


Figure 5.11 - Variations en temps et en période de la DSP du signal entier associée au couple 2/0 sur la zone 45°S-50°S pour les anomalies de hauteur de mer (à gauche) et la chlorophylle (à droite) sur la période comprise entre septembre 1997 et mai 2006. La DSP est représentée sous forme de pourcentage de la valeur maximale (bleu : 0% - rouge : 100%). L'arête (i.e. les périodes instantanées) est représentée par les lignes noires en gras. La demi-largeur de la fenêtre analysante, représentée par les pointillés, est de 78 semaines (26 mois, 1.5 an). L'intersection des deux lignes fines continues indique le maximum.

En ce qui concerne le mode ZWN3, les résultats des analyses d'Hayashi (Figure 5.12, à gauche) et de Céron (Figure 5.13) appliquées aux données altimétriques sur la période entre septembre 1997 et mai 2006 vont dans le même sens que ceux sur la période d'octobre 1992 à mai 2006 (partie précédente), à savoir que le mode ZWN3 a une propagation vers l'ouest très marquée, avec une période ~ 1 an. L'évolution temporelle de la DSP du signal entier montre la présence de trois noyaux (juin 2000, juin 2002 et avril 2004) dont l'intensité varie avec la bande de latitude : l'onde a été plus importante en juin 2002 et avril 2004 sur la bande 35°S-45°S, mais plus importante en juin 2000 et juin 2002 sur la bande 45°S-50°S. Une partie stationnaire est notable en février 2002.

Contrairement aux données altimétriques, si on retrouve cette période de 1 an environ dans les données de concentration en chlorophylle, on peut constater, dans un premier temps sur les résultats de l'analyse d'Hayashi (Figure 5.12, à droite) que le sens de propagation n'est pas toujours le même. En effet, la propagation de cette onde se fait vers l'est sur la bande 35°S-45°S et vers l'ouest, comme pour l'altimétrie, sur la bande 45°S-50°S. On peut aussi constater la présence d'une période à 6 mois associée à une propagation vers l'est, non visible pour l'altimétrie et dont la présence est plus importante sur la bande 45°S-50°S. Les résultats de l'analyse de Céron (Figure 5.14) confirment cette tendance.

L'évolution temporelle de la DSP du signal entier sur le domaine 35°S-50°S (Figure 5.14, en haut à gauche) est très intéressante : celle-ci comporte 4 maxima (jusqu'à avril 2000, entre avril 2000 et novembre 2003, de novembre 2003 à mars 2004 et à partir de mars 2004) avec un noyau chacun (resp. mai 1999, février 2001, septembre 2003 et novembre 2004). Ces maxima sont présents quelle que soit la bande de latitude d'intérêt, mais avec des intensités différentes en fonction de la zone. Notons aussi la présence de minima en avril 2000, novembre 2003 et mars 2004. Deux noyaux stationnaires sont aussi à remarquer, le premier en janvier 2001 et le second en juin 2003.

Anomalies de hauteur de mer

Chlorophylle

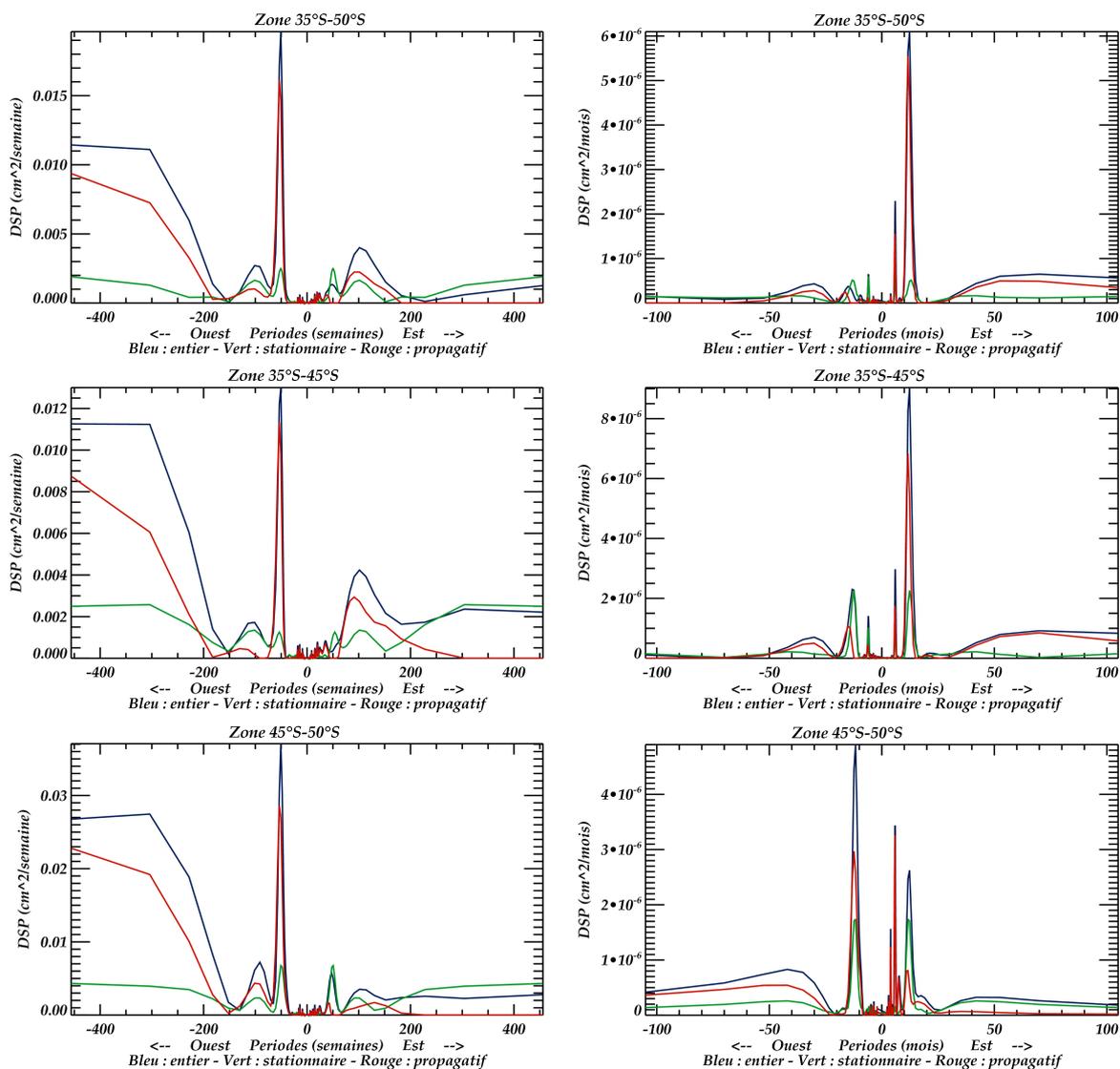


Figure 5.12 – Variations de la DSP associée au couple 3/0 selon les périodes fournies par l'analyse d'Hayashi sur les domaines 35°S-50°S (en haut), 35°S-45°S (au milieu) et 45°S-50°S (en bas) pour le signal entier (bleu), la partie stationnaire (vert) et la partie propagative (rouge) et pour les données altimétriques (à gauche) et les données de chlorophylle mensuelle SeaWiFS (à droite) pour la période comprise entre septembre 1997 et mai 2006.

Anomalies de hauteur de mer

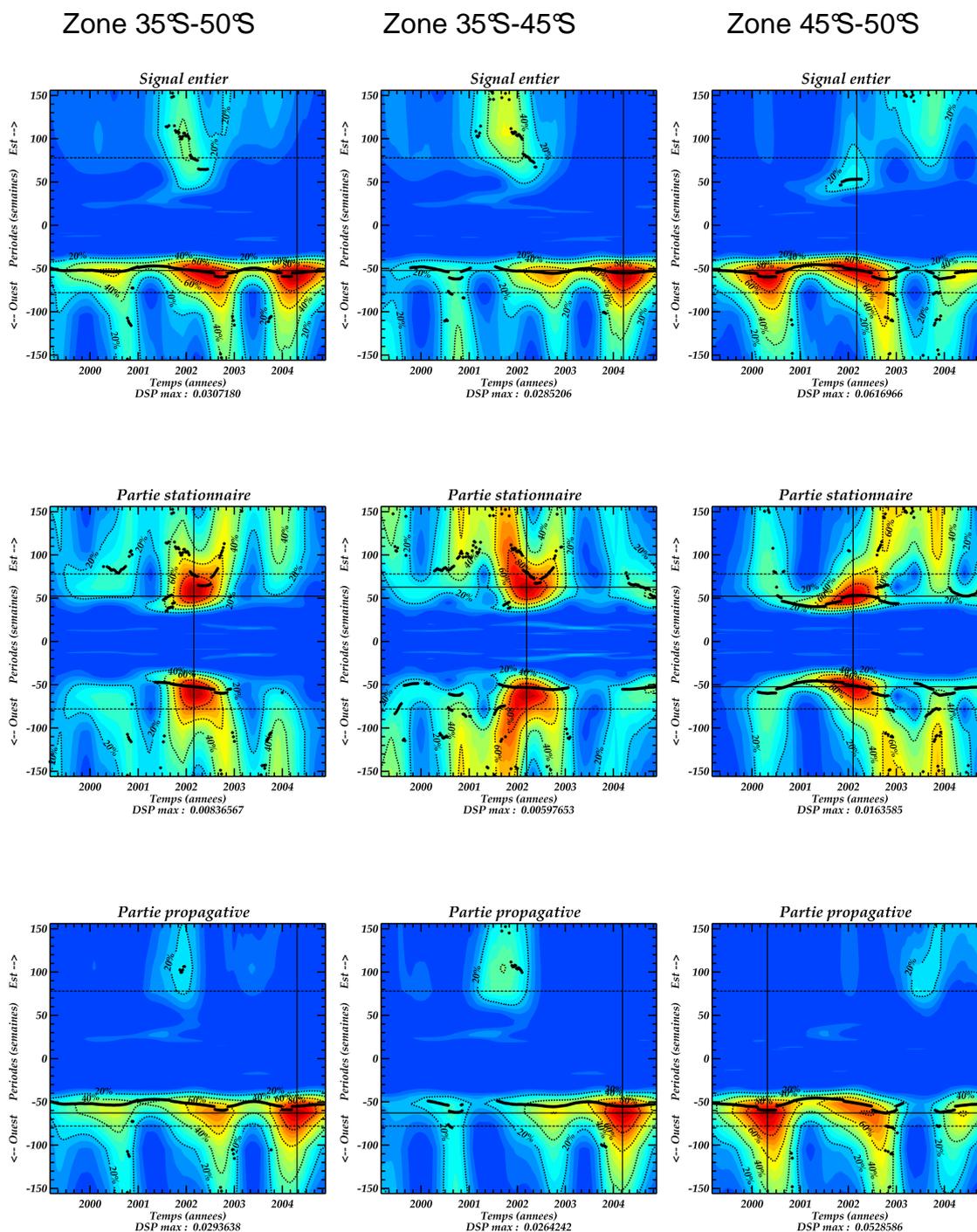


Figure 5.13 – Variations en temps et en période de la DSP associée à la paire 3/0 du signal entier (en haut) et des parties stationnaire (au milieu) et propagative (en bas) pour les données altimétriques, sur les domaines 35°S-50°S (à gauche), 35°S-45°S (au milieu) et 45°S-50°S (à droite) et pour la période comprise entre septembre 1997 et mai 2006. La demi-largeur de la fenêtre analysante est de 78 semaines = 1.5 ans.

Chlorophylle

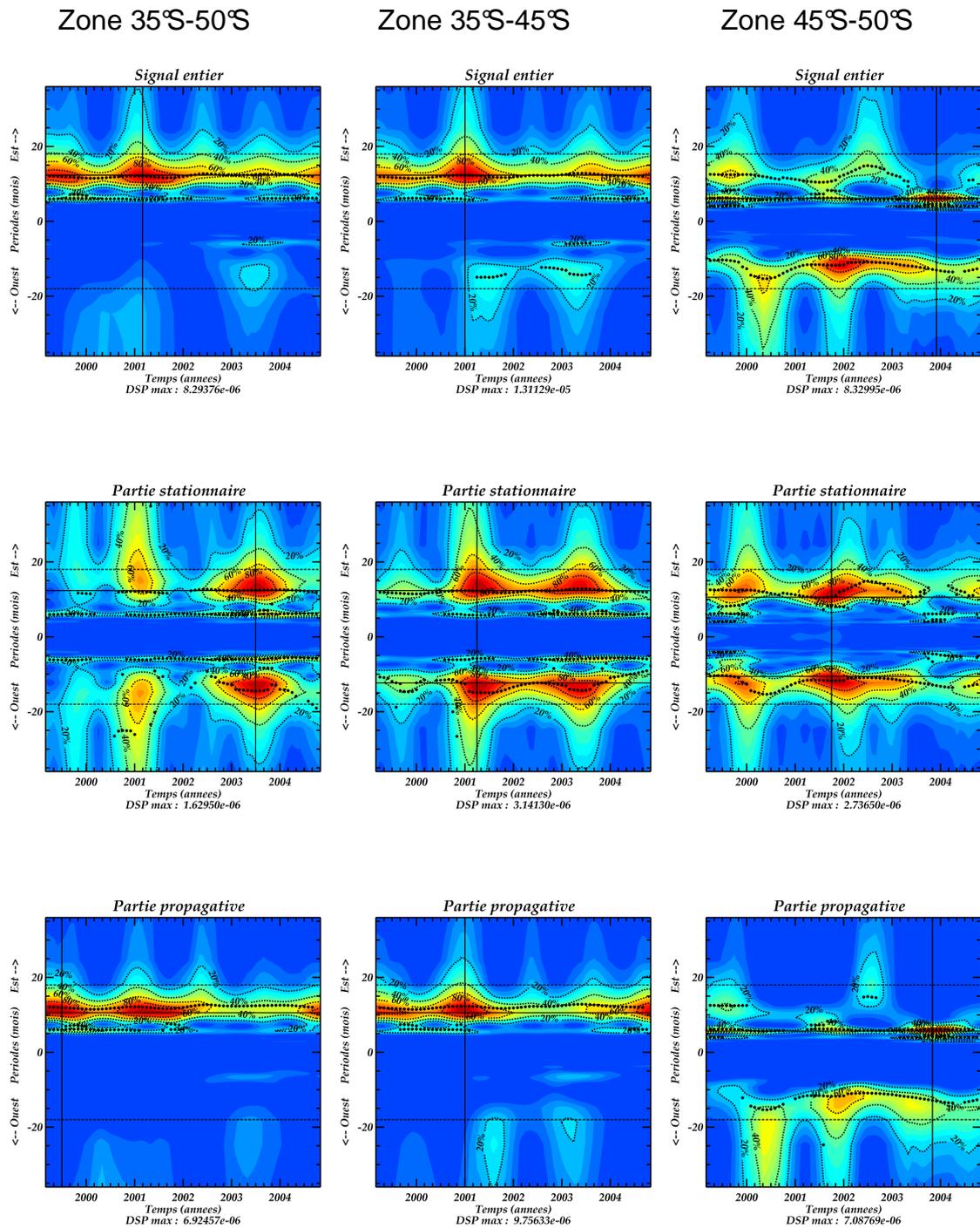


Figure 5.14 – Variations en temps et en période de la DSP associée à la paire 3/0 du signal entier (en haut) et des parties stationnaire (au milieu) et propagative (en bas) pour les données de chlorophylle, sur les domaines 35°S-50°S (à gauche), 35°S-45°S (au milieu) et 45°S-50°S (à droite) et pour la période comprise entre septembre 1997 et mai 2006. La demi-largeur de la fenêtre analysante est de 18 mois = 1.5 ans.

4.3 - Application aux données hebdomadaires issues de la combinaison de SeaWiFS et MODIS/Aqua

4.3.1 - Les données utilisées

Les données analysées sont des cartes de concentration en chlorophylle-a issues de la combinaison des données SeaWiFS et MODIS/Aqua par analyse objective (cf. Chapitre 3-3). Notons que l'on a considéré que les statistiques, alors calculées sur l'année 2003, étaient semblables pour les autres années, ce qui peut constituer une hypothèse forte. La sortie de l'analyse objective fournissant des données journalières, celles-ci ont été moyennées sur 7 jours afin d'obtenir des cartes hebdomadaires, c'est-à-dire la même résolution temporelle que les données altimétriques (cf. partie 3.2). Nous avons commencé l'analyse au 4 juillet 2002, début de la période commune SeaWiFS – MODIS/Aqua (cf. Tableau 1.2 au Chapitre 1).

Ces données ont une résolution spatiale de **1/12°** et une résolution temporelle de **1 semaine**. La durée temporelle de la série à analyser est d'à peine 4 ans, plus exactement du 4 juillet 2002 au 30 mai 2006, soit **204 observations**. Une observation sur l'ensemble de la zone d'étude, le continent Antarctique mis à part, c'est-à-dire de 35°S à 50°S, amène une étude en latitude sur **180 points** avec une résolution de **9.2 km**. Une observation sur 360° en longitude amène une étude sur **4320 points** avec une résolution variant entre **7.6 km** (pour 35°S) et **6 km** (pour 50°S).

Là encore, cette série présente des valeurs manquantes localisées sur les continents et les glaces (comme les données altimétriques). La méthode utilisée pour remplir ces trous sans altérer le spectre du signal est basée sur l'algorithme LMS 1D (Least Mean Square). Là encore, les données trop côtières, présentant une concentration en chlorophylle élevée, ont préalablement été supprimées pour ne pas altérer l'algorithme.

4.3.2 - Résultats

Le Tableau 5.3 montre les maxima des DSP du signal entier et des parties stationnaire et propagative pour le signal analysé sur les bandes 35°S-50°S, 35°S-45°S et 45°S-50°S. On constate que le couple 3/0 émerge toujours nettement sur le domaine entier, particulièrement sur la bande 35°S-45°S. Ce mode ZW N3 est aussi caractérisé par une propagation zonale.

Tableau 5.3 - Maxima des DSP (cm²/semaine) du signal entier et des parties stationnaire et propagative pour les dix paires les plus significatives obtenues par l'analyse d'Hayashi sur toute la zone 35°S-50°S. Les résultats obtenus sur les bandes 35°S-45°S et 45°S-50°S sont aussi indiqués.

Couples de nombres d'onde (zonal / méridional)	Bande 35°S-50°S			Bande 35°S-45°S			Bande 45°S-50°S		
	Total	Sta.	Prop.	Total	Sta.	Prop.	Total	Sta.	Prop.
2/0	7.4e-6	1.8e-6	5.5e-6	10.7e-6	1.4e-6	9.6e-6	17.0e-6	6.5e-6	9.8e-6
2/1	6.0e-6	2.7e-6	3.2e-6	4.5e-6	1.9e-6	2.4e-6	0.8e-6	0.2e-6	0.5e-6
2/-1	7.7e-6	1.3e-6	5.9e-6	5.9e-6	1.7e-6	4.0e-6	1.5e-6	0.5e-6	1.0e-6
3/0	14.1e-6	1.4e-6	12.1e-6	23.7e-6	1.9e-6	18.5e-6	8.9e-6	2.9e-6	5.7e-6
3/1	6.7e-6	1.4e-6	5.2e-6	4.0e-6	1.2e-6	2.8e-6	0.9e-6	0.4e-6	0.5e-6
3/-1	7.9e-6	0.9e-6	6.6e-6	6.6e-6	0.8e-6	5.7e-6	0.5e-6	0.2e-6	0.3e-6
4/0	2.0e-6	0.4e-6	1.6e-6	4.1e-6	0.9e-6	2.9e-6	1.2e-6	0.3e-6	1.0e-6
4/-1	4.2e-6	0.3e-6	3.7e-6	5.2e-6	0.4e-6	4.6e-6	0.1e-6	0.03e-6	0.08e-6
5/0	2.0e-6	0.3e-6	1.7e-6	2.5e-6	1.0e-6	1.5e-6	9.5e-6	2.4e-6	6.8e-6
5/-1	6.5e-6	0.4e-6	5.9e-6	5.6e-6	0.6e-6	4.8e-6	0.2e-6	0.08e-6	0.09e-6

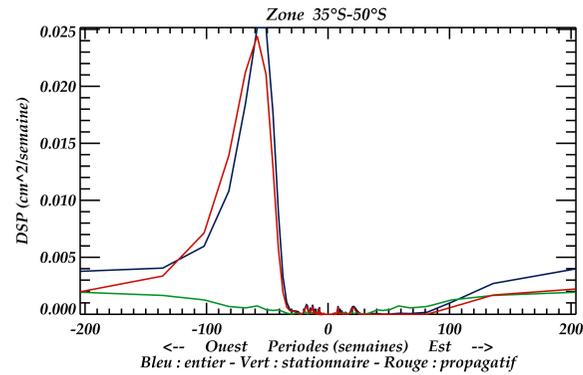
La Figure 5.15 montre les résultats obtenus par l'analyse d'Hayashi, i.e. les variations de la DSP en fonction des périodes, tandis que les Figures 5.16, 5.17 et 5.18 exposent ceux obtenus par l'analyse de Céron, i.e. les variations en temps et en période de la DSP. L'analyse d'Hayashi permet l'étude de périodes allant jusqu'à 204/2 semaines = 102 semaines ≈ 2 ans, tandis que l'analyse de Céron analyse des périodes jusqu'à 204/4 = 52 semaines ≈ 1 an.

L'analyse des données se faisant sur une période vraiment trop courte, le mode ZWN2 ne peut absolument pas être étudié.

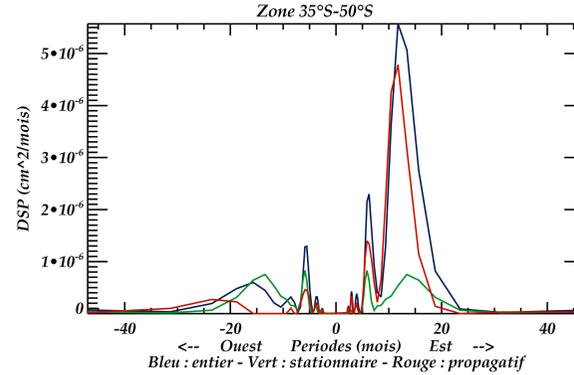
Pour le mode ZWN3, les résultats des analyses d'Hayashi (Figure 5.15, à gauche) et de Céron (Figures 5.16, 5.17 et 5.18, à gauche) appliquées aux données altimétriques sur la période entre juillet 2002 et mai 2006 vont dans le même sens que ceux des parties précédentes, à savoir que le mode ZWN3 a une propagation vers l'ouest très marquée, avec une période ~1 an. Sur la longueur temporelle analysée, un noyau maximum est visible sur la DSP du signal entier de la bande 35°S-50°S (Figure 5.16, en haut à gauche), allant d'octobre 2003 à novembre 2004, avec un maximum en avril 2004.

Pour les données mensuelles de concentration en chlorophylle SeaWiFS, on retrouve les résultats de la partie précédente, à savoir une période de 1 an environ, propagative vers l'est sur le domaine entier 35°S-50°S, propagative vers l'est sur la bande 35°S-45°S mais vers l'ouest, comme pour l'altimétrie, sur la bande 45°S-50°S. On retrouve aussi la période à 6 mois propagative vers l'est, particulièrement sur la bande 45°S-50°S.

Anomalies de hauteur de mer



Chlorophylle SeaWiFS



Chlorophylle combinée

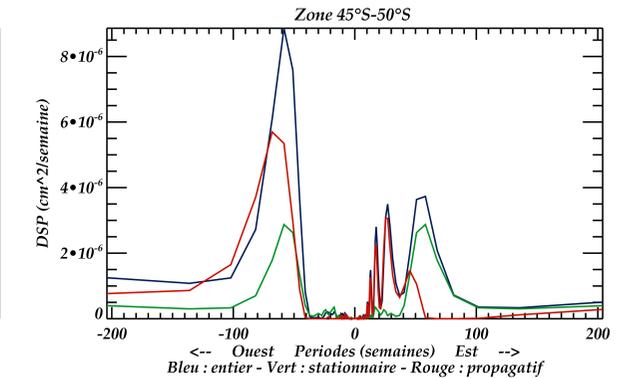
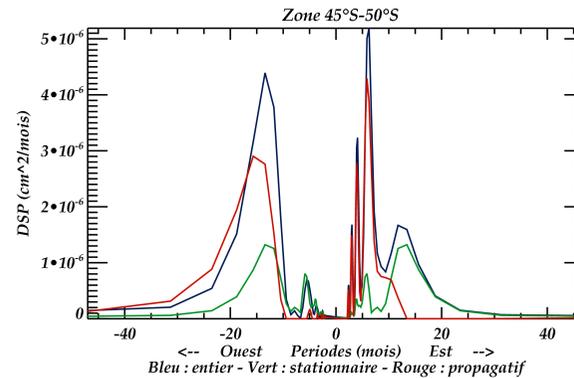
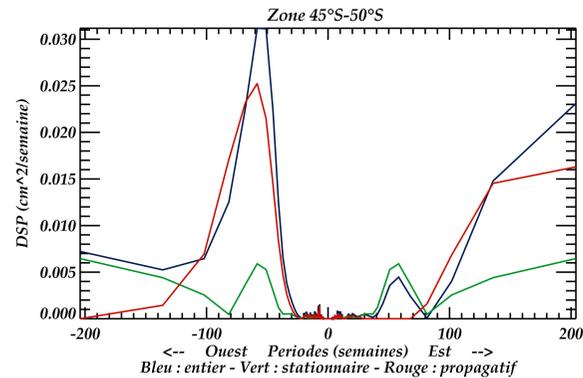
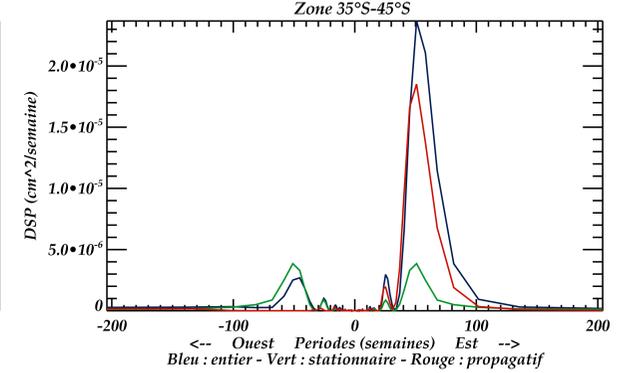
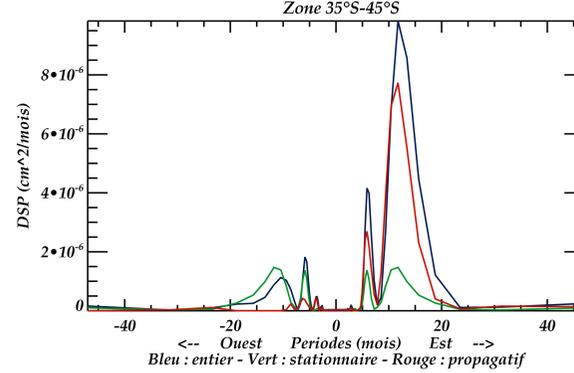
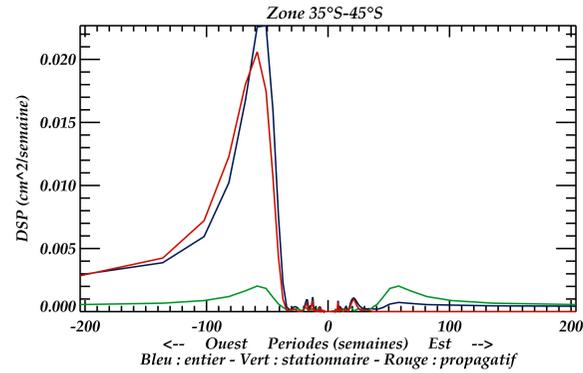
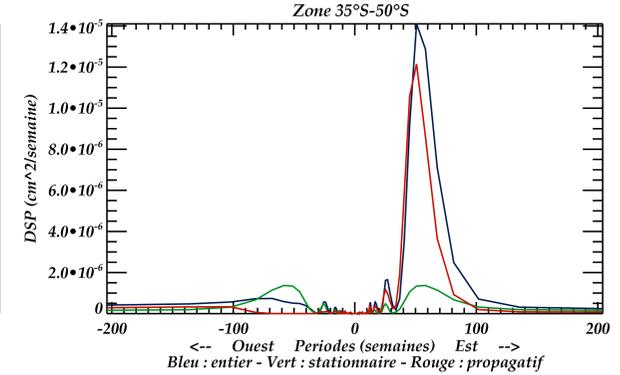


Figure 5.15 – Variations de la DSP associée au couple 3/0 selon les périodes fournies par l'analyse d'Hayashi sur les domaines 35°S-50°S (en haut), 35°S-45°S (au milieu) et 45°S-50°S (en bas) pour le signal entier (bleu), la partie stationnaire (vert) et la partie propagative (rouge) et pour les données altimétriques (à gauche), les données mensuelles de chlorophylle SeaWiFS (au milieu) et les données hebdomadaires de chlorophylle combinée SeaWiFS + MODIS/Aqua pour la période comprise entre juillet 2002 et mai 2006.

Zone 35°S – 50°S

Anomalies
de hauteur de mer

Chlorophylle SeaWiFS

Chlorophylle combinée

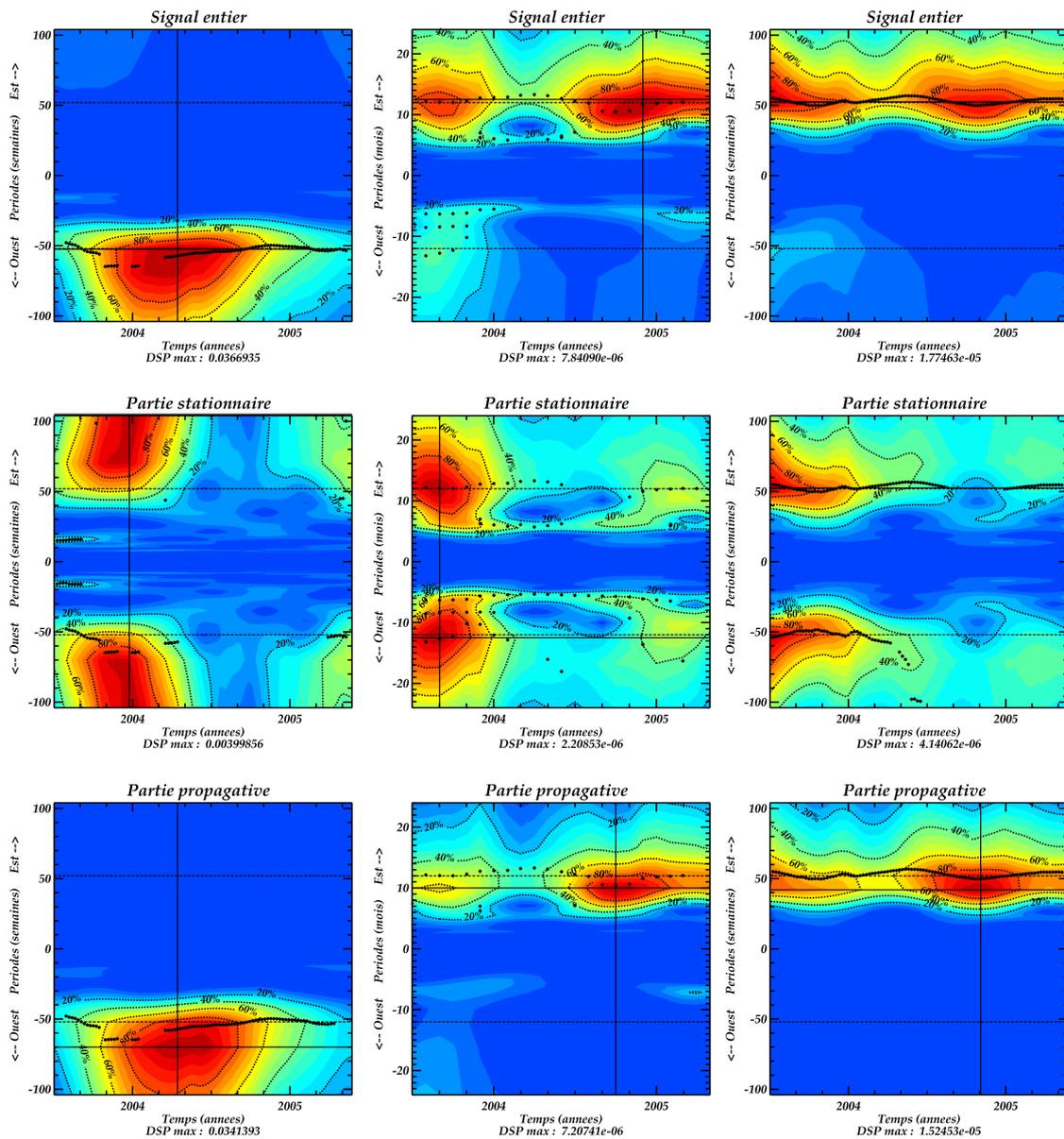


Figure 5.16 – Variations en temps et en période de la DSP associée à la paire 3/0 du signal entier (en haut) et des parties stationnaire (au milieu) et propagative (en bas) pour les données altimétriques (à gauche), les données mensuelles de chlorophylle SeaWiFS (au milieu) et les données hebdomadaires de chlorophylle combinée SeaWiFS + MODIS/Aqua (à droite) sur le domaine 35°S-50°S et pour la période comprise entre juillet 2002 et mai 2006. La demi-largeur de la fenêtre analysante est de 1 an.

Zone 35°S – 45°S

Altimétrie
de hauteur de mer

Chlorophylle SeaWiFS

Chlorophylle combinée

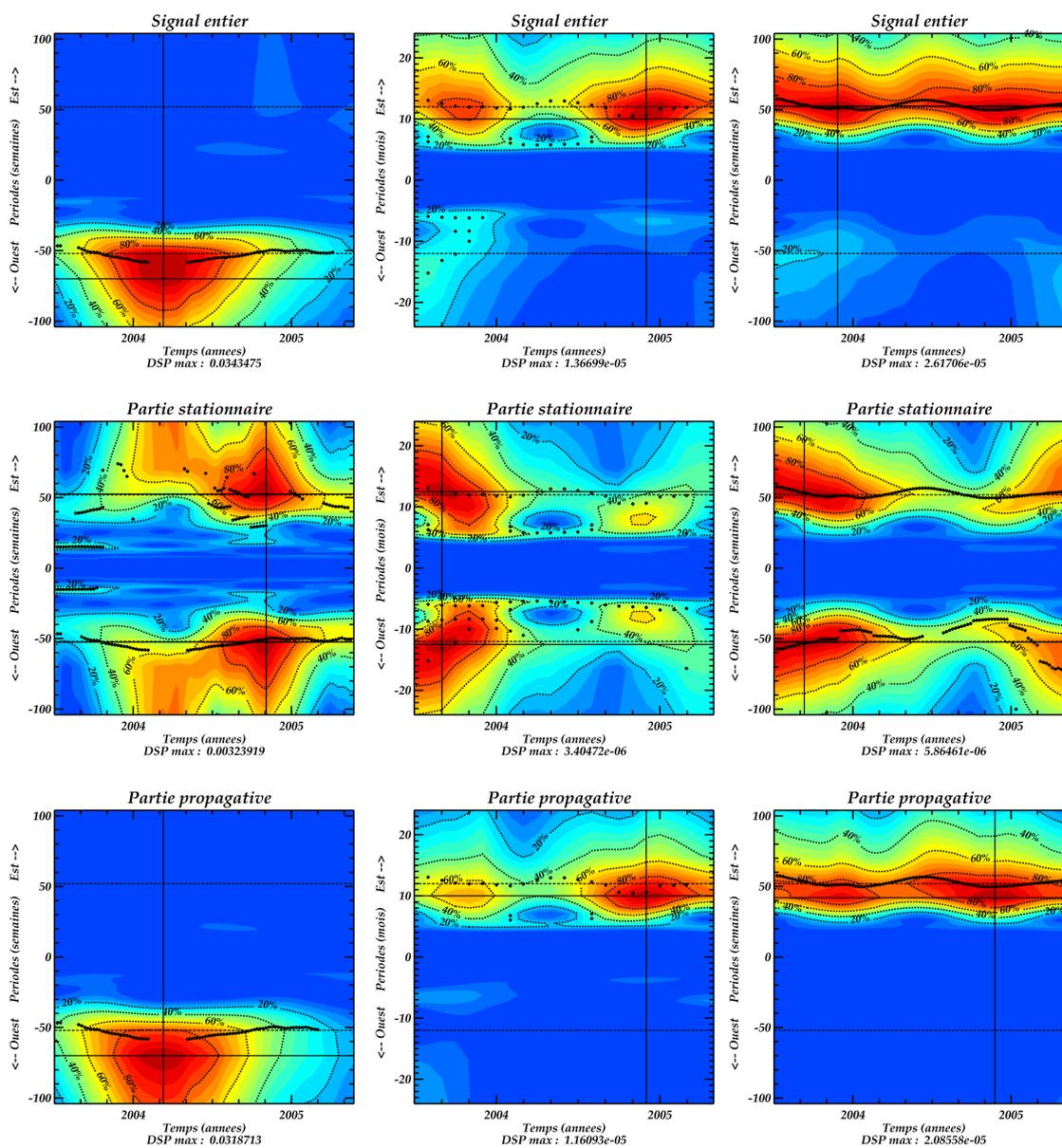


Figure 5.17 – Variations en temps et en période de la DSP associée à la paire 3/0 du signal entier (en haut) et des parties stationnaire (au milieu) et propagative (en bas) pour les données altimétriques (à gauche), les données mensuelles de chlorophylle SeaWiFS (au milieu) et les données hebdomadaires de chlorophylle combinée SeaWiFS + MODIS/Aqua (à droite) sur le domaine 35°S-45°S et pour la période comprise entre juillet 2002 et mai 2006. La demi-largeur de la fenêtre analysante est de 1 an.

Zone 45°S – 50°S

Anomalies
de hauteur de mer

Chlorophylle SeaWiFS

Chlorophylle combinée

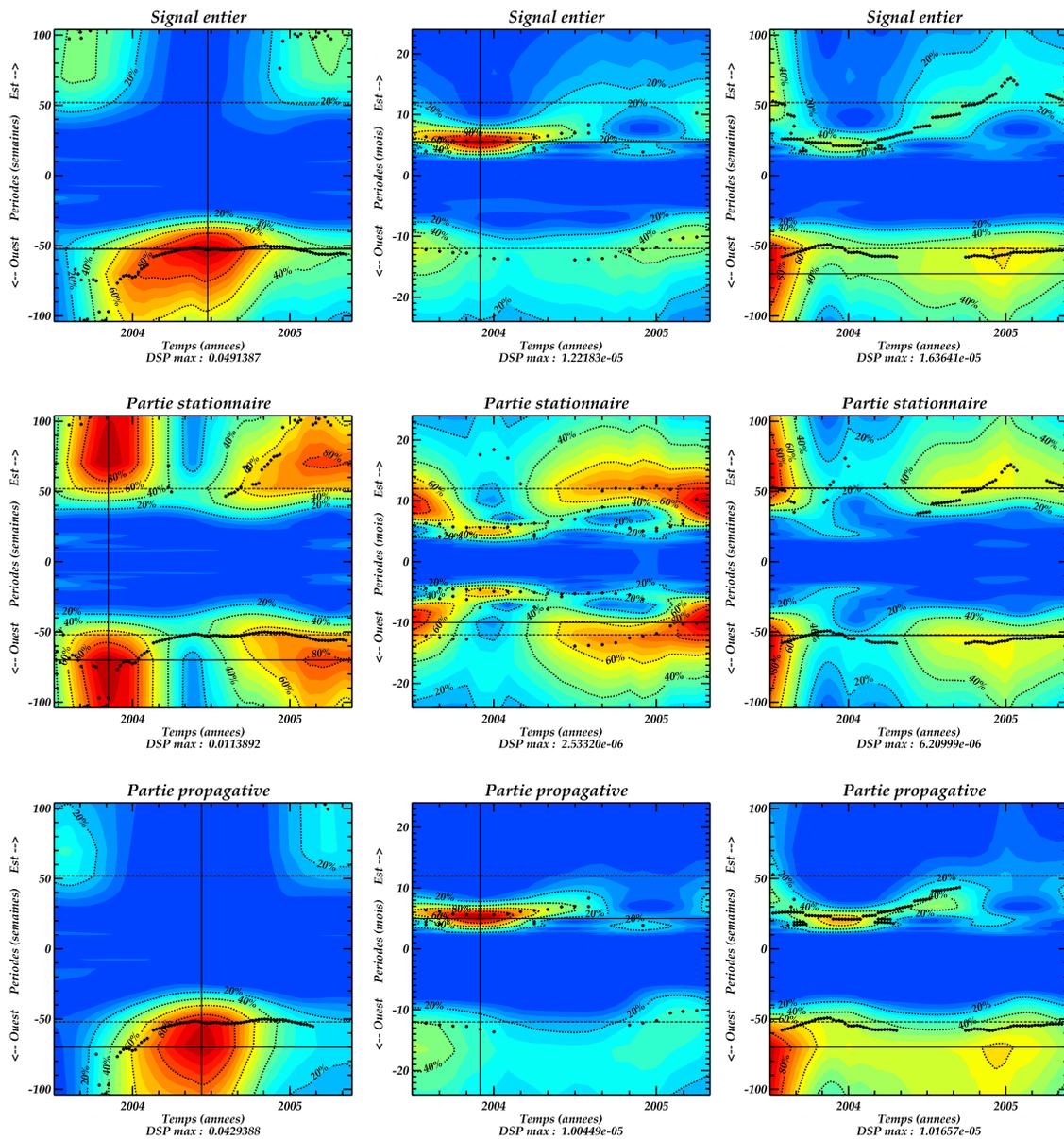


Figure 5.18 – Variations en temps et en période de la DSP associée à la paire 3/0 du signal entier (en haut) et des parties stationnaire (au milieu) et propagative (en bas) pour les données altimétriques (à gauche), les données mensuelles de chlorophylle SeaWiFS (au milieu) et les données hebdomadaires de chlorophylle combinée SeaWiFS + MODIS/Aqua (à droite) sur le domaine 45°S-50°S et pour la période comprise entre juillet 2002 et mai 2006. La demi-largeur de la fenêtre analysante est de 1 an.

Les résultats de l'analyse de Céron (Figures 5.16, 5.17 et 5.18, au milieu) confirment cette tendance. Sur la bande 35°S-50°S (Figure 5.16, au milieu en haut), la DSP du signal entier comporte deux noyaux (jusqu'à janvier 2004 et à partir de juin 2004) avec un maximum chacun (resp. septembre 2003 et novembre 2004). A l'identique, sur la bande 35°S-45°S (Figure 5.17, au milieu en haut), la DSP du signal entier comporte ces deux mêmes noyaux. Sur la bande 45°S-50°S (Figure 5.18, au milieu en haut), la DSP du signal entier montre que la période à ~1 an est moins marquée que la période à 6 mois. La première présente 2 maxima en juillet 2003 et février 2005, la seconde 1 noyau de juillet 2003 à juillet 2004 avec un maximum en décembre 2003.

Les données hebdomadaires combinées SeaWiFS + MODIS/Aqua de concentration en chlorophylle présentent des caractéristiques similaires aux données mensuelles de chlorophylle SeaWiFS. L'énorme avantage apporté par les données combinées pour ce type d'analyse est le gain en précision dans la DSP (les valeurs sont aussi plus fortes car le nombre d'observations est multiplié par 4) et le suivi de l'arête (sur les Figures 5.16, 5.17 et 5.18, à droite, on obtient de « jolies » arêtes continues). Sur la bande 35°S-50°S (Figure 5.16, en haut à droite), deux noyaux sont toujours présents dans la DSP du signal entier pour la période ~1 an propagative vers l'est : jusqu'à et à partir de mars 2004, avec des maxima en septembre 2003 et novembre 2004. De même, sur la bande 35°S-45°S (Figure 5.17, en haut à droite), la DSP du signal entier présente cette caractéristique à deux noyaux. Sur la bande 45°S-50°S (Figure 5.18, en haut à droite), deux noyaux sont aussi présents dans la DSP du signal entier pour la période ~1 an vers l'ouest (jusqu'à octobre 2003 et d'octobre 2004 à mai 2005) et dont les maxima ont respectivement lieu en juillet 2003 et décembre 2004. On peut aussi constater la présence d'une période à 6 mois propagative vers l'est, dont le noyau de DSP du signal entier associé s'étend entre septembre 2003 et février 2004, avec un maximum en décembre 2003 (Figure 5.18, en haut à droite). Notons que cette même période à 6 mois propagative vers l'est, vaguement visible sur les bandes 35°S-50°S et 35°S-45°S pour les données mensuelles SeaWiFS, ne l'est plus du tout sur les données combinées.

4.4 - Discussion et conclusion

Les analyses d'Hayashi et de Céron sont appliquées aux données de chlorophylle mensuelles SeaWiFS et aux données de chlorophylle hebdomadaires combinées SeaWiFS + MODIS/Aqua sur la période comprise entre septembre 1997 et mai 2006 (pour les données mensuelles SeaWiFS uniquement) puis la période comprise entre juillet 2002 et

mai 2006 (pour les deux jeux de données). Pour la première série temporelle étudiée, la méthode d'Hayashi permet une analyse des périodes allant jusqu'à 4.4 ans (vers l'ouest et vers l'est) alors que l'analyse de Céron ne permet une analyse que jusqu'à 2.2 ans. Pour la seconde période temporelle, le maximum des périodes analysées est de 2 ans pour la méthode d'Hayashi et de 1 an pour la méthode de Céron.

Nous avons tout d'abord pu constater l'amélioration apportée par les données de chlorophylle combinées SeaWiFS + MODIS/Aqua en comparaison aux données mensuelles SeaWiFS uniquement. En effet, non seulement l'énergie du signal est augmentée, et donc plus facile à analyser (moins de risque d'erreurs d'approximation informatique), mais l'analyse se fait sur plus de données (résolution temporelle plus fine) et elle est donc plus fiable : comme on peut le voir sur les Figures 5.16, 5.17 et 5.18 (à droite), l'arête est clairement plus nette dans le cas des données combinées.

Intéressons-nous maintenant au mode ZWN2. Si une analyse sur la période comprise entre juillet 2002 et mai 2006 ne nous permet certainement pas d'étudier ce mode, on devrait pouvoir commencer à le voir avec une analyse des données de chlorophylle sur la période comprise entre septembre 1997 et mai 2006, comme c'est le cas pour les anomalies de hauteur de mer (Figure 5.11, à gauche). Or, cette propagation vers l'est avec une période supérieure à 4 ans n'est pas du tout visible dans l'analyse des données de chlorophylle mensuelles SeaWiFS (Figure 5.11, à droite). Il semble donc que l'ACW ne puisse être résolue dans les données biologiques de concentration en chlorophylle-a sur l'archive 1997-2006. Ceci contredit les travaux de Le Quéré *et al.* [2002] qui, sur la base de seulement quatre années de données SeaWiFS (1997-2001), parlent de détection de la signature de l'ACW.

En ce qui concerne le mode ZWN3, nous avons vu qu'il était associé à deux propagations, l'une vers l'est avec une période de 6 mois, l'autre vers l'est (bande entière 35°S-50°S et bande 35°S-45°S) ou vers l'ouest (bande 45°S-50°S) avec une période de ~1 an, contrairement à l'altimétrie où il n'était associé qu'à une seule période de ~1 an propagative vers l'ouest sur toute la bande 35°S-50°S de surcroît.

La période à 6 mois propagative vers l'est n'est pas visible sur la bande 35°S-50°S, que ce soit sur les données de chlorophylle mensuelles SeaWiFS ou les données de chlorophylle hebdomadaires combinées SeaWiFS + MODIS/Aqua (Figure 5.15). Si on regarde par bande de latitude, on remarque que cette période n'est pas visible sur la bande

35°S-45°S mais est très nette sur la bande 45°S-50° S. Cette période, relative au mode ZWN3, pourrait être liée à l'oscillation semi-annuelle (SAO : Semi-Annual Oscillation). La SAO consiste en une contraction et une augmentation du creux de pression autour du continent Antarctique deux fois l'année, en réponse à des différences d'accumulation de chaleur entre le continent Antarctique et les océans environnants [van Loon, 1967]. Hurrell et van Loon [1994] ont montré que l'amplitude de cette oscillation semi-annuelle est la plus forte sur la bande de latitude 45°S-50°S dans les trois océans (avec des valeurs de 2-3 hPa) puis montre un minimum entre 55°S et 60°S (0.5 hPa) et augmente au niveau des côtes du continent Antarctique (4-5 hPa). L'augmentation de la pression cause une amplification de la structure de mode 3 atmosphérique dans la circulation autour du continent Antarctique [van den Broeke, 1998a]. Cela augmente le transport d'air des basses latitudes vers le continent Antarctique, un processus liant directement la SAO aux températures de surface au niveau du continent Antarctique. Nos résultats semblent montrer que la SAO a aussi un impact sur la biologie, phénomène particulièrement net tous les deux ans environ (signal entier, Figure 5.14) : août 1999, septembre 2001 et décembre 2003, ce dernier événement étant le plus marqué.

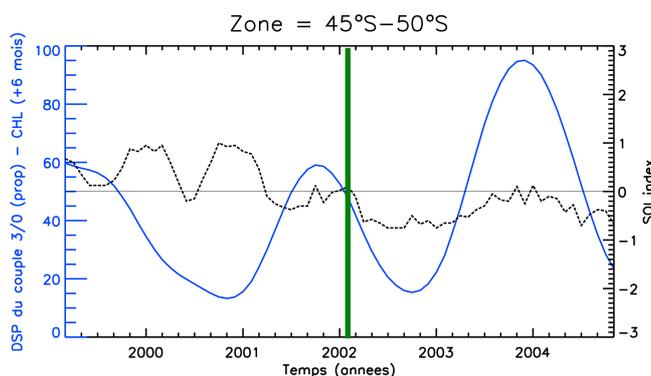


Figure 5.19 – Variations temporelles de la DSP de la partie propagative du mode ZWN3 pour les données mensuelles de chlorophylle-a SeaWiFS (en bleu) sur la bande 45°S-50°S. En pointillés noirs : variations temporelles de l'indice SOI du NOAA Climate Prediction Center [http://www.cpc.noaa.gov/data/indices/], lissé sur 3 mois.

La Figure 5.19 montre les variations temporelles de la DSP associée à la période de 6 mois propagative vers l'est et de l'indice SOI. Jusqu'à janvier 2002 (indiqué par un trait vert sur la Figure 5.19), la DSP et l'indice SOI sont anti-corrélés (-44%) : un accroissement de l'oscillation australe provoque une perte d'énergie de l'onde de période 6 mois propagative vers l'est, et vice-versa. A l'inverse, à partir de cette date, les deux mêmes signaux sont corrélés (79%) : une augmentation de l'oscillation australe a lieu en même temps qu'un renforcement de l'onde de période 6 mois propagative vers l'est et vice-versa. Ces derniers

résultats peuvent être comparés à ceux de van den Broeke [1998b] qui a montré que l'intensité de la SAO était liée à l'indice SOI en analysant les données de température de surface du continent Antarctique. Plus l'indice SOI est faible, moins la SAO est marquée. Par contre, la SAO ne semble pas avoir d'impact sur la physique de l'Océan Austral, car aucune période de 6 mois n'est visible dans les résultats de la Figure 5.13.

Pour les données de chlorophylle, la période à ~ 1 an a un comportement assez curieux : elle se propage vers l'est sur la bande de latitude 35°S - 45°S et vers l'ouest sur la bande 45°S - 50°S , alors qu'elle se propage vers l'ouest sur tout le domaine pour les données d'anomalies de hauteur de mer. Une importante question se pose alors : pourquoi la concentration en chlorophylle-*a* (i.e. la biologie) se propage dans le sens opposé aux anomalies de hauteur de mer (i.e. la physique) sur la bande 35°S - 45°S , alors que les deux vont dans le même sens sur la bande 45°S - 50°S ? La Figure 5.20, représentant les variations de l'indice SAM et de l'énergie de notre période à ~ 1 an en fonction du temps, peut aider à répondre à cette question. Sur la bande 35°S - 50°S , on constate une différence de comportement de nos signaux avant et après juin 2002 (Figure 5.20 et Tableau 5.4). Avant juin 2002, les DSP relatives aux anomalies de hauteur de mer (en bleu) et à la chlorophylle (en rouge) sont en opposition avec l'indice SAM : quand cet indice est important, les DSP diminuent, et vice-versa ; après juin 2002, c'est le contraire qui se passe : les DSP relatives aux anomalies de hauteur de mer et à la chlorophylle sont en phase avec l'indice SAM, i.e. les DSP augmentent avec l'indice SAM. D'autre part, les DSP relatives aux anomalies de hauteur de mer et à la chlorophylle sont en opposition de phase jusqu'en juin 2002 (anticorrélation de -73%) alors qu'elles sont en phase après juin 2002 (corrélations de 48%). Regardons maintenant par sous-bande de latitude. Dans la bande 35°S - 45°S , jusqu'en avril 2002, les DSP des deux signaux sont en opposition de phase avec l'indice SAM et anticorrélées (-60%). A partir d'avril 2002, les DSP des deux signaux sont en phase avec l'indice SAM et sont corrélées (corrélations supérieures à 50%). Dans la bande 45°S - 50°S , l'inverse se produit : avant septembre 2002, les DSP sont corrélées avec le SAM et entre elles (30%), et après septembre 2002, ces signaux sont anticorrélés (jusqu'à 80%). Les corrélations présentes dans les deux bandes 35°S - 45°S et 45°S - 50°S montrent sans ambiguïté un comportement très distinct entre ces deux bandes de latitude. Le comportement de la bande entière 35°S - 50°S se calcule sur celui de la bande la plus étendue en superficie, i.e. la bande 35°S - 45°S . Dans le courant de l'année 2002, il se produit une inversion dans le signe des corrélations pour ces deux bandes (Tableau 5.4). A la même époque, on peut remarquer un maximum très clair de la DSP stationnaire dans les données de hauteur de mer sur les deux bandes 35°S - 45°S et 45°S - 50°S (Figures 5.8 et 5.13).

Analyse de septembre 1997 à mai 2006

Analyse de juillet 2002 à mai 2006

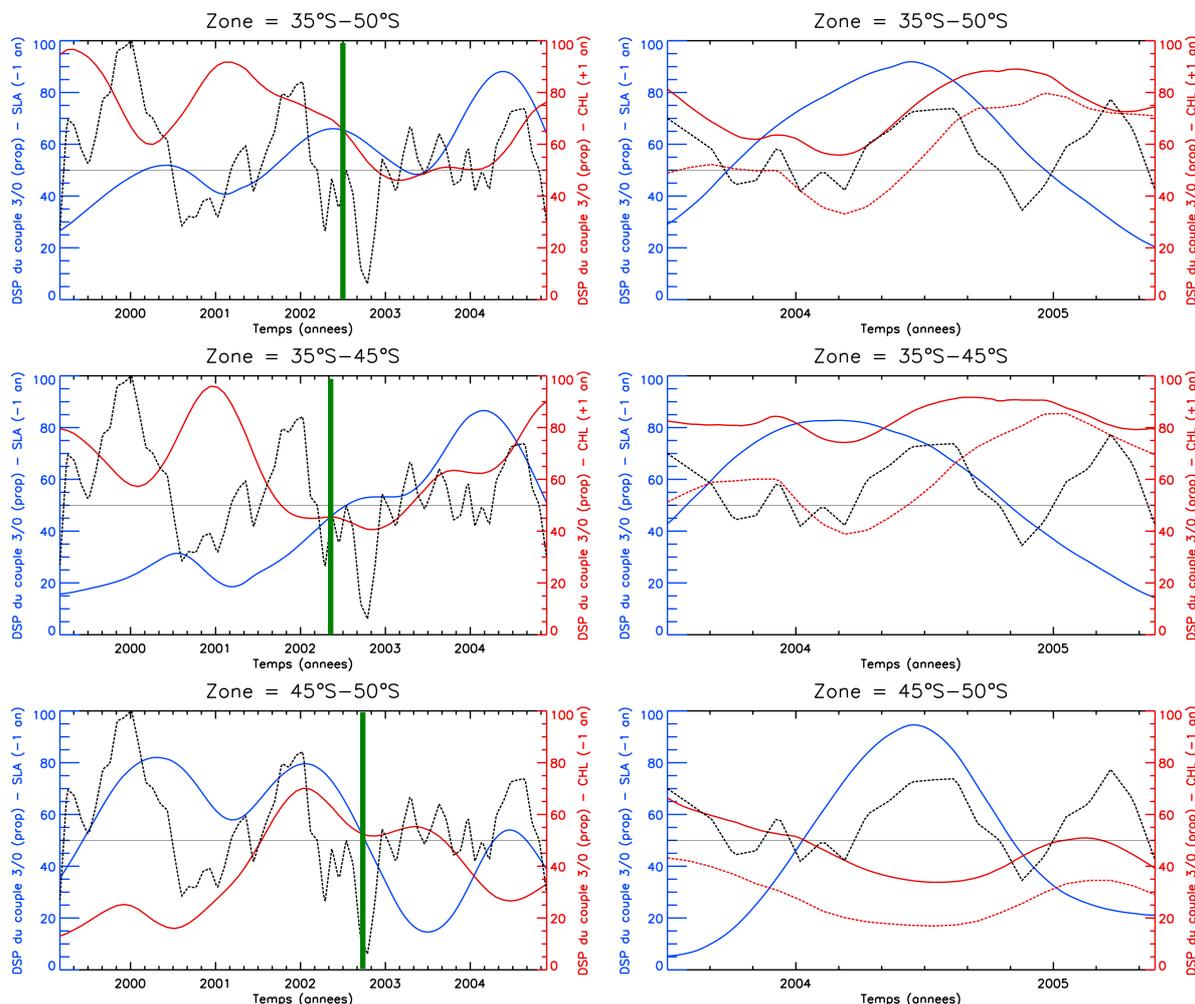


Figure 5.20 – A gauche : variations temporelles de la DSP de la partie propagative du mode ZWN3 pour les données altimétriques (en bleu) et SeaWiFS (en rouge) sur les régions 35°S-50°S (en haut), 35°S-45°S (au milieu) et 45°S-50°S (en bas). A droite : variations temporelles de la DSP de la partie propagative du mode ZWN3 pour les données altimétriques (en bleu), les données SeaWiFS (pointillés rouges) et les données combinées SeaWiFS + MODIS/Aqua (ligne continue rouge) sur les régions 35°S-50°S (en haut), 35°S-45°S (au milieu) et 45°S-50°S (en bas). En pointillés noirs : variations temporelles de l'indice SAM de Marshall [2003]. La ligne noire horizontale représente le 0 de l'indice SAM. La barre verticale verte est utilisée dans le calcul des corrélations du Tableau 5.4. A noter que les axes des abscisses entre les graphes de gauche et de droite ne vont pas jusqu'à la même date en raison de la taille de la fenêtre analysante.

Tableau 5.4 – Corrélations entre la DSP propagative des anomalies de hauteur de mer et l'indice SAM (SLA/SAM), la DSP propagative de la chlorophylle avec l'indice SAM (CHL/SAM) et entre les DSP propagatives des anomalies de hauteur de mer et de la chlorophylle (SLA/CHL) pour les bandes de latitudes 35°S-50°S, 35°S-45°S et 45°S-50°S, avant et après la barre verticale verte en 2002 indiquée Figure 5.20. Les corrélations négatives sont en **bleu**, les corrélations positives en **orange**.

	35°S – 50°S		35°S – 45°S		45°S – 50°S	
	Avant	Après	Avant	Après	Avant	Après
SLA / SAM	-10%	29%	-16%	42%	28%	-13%
CHL / SAM	-22%	10%	-55%	52%	15%	-37%
SLA / CHL	-73%	48%	-60%	52%	26%	-78%

Hall et Visbeck [2002] décrivent, en cas d'indice SAM positif, les changements attendus de la circulation océanique en réponse à l'intensification et déplacement des vents vers le pôle. Une anomalie dans le transport d'Ekman génère des anomalies négatives de SST au sud du Front Subantarctique et des anomalies positives de SST dans la zone du Front Subtropical. Dans cette zone (bande 35°S-45°S), la stratification accrue résultant des anomalies chaudes de SST réduit l'apport en surface des macronutriments nécessaires à la croissance phytoplanctonique. Les concentrations en chlorophylle sont ainsi peu élevées. A ces latitudes, pendant l'hiver austral, la lumière peut s'avérer être aussi un facteur limitant le développement du phytoplancton [Llido *et al.*, 2004]. Dans la bande plus au sud (45°S-50°S), un indice SAM positif s'accompagne de couches de mélange beaucoup plus profondes et donc d'une limitation plus forte de la production via le manque de disponibilité en lumière pour les cellules phytoplanctoniques. La lumière semble être le facteur limitant qui prédomine tout au long de l'année dans cette bande de latitude, en particulier au sud de l'Afrique du Sud [Llido *et al.*, 2004]. Il en résulte aussi des niveaux peu élevés de chlorophylle, sensiblement supérieurs cependant à ceux observés aux latitudes plus près de l'Equateur (Figure 5.21). Lovenduski et Gruber [2005] ont calculé des anticorrélations faibles de -0.2% et -0.16% entre les anomalies de logarithme de chlorophylle et l'indice SAM dans les zones subtropicale et subantarctique respectivement, sur la période 1997-2004.

Notre étude montre que le lien entre l'indice SAM et l'abondance en phytoplancton dans l'Océan Austral présente des modulations au cours du temps, et que la réponse de la région du front subtropical diffère nettement de celle du front subantarctique. Des questions restent cependant en suspens. La propagation de période ~1 an associée au mode ZWN3 visible sur les anomalies de hauteur de mer se fait vers l'ouest sur les deux bandes de latitude 35°S-45°S et 45°S-50°S. Pourquoi la propagation de période ~1 an associée au mode ZWN3 visible sur les concentrations en chlorophylle-*a* se fait vers l'ouest dans la bande 45°S-50°S et vers l'est dans la bande 35°S-45°S ? Comment les facteurs co-limitant la croissance du phytoplancton – à savoir la disponibilité en macro- ou micronutriments (le fer, par exemple) et l'énergie lumineuse disponible pour la photosynthèse – interagissent-ils et sont-ils liés à la propagation du mode ZWN3 ? Quels sont les mécanismes régulant la modulation de l'évolution spatio-temporelle des DSP des parties propagatives associées au mode ZWN3 visibles sur les concentrations de chlorophylle-*a* ? Il est probable aussi que l'océan ne réponde pas au forçage atmosphérique de type SAM de façon purement zonale, montrant des particularités par bassin océanique, qui ne sont pas accessibles par les méthodes mises en œuvre dans cette étude.

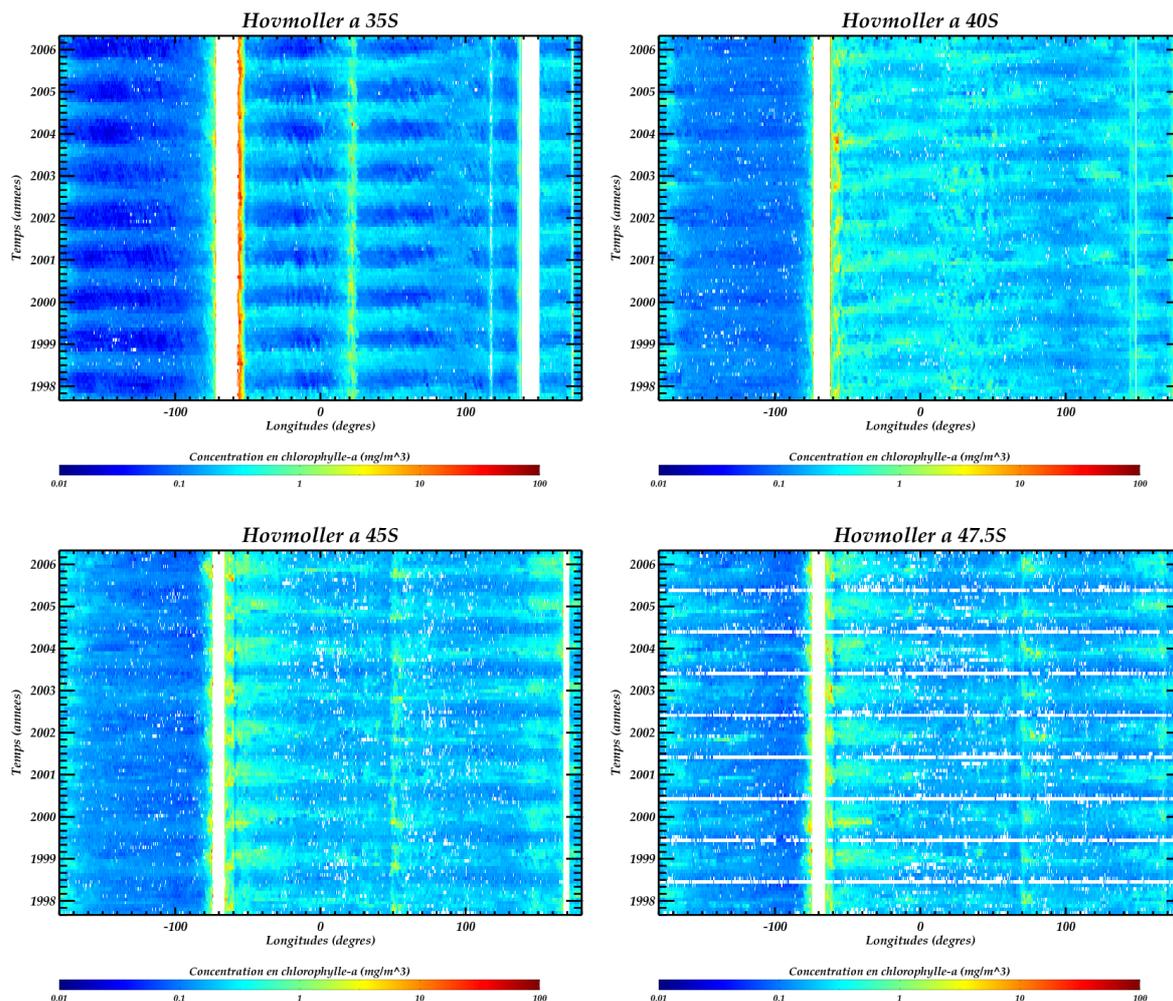


Figure 5.21 – Diagrammes Hovmoller (i.e. variations en longitude et en temps pour une latitude donnée) pour les latitudes 35°S, 40°S, 45°S et 47.5°S.

5 - Conclusion du chapitre

L'objectif de l'étude menée dans ce chapitre a été de déterminer les caractéristiques des modes dominants dans l'Océan Austral, tant sur les données physiques d'anomalies de hauteur de mer que sur les données biologiques de concentration en chlorophylle-a. Pour cela, deux méthodes prévues à l'origine pour des études météorologiques permettent d'extraire les modes dominants dans le signal, de décomposer leur énergie en une partie stationnaire et une partie propagative, et de visualiser leur(s) période(s) ou leur modulation en amplitude et en fréquence, selon la méthode.



Deux modes principaux ont été spécifiés.

Le mode ZWN2 est propagatif vers l'est avec une période de 5.5 ans environ. Il correspond à l'ACW. Si la signature de cette onde est visible dans les anomalies de hauteur de mer, ce n'est pas le cas pour la chlorophylle, même sur l'archive la plus longue de septembre 1997 à mai 2006.

Pour les anomalies de hauteur de mer, le mode ZNW3 est propagatif vers l'ouest avec une période de 1 an environ. Quant aux données de chlorophylle, ce mode est associé à deux périodes. La première, de 6 mois, est propagative vers l'est et liée à la SAO ; son importance est cyclique et revient tous les 2 ans environ. Elle semble être liée à l'oscillation australe et elle n'est pas visible sur les données physiques. La seconde, de 1 an environ, est liée au SAM ; elle se propage vers l'est (en opposition à l'altimétrie) sur la bande 35°S-45°S et vers l'ouest (dans le sens de l'altimétrie) sur la bande 45°S-50°S. Pour ces deux périodes (6 mois et ~1 an), il semble se passer un événement en 2002, qui modifie leur comportement par rapport au phénomène les causant (resp. oscillation australe et SAM). On a remarqué que cette année-là aussi, il y a un maximum très clair de la DSP stationnaire dans les données de hauteur de mer sur toutes les bandes.

L'intérêt d'utiliser des données combinées hebdomadaires de concentration en chlorophylle à la place des données mensuelles en provenance d'un seul capteur a aussi été montré. En effet, nous avons montré qu'avoir des données de chlorophylle à la même résolution que les données d'anomalies de hauteur de mer a rendu notre analyse plus fiable, non seulement par rapport à la DSP qui est plus élevée, mais aussi en ce qui concerne la continuité des arêtes qui s'en trouve nettement améliorée.

Conclusion Générale
et
Perspectives

CONCLUSION GENERALE ET PERSPECTIVES

L'objectif de ce travail de thèse était de concevoir, développer et tester des méthodes de combinaison de données couleur de l'eau, provenant des capteurs SeaWiFS et MODIS/Aqua, afin d'élaborer des produits de couleur de l'eau pour des applications en temps réel relevant de l'océanographie opérationnelle. Dans le cadre de cette étude, nous avons décidé de nous intéresser directement aux données de concentration en chlorophylle, puisqu'il s'agit de la variable que nous souhaitons cartographier.

Avant toute chose, il est nécessaire de caractériser les données sur lesquelles on travaille. C'est pourquoi nous avons, dans un premier temps, qualifié et quantifié la concentration en chlorophylle-*a* à l'échelle de l'océan global et par province biogéochimique, qui est une région ayant un comportement biologique et physique qui lui est propre. Dans un premier temps, nous avons pu constater que la distribution de la concentration en chlorophylle-*a* a la particularité d'être lognormale, i.e. c'est le logarithme base 10 de la concentration en chlorophylle qui a une distribution dite « normale » ou « gaussienne ». Cette propriété amène des contraintes dans les calculs, notamment dans le cadre des « matchups », i.e. la comparaison des données satellitaires avec les mesures *in situ* : les statistiques (i.e. erreur de mesure et biais) doivent être calculées sur le logarithme base 10 de la chlorophylle et non sur la chlorophylle directement. Les matchups ont montré que les données sont plus proches des observations *in situ* en océan ouvert qu'en côtier. Les données sont aussi plus fiables dans le biome des coups de vent d'ouest, puis dans les biomes polaire et enfin, des alizés. De plus, pour les reprocessings étudiés (i.e. #5.1 pour SeaWiFS et #1.1 pour MODIS/Aqua), les données issues de MODIS/Aqua sont globalement moins entachées d'erreur que celles de SeaWiFS, mais leur biais par rapport aux observations *in situ* est, lui, plus important. La différence de qualité entre les deux jeux de données est principalement notable dans l'Atlantique Tropical, où SeaWiFS surestime nettement la concentration en chlorophylle-*a*.

Après ces études quantitatives, nous nous sommes penchés sur les méthodes de combinaison proprement dites.



La première approche est la moyenne pondérée par l'erreur capteur. Bien que souvent considérée comme basique, elle améliore la couverture spatiale en ne prenant en compte que les valeurs satellitaires existantes. Rapide, nous avons vu que sa principale limitation est la discontinuité dans la précision du produit combiné. L'application de cette méthode en global sur l'année 2003 a montré que la qualité des données combinées varie très peu entre l'application sur les données brutes et l'application sur le logarithme base 10 des données. Néanmoins, l'application sur le logarithme base 10 des données et un produit combiné sur une grille $1/12^\circ \times 1/12^\circ$ (la résolution des données les plus grossières) a été sélectionnée, pour des raisons de rapidité d'exécution de l'algorithme.

La seconde approche est l'analyse objective. Etant donné un champ « first-guess » et des paramètres statistiques (variance, fonction de corrélation et rayons de corrélation) calculés directement à partir des données satellitaires, cette méthode permet d'interpoler des valeurs là où il n'y en a pas. En conséquence, la couverture spatiale des données combinées s'en trouve améliorée, par rapport à la moyenne pondérée. Son application aux données brutes a donné des résultats pauvres voire erronés, en particulier en raison de la distribution lognormale et de la dynamique temporelle de la chlorophylle. L'analyse objective a été et doit donc être appliquée au logarithme base 10 des données ; tous les paramètres statistiques doivent donc l'être aussi. Le champ « first-guess » utilisé est une climatologie mensuelle calculée par filtrage à un an par transformée de Fourier sur les données mensuelles SeaWiFS de toute l'archive. Les paramètres statistiques ont été calculés pour l'année 2003. Concernant la fonction de corrélation, nous avons choisi de modéliser la fonction empirique par un modèle inverse, qui est une fonction en forme de « chapeau pointu » ; nous avons vu que la caractéristique plus ou moins « pointue » de la fonction de corrélation dépendait de l'emplacement : les observations sont mieux corrélées en côtier (fonction « moins pointue ») qu'en océan ouvert (fonction « plus pointue »). Des tests de sensibilité aux paramètres ont été réalisés. Ces paramètres jouent sur le lissage et la qualité du produit combiné, ainsi que la rapidité du calcul : seule une estimation rigoureuse des paramètres d'entrée permet d'obtenir un produit combiné le plus fiable possible.

La dernière approche est une technique innovante et prometteuse basée sur la transformée en ondelettes. Elle se décompose en deux étapes. La première est une estimation des données manquantes de chaque image à combiner (i.e. relative à une journée et à un capteur), basée sur le phénomène de persistance des détails : il existe une relation entre les coefficients d'ondelette père et fils pour chaque orientation (i.e. horizontale, verticale et diagonale, suivant le concept d'algorithme pyramidal). La seconde est la fusion d'images proprement dite, qui consiste à combiner par moyenne pondérée par l'erreur les coefficients d'ondelettes des images par niveau et orientation. Plusieurs tests nous ont

révéle que l'utilisation de l'ondelette de Battle-Lemarié d'ordre 3 convenait le mieux à nos données. De plus, une application sur le logarithme base 10 des données est préférable afin d'éviter d'obtenir numériquement d'éventuelles concentrations en chlorophylle négatives. A travers une première application, nous avons montré les possibilités qu'une telle méthode pouvait nous apporter : non seulement la couverture spatiale est améliorée au maximum mais les structures sont conservées le plus possible. De plus, contrairement à la moyenne pondérée et l'analyse objective, l'avantage supplémentaire de cette méthode est que l'on récupère une image combinée de la résolution la plus fine et non de la résolution la plus grossière.

Dans un dernier chapitre, nous avons montré l'intérêt de la combinaison des données couleur de l'eau à travers la caractérisation des modes dominants dans l'Océan Austral (i.e. la bande circumpolaire 35°S-50°S) sur les données biologiques de concentration en chlorophylle-a et de comparer aux données physiques d'anomalies de hauteur de mer. Pour cela, deux méthodes prévues à l'origine pour des études météorologiques permettent d'extraire les modes dominants dans le signal, de décomposer leur énergie en une partie stationnaire et une partie propagative, et de visualiser leur(s) période(s) ou leur modulation en amplitude et en fréquence, selon la méthode.

Deux modes principaux ont été spécifiés. Un mode zonal de longueur d'onde 180° et de période 5.5 ans environ, correspondant à l'ACW, est visible dans les anomalies de hauteur de mer, mais pas dans les données de concentration en chlorophylle-a. Le second mode est de longueur d'onde 120° et de période 1 an. Son comportement est différent sur les anomalies de hauteur de mer et sur les données de concentration en chlorophylle-a. Pour la hauteur de mer, il est propagatif vers l'ouest avec une période de 1 an environ. Quant aux données de chlorophylle, ce mode est associé à deux périodes. La première, de 6 mois, est propagative vers l'est et liée à la SAO ; son importance est cyclique et revient tous les 2 ans environ. Elle semble être liée à l'oscillation australe et elle n'est pas visible sur les données physiques. La seconde, de 1 an environ, est liée au SAM ; elle se propage vers l'est (en opposition à l'altimétrie) sur la bande 35°S-45°S et vers l'ouest (dans le sens de l'altimétrie) sur la bande 45°S-50°S. Pour ces deux périodes (6 mois et ~1 an), il semble se passer un événement en 2002, qui modifie leur comportement par rapport au phénomène les causant (resp. oscillation australe et SAM).

Nous avons montré qu'avoir des données de chlorophylle à la même résolution que les données d'anomalies de hauteur de mer a rendu notre analyse plus fiable pour le calcul de la densité spectrale de puissance et l'extraction d'arêtes.

Les perspectives portent sur trois points.

En ce qui concerne la moyenne pondérée par l'erreur capteur et l'analyse objective, les algorithmes sont opérationnels. Les statistiques (erreur de mesure, biais, variance, fonctions et rayons de corrélation) ont été déterminées pour l'année 2003 uniquement, d'une part car il s'agit de notre année de test, et d'autre part pour des questions d'espace disque (les données journalières sont extrêmement volumineuses) et de temps de calcul. Bien entendu, il sera nécessaire de recalculer ces paramètres sur toute l'archive pour une application des algorithmes sur d'autres années. Les cartes d'erreur de mesure et de biais pour chaque capteur devront être estimées à partir de toute l'archive, afin d'obtenir des valeurs les plus correctes possibles. Il est possible que ces valeurs varient mensuellement, et cela sera à vérifier pour toute l'archive. Pour effectuer ces matchups, il faudra d'abord construire une base de données internationale de concentration en chlorophylle-*a in situ* de surface de 1997 à nos jours la plus complète possible, ce qui représente, il est vrai, un travail considérable puisqu'à ce jour, une telle base n'existe pas. Quant à la variance et aux fonctions et rayons de corrélation, il faudra les calculer sur l'archive SeaWiFS des anomalies de 1997 à nos jours ; une comparaison des résultats obtenus avec l'archive MODIS/Aqua pourrait se révéler intéressante. Enfin, il a été dommage de ne pas pouvoir ajouter les données issues du capteur européen MERIS à la combinaison de SeaWiFS et MODIS/Aqua durant ce travail de thèse : un maximum de capteurs à combiner est toujours préférable.

Quant à la méthode basée sur la transformée en ondelettes, de nombreuses portes ont été ouvertes lors de ce travail de thèse mais il reste encore beaucoup d'étapes à franchir avant de faire tourner cet algorithme très prometteur au niveau global ! Tout d'abord, nous avons utilisé une ondelette connue la plus adaptée à nos données (ondelette de Battle-Lemarié d'ordre 3). Dans un premier temps, il faudrait calculer l'ondelette idéale propre à la concentration en chlorophylle-*a* en se basant sur les travaux de Turiel et Parga [2000b]. Puis il faudrait effectuer d'autres tests de reconstruction, avec diverses localisations pour les données manquantes (nuage de points à différents pourcentages, « gros trous » positionnés à divers endroits, combinaison des deux, etc.), pour bien analyser le comportement de l'algorithme de reconstruction. Ensuite, le calcul de l'erreur reste à déterminer. Enfin, l'algorithme de combinaison reste à écrire pour le global (gestion des continents) et l'opérationnel (pour le moment, l'algorithme est gourmand en temps de calcul). Il n'est pas possible de considérer l'océan global en une fois ; il faut donc appliquer l'algorithme par sous-image dont la taille optimale (rapport qualité / temps de calcul ; il faut considérer la taille et la position des traces des satellites) est à déterminer : c'est ce qu'on appelle le « tuilage ». Quelques tests préliminaires ont été réalisés pour trouver cette taille adaptée : 128 x 128 pixels pour l'image MODIS/Aqua et 64 x 64 pixels pour l'image SeaWiFS semble

convenable, mais ceci est à approfondir. La méthode d'assemblage des « tuiles » est aussi à déterminer : une juxtaposition directe des imagerie reconstruites et combinées n'est pas possible car les intersections seraient par trop visibles. Une technique pourrait être un lissage par la méthode du gradient, déjà utilisée dans notre méthode de reconstruction. Evidemment, on peut penser que, moyennant des études statistiques préliminaires comparables à celles réalisées dans ce travail de thèse, cet algorithme puissant peut être appliqué à toute donnée se présentant sous forme d'image pixellisée : les radiances, la température de surface de la mer, etc.

Le dernier point concerne les ondes dominantes dans l'Océan Austral. Appliquer les méthodes d'analyse d'Hayashi et de Céron aux données de température de surface de la mer pourrait ouvrir des pistes quant à l'explication concernant le sens de propagation du mode ZWN3. On pourrait aussi modifier la méthode d'analyse en remplaçant l'actuelle étape 1 (i.e. l'analyse de Fourier spatiale) par une analyse en ondelettes 2D, qui permettrait de récupérer l'emplacement exact des périodes analysées. Enfin, pour suivre cette idée concernant la position des ondes, il est probable que le mode ZWN3 ne soit pas strictement propagatif zonalement, mais qu'il existe un déplacement méridional : une technique d'analyse pourrait consister à considérer comme matrice d'étude une ceinture circumpolaire dont l'intervalle de latitude varie en fonction de la longitude.

Dans le cadre du projet européen MERSEA, trois méthodes de combinaison des données couleur de l'eau ont été développées et testées, dont deux fonctionnent de manière opérationnelle : la moyenne pondérée par l'erreur capteur et l'analyse objective. La troisième méthode, basée sur la transformée en ondelettes, est très puissante et offre de grandes possibilités. C'est pourquoi des efforts doivent être faits pour continuer le travail de recherche et développement dans ce domaine, dans le cadre du prochain GMES Marine Core Services.

Bibliographie

BIBLIOGRAPHIE

- AIKEN, J., N. REES, S. HOOKER, P. HOLLIGAN, A. BALE, D. ROBINS, G. MOORE, R. HARRIS, et D. PILGRIM, 2000: The Atlantic Meridional Transect: overview and synthesis of the data, *Progress in Oceanography*, **45**, 257-312.
- ANTOINE, D., 1998: pports de la télédétection spatiale de la "couleur de l'océan" à l'océanographie, *Océanis*, **24 (2)**, 81-150.
- BATTLE, G., 1987: A block spin construction of ondelettes. Part 1: Lemarié functions, *Communication in Mathematics and Physics*, **110**, 601-615.
- BEHRENFELD, M.J., W.E. ESAIAS, et K.R. TURPIE, 2002: Assessment of Primary Production at the Global Scale, In: Phytoplankton Productivity Carbon Assimilation in Marine and Freshwater Ecosystems, Chap. 7, *Blackwell Sci.*, P.J. le B. Williams, D.N. Thomas, et C.S. Reynolds, Eds., 156-186.
- BELAMARI, S., 2002: Rôle des coups de vent d'Ouest dans le déclenchement d'évènements El Niño, *thèse de l'Université de Toulouse III*, 236 p.
- BROWN, J., COLLING, A., PARK, D., PHILLIPS, J., ROTHERY, D., et WRIGHT, J., 1989: Ocean Circulation, *The Open University*, 238 pp.
- BUCCIGROSSI, R.W., et E.P. SIMONCELLI, 1997: Progressive Wavelet Image Coding based on a Conditional Probability Model, *Proceedings ICASSP, Munich, Germany*, April 21-24.
- BUCCIGROSSI, R.W., et E.P. SIMONCELLI, 1999: Image Compression via Joint Statistical Cgacterization in the Wavelet Domain, *IEEE Transactions on Image Processing*, **8**, **12**, 1688-1701.
- CAI, W., et P.G. BAINES, 2001: Forcing of the Antarctic Circumpolar Wave by El Niño-Southern Oscillation teleconnections, *Journal of Geophysical Research*, **106 (C5)**, 9019-9038.
- CAMPBELL, J. W., 1995: The lognormal distribution as a model for bio-optical variability in the sea, *Journal of Geophysical Research*, **100 (C7)**, 13237-13254.
- CARRIL, A.F. et A. NAVARRA, 2001: Low-frequency variability of the Antarctic Circumpolar Wave, *Geophysical Research Letters*, **28 (24)**, 4623-4626.
- CERON, J.P., 2000 : L'Analyse de Gabor et son application à l'étude de l'évolution temporelle en champs 2D, *Note interne de la Direction de la Climatologie*.
- CÉRON, J.P., et GUÉRÉMY, J.F., 1999: Validation of the Space-Time Variability of African Easterly Waves Simulated by the CNRM GCM, *Journal of Climate*, **12**, 2831-2855,.



- CÉRON, J.-P., S. BELAMARI, et L. AUGER, 2000: Gabor analysis and its application to the characterization of the time evolution of a 2D field, EGS Meeting, Nice, France.
- COHEN, A., I. DAUBECHIES, B. JAWERTH et P. VIAL, 1993: Multiresolution analysis, wavelets, and fast wavelet transform on an interval, *CRAS Paris, Ser. A.*, **312**, 417-421.
- DAUBECHIES, I., 1988: Orthonormal bases of compactly supported wavelets, *Communication on Pure and Applied Mathematics*, XLI:909-996.
- DUCET, N., P.-Y. LE TRAON, et G. REVERDIN, 2000: Global high resolution mapping of ocean circulation from Topex/Poseidon and ERS-1 and -2, *Journal of Geophysical Research*, **105 (C8)**, 19,477-19,478.
- EMERY, W.J., ET THOMSON, R.E., 2001: Data Analysis Methods in Physical Oceanography – Second and Revised Edition, *Elsevier*, 638 pp.
- EPLER, R.E. JR, R.A. BARNES, S.W. BAILEY, et P.J. WERDELL, 2003: Changes to the vicarious calibration of SeaWiFS, in: PATT, F.S., R.A. BARNES, R.E. EPLER, JR, B.A. FRANZ, W.D. ROBINSON, G.C. FELDMAN, S.W. BAILEY, P.J. WERDELL, R. FROUIN, R.P. STUMPF, R.A. ARNONE, R.W. GOULD, JR, P.M. MARTINOLICH, et V. RANSIBRAHMANAKUL, Algorithm Updates for the Fourth SeaWiFS Data Reprocessing, NASA Technical Memorandum 2003-206892, Vol. 22, S.B. Hooker et E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, (in press).
- EVANS, R.H. et H.R. GORDON, 1994: Coastal zone color scanner “system calibration”: A retrospective examination, *Journal of Geophysical Research*, **99 (C4)**, 7293-7308.
- FELDMANN, G. C., 1994: The Living Ocean. SeaWiFS: Studying ocean color from Space, *National Aeronautics and Space Administration*, EP-307.
- FRANZ, B.A., 2003: A long-term Intercomparison of Oceanic Optical Property Retrievals from MODIS-Terra and SeaWiFS, in *MODIS validation, Data Merger, and other activities accomplished by SIMBIOS Project: 2002-2003*, NASA-TM-2003 (Greenbelt, Maryland, NASA Goddard Space Flight Center), 19-31.
- FRANZ, B.A., 2003: Methods for Assessing the Quality and Consistency of Ocean Color Products, http://oceancolor.gsfc.nasa.gov/DOCS/methods/sensor_analysis_methods.html.
- FRISCH, U., 1996: Turbulence: The Legacy of A. N. Kolmogorov, *Cambridge University Press*, 310 p.
- GABOR, D., 1946: Theory of communication, *Journal of the Institution of Electrical Engineers*, **93 (3)**, 429-457.
- GORDON, H.R., et M.H. WANG, 1994: Retrieval of water-leaving radiance and aerosol optical-thickness over the oceans with SeaWiFS – a preliminary algorithm, *Applied Optics*, **33 (3)**, 443-452.

- GREENBERG, A.E., 1995: Standard Methods for the Examination of Water and Wastewater (19th Edition), *American Public Health Association*, 1325 p.
- GREGG, W.W., W.E. ESAIS, G.C. FELDMAN, R. FROUIN, S.B. HOOKER, C.R. MCCLAIN, and R.H. WOODWARD, 1998: Coverage opportunities for Global Ocean Color in a Multimission Era, *IEEE Transactions on Geoscience and Remote Sensing*, **36 (5)**, 1620-1627.
- GREGG, W.W., and M.E. CONKRIGHT, 2001: Global seasonal climatologies of ocean chlorophyll: Blending *in situ* and satellite data for the Coastal Zone Color Scanner era, *Journal of Geophysical Research*, **106 (C2)**, 2499-2516.
- FOUGNIE, B., P. HENRY, A. MOREL, D. ANTOINE et F. MONTAGNER, 2002 : Identification and Characterization of Stable Homogeneous Oceanic Zones : Climatology and Impact on In-Flight Calibration of Space Sensors over Reyleigh Scattering, *Ocean Optics XVI*, Santa Fe, NM, November 18-22, 2002.
- HALL, A., et M. VISBECK, 2002: Synchronous Variability in the Southern Hemisphere Atmosphere, Sea Ice, and Ocean Resulting from the Annular Mode, *Journal of Climate*, **15**, 3043-3057.
- HAYASHI, Y., 1977: On the Coherence between Progressive and Retrogressive Waves and a Partition of Space-Time Power Spectra into Standing and Traveling Parts, *Journal of Applied Meteorology*, **16**, 368-373.
- HAYASHI, Y., 1979: A Generalized Method of Resolving Transient Disturbances into Standing and Traveling Waves by Space-Time Spectral Analysis, *Journal of the Atmospheric Sciences*, **36**, 1017-1029.
- HAYASHI, Y., 1982: Space-Time Spectral Analysis and its Applications to Atmospheric Waves, *Journal of the Meteorological Society of Japan*, **60 (1)**, 156-171.
- HAYASHI, Y., et GOLDBERGER, D.G., 1993: Tropical 40-50- and 25-30-Day Oscillations Appearing in Realistic and Idealized GFDL Climate Models and the ECMWF Dataset, *Journal of the Atmospheric Sciences*, **50 (3)**, 464-494.
- HARTMANN, D.L., et F. LO, 1998: Wave-Driver Zonal Flow Vascillation in the Southern Hemisphere, *Journal of Atmospheric Sciences*, **55**, 1303-1312.
- HUGHES, C.W., P.L. WOODWORTH, M.P. MEREDITH, V. STEPANOV, T. WHITWORTH, et A.R. PYNE, 2003 : Coherence of Antarctic sea levels, Southern Hemisphere Annular Mode, and flow through Drake Passage, *Geophysical Research Letters*, **30 (9)**, 1464, 10.1029/2003GL017240.
- HURRELL, J.W., et H. VAN LOON, 1994: A modulation of the atmospheric annual cycle in the Southern Hemisphere, *Tellus*, **46A**, 325-338.



- JACOBS, G.A., et J.L. MITCHELL, 1996: Ocean circulation variations associated with the Antarctic Circumpolar Wave, *Geophysical Research Letters*, **23 (21)**, 2947-2950.
- JERLOV, N.G., 1976: Marine Optics, Vol. 14 of *Elsevier Oceanography Series*, Elsevier, Amsterdam, 231 pp.
- JGOFS, 1994: Protocols for the Joint Global Ocean Flux Study (JGOFS) Core Measurements, *Manual and Guides 29, UNESCO*, 170 p.
- KWIATKOWSKA, E.J., and G.S. FARGION, 2001: Merger of Ocean Color Data from Multiple Satellite Missions within the SIMBIOS project, *Remote Sensing of the Atmosphere, Ocean, Environment and Space, SPIE Symposium*, invited paper.
- KWIATKOWSKA, E.J., and G.S. FARGION, 2002a: Merger of Ocean Color Data from Multiple Satellite Missions within the NASA SIMBIOS Project Office, *Proceedings of the 5th IEEE International Conference on Information Fusion*, **1**, 291-298.
- KWIATKOWSKA, E.J., and G.S. FARGION, 2002b: Application of machine-learning techniques toward the creation of a consistent and calibrated global chlorophyll concentration baseline dataset using remotely sensed ocean color data, *IEEE Transactions on Geoscience and Remote Sensing*, **41(12)**, 2844-2860.
- KWIATKOWSKA, E.J., 2003: Comparisons of Daily Global Ocean Color Data Sets: MODIS-Terra/Aqua and SeaWiFS, in *MODIS validation, Data Merger, and other activities accomplished by SIMBIOS Project: 2002-2003*, NASA-TM-2003 (Greenbelt, Maryland, NASA Goddard Space Flight Center), 2-18.
- LEMARIE, P.G., 1988: Ondelette à localisation exponentielle, *Journal de Mathématiques Pures et Applications*, **67**, 227-236.
- LE QUÉRÉ, C., L. BOPP, ET I. TEGEN, 2002: Antarctic circumpolar wave impact on marine biology : A natural laboratory for climate change study, *Geophysical Research Letters*, **29 (0)**, doi:10.1029/2001GL014585.
- LE TRAON, P.-Y., F. NADAL, et N. DUCET, 1998: An Improved Mapping Method of Multisatellite Altimeter Data, *Journal of Atmospheric and Oceanic Technologies*, **15**, 522-534.
- LLIDO, J., E. MACHU, J. SUDRE, I. DADOU et V. GARÇON, 2004: Variability of the biological front south of Africa from SeaWiFS and a coupled physical-biological model, *Journal of Marine Research*, **62 (4)**, 595-609.
- LONGHURST, A., 1995: Seasonal cycles of pelagic production and consumption, *Progress in Oceanography*, **36**, 77-167.
- LONGHURST, A., 1998: Ecological Geography of the Sea, *Academic Press*, San Diego, 398 p.
- MALLAT, S., 1989: A theory for Multiresolution Signal Decomposition: the Wavelet Representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11 (7)**, 674-693.

- MALLAT, S., 1998: A wavelet tour of signal processing, Academic Press, second edition.
- MALLAT, S., et W.L. HWANG, 1992: Singularity detection and processing with wavelets, *IEEE Transactions on Information Theory*, **38 (2)**, 617-643.
- MARITORENA, S., and D.A. SIEGEL, 2005: Consistent merging of satellite ocean color data sets using a bio-optical model, *Remote Sensing of Environment*, **94**, 429-440.
- MARSHALL, G.J., 2003 : Trends in the Southern Annular Mode from observations and reanalyses, *Journal of Climate*, **14**, 4134-4143.
- MCCLAIN, C.R., W. ESAIAS, G. FELDMAN, R. FROUIN, W. GREGG, and S. HOOKER, 2002: The Proposal for the NASA Sensor Intercalibration and Merger for Biological and Interdisciplinary Oceanic Studies (SIMBIOS) Program, NASA TM-2002-210008, NASA/GSFC, Greenbelt, MD, USA, 54 p.
- MEISTER, G., E.J. KWIATKOWSKA, B.A. FRANZ, F.S. PATT, G.C. FELDMAN, and C.R. MCCLAIN, 2005: Moderate-Resolution Imaging Spectroradiometer ocean color polarization correction, *Applied Optics*, **44 (26)**, 5524-5535.
- MELIN, F., 2003: Potentiel de la télédétection pour l'analyse des propriétés optiques du système océan-atmosphère et application à l'estimation de la photosynthèse phytoplanctonique, *Thèse de doctorat de l'Université de Toulouse III*, 514 p.
- MINSTER, J.-F., 1997: La Machine Océan, *Flammarion*, 304 p.
- MOBLEY, C.D., 1994: Light and water, *Academic Press*, 592 p.
- MOREL, A., ET B. GENTILI, 1991: Diffuse reflectance of oceanic waters: its dependence on sun angle as influenced by the molecular scattering contribution, *Applied Optics*, **30**, 4427-4438.
- MOREL, A., et L. PRIEUR, 1977: Analysis of variations in ocean colour, *Limnology and Oceanography*, **22**, 709-722.
- MORLET, J., 1994: Sampling theory and wave propagation, in *12th International Conference on Pattern Recognition*.
- O'REILLY, J.E., S. MARITORENA, B.G. MITCHELL, D.A. SIEGEL, K.L. CARDER, S.A. GARVER, M. KAHRU, and C. MCCLAIN, 1998: Ocean color chlorophyll algorithms for SeaWiFS, *Journal of Geophysical Research*, **103C**, 24937-24953.
- PETERSON, R.G., et W.B. WHITE, 1998: Slow oceanic teleconnections linking the Antarctic Circumpolar Wave with tropical ENSO, *Journal of Geophysical Research*, **103 (C11)**, 24573-24583.
- POTTIER, C., J.-P. CÉRON, J. SUDRE, I. DADOU, S. BELAMARI, et V. GARÇON, 2004: Dominant propagating signals in sea level anomalies in the Southern Ocean, *Geophysical Research Letters*, **31**, L11305, doi:10.1029/2004GL019565.



- POTTIER, C., V. GARÇON, G. LARNICOL, J. SUDRE, P. SCHAEFFER, et P.-Y. LE TRAON, 2006: Merging SeaWiFS and MODIS/Aqua Ocean Color Data in North and Equatorial Atlantic using Waighted Averaging and Objective Analysis, *IEEE Transactions on Geoscience and Remote Sensing*, Volume 44, Issue 11, Part 2, doi:10.1109/TGRS.2006.878441, 3436-3451.
- PRENTICE, I.C., et al., 2001: The Scientific Basis: Contribution of Working Group 1 to the Third Assessment Report of the IPCC, in *Climate Change 2001*, J.T. Houghton et al. (Eds.), *Cambridge University Press*, UK, 183-237.
- REYNOLDS, R.W., et T.M. SMITH, 1994: Improved Global Sea Surface Temperature Analyses Using Optimum Interpolation, *Journal of Climate*, **7**, 929-948.
- RINTOUL, S.R., C.W. HUGHES, et D. OLBERS, 2001: The Antarctic Circumpolar Current System, in *Ocean Circulation and Climate 2001*, G. Siedler, J. Church, et J. Gould (Eds.), *Academic Press*, 271-302.
- ROBINS, D.B., et J. AIKEN, 1996: The Atlantic Meridional Transect: an oceanographic research programme to investigate physical, chemical, biological and optical variables of the Atlantic Ocean, *Underwater Technology*, **21(4)**, 8-14.
- RUIZ, S. et G. LARNICOL, 2004: Multivariate analysis of SST combining Infrared and Microwave satellite data, in The 5th Science Team Meeting of the Global Ocean Data Assimilation Experiment High Resolution Sea Surface Temperature Pilot Project (GHRSSST – PP) meeting, Townsville.
- SABINE, C. L., R. FEELY, N. GRUBER, R. M. KEY, K. LEE, J. L. BULLISTER, R. WANNINKHOF, C. S. WONG, D. W. R. WALLACE, B. TILBROOK, F. J. MILLERO, T.-H. PENG, A. KOZYR, T. ONO, et A. F. RIOS, 2004: The Oceanic Sink for Anthropogenic CO₂, *Science*, **305**, 367-371.
- SARMIENTO, J.L., T.M.C. HUGUES, R.J. STOUFFER et S. MANABE, 1998: Simulated response of the ocean carbon cycle to anthropogenic climate warming, *Nature*, **393**, 245-249.
- SHETTLE, E.P., et R.W. FENN, 1979: Models for the aerosols of the lower atmosphere and the effects of humidity variations on their optical properties, *AFGL-TR-79-0214*, *Environmental Research Paper*, **675**, NTIS, ADA 085851, 94 p.
- SIMONCELLI, E.P., 1999: Modeling the Joint Statistics of Images in the Wavelet Domain, *Proceedings SPIE 44th Annual Meeting, Denver, Colorado*, **3813**, 188-195.
- SIMONCELLI, E.P., et R.W. BUCCIGROSSI, 1997: Embedded Wavelet Image Compression Based on a Joint Probability Model, *4th IEEE International Conference on Image Processing, Santa Barbara, CA., October 26-29*.
- STRANG, G., et T. NGUYEN, 1996: Wavelets and filter banks, *Wellesley-Cambridge Press*.

- TCHERNIA, P., 1978: Océanographie Régionale : Description Physique des Océans et des Mers, *Centre d'Édition et de Documentation de l'École Nationale de Techniques Avancées*.
- THOMPSON, D.W.J., et J.M. WALLACE, 2000: Annular Modes in the Extratropical Circulation. Part I: Month-to-month Variability, *Journal of Climate*, **13**, 1000-1016.
- TORRÉSANI, B., 1995: Analyse continue par ondelettes, *InterEditions / CNRS Editions*, 239 p.
- TURIEL, A., et N. PARGA, 2000a: The multifractal structure of contrast changes in natural images: from sharp edges to textures, *Neural Computation*, **12**, 763-793.
- TURIEL, A., et N. PARGA, 2000b: Multifractal wavelet filter of natural images, *Physical Review Letters*, **85**, 3325-3328.
- TURIEL, A., et A. DEL POZO, 2002: Reconstructing Images From Their Most Singular Fractal Manifold, *IEEE Transactions on Image Processing*, **11 (4)**, 345-350.
- TURIEL, A., J.-P. NADAL, et N. PARGA, 2003: Orientational minimal redundancy wavelets: from edge detection to perception, *Vision Research*, **43 (9)**, 1061-1079.
- VAN DEN BROEKE, M., 1998a: The semi-annual oscillation and Antarctic climate. Part1: influence on near surface temperatures (1957-79), *Antarctic Science*, **10 (2)**, 175-183.
- VAN DEN BROEKE, M., 1998b: The semi-annual oscillation and Antarctic Climate. Part 2: recent changes, *Antarctic Science*, **10 (2)**, 184-191.
- VAN LOON, H., 1967: The half-yearly oscillation in the middle and high southern latitudes and coreless winter, *Journal of Atmospheric Science*, **24**, 472-486.
- VAN LOON, H., 1972: Pressure in the Southern Hemisphere, *American Meteorology Society, Meteorological Monographs*, **13 (35)**, 59-86.
- VENEGAS, S.A., 2003: The Antarctic Circumpolar Wave: A Combination of Two Signals?, *Journal of Climate*, **16 (15)**, 2509-2525.
- WALD, L., 2002: Data Fusion – Definition and Architectures – Fusion of images of different spatial resolutions, *Presses de l'École des Mines de Paris*, 200 p.
- WATTERSON, I.G., 2000 : Southern Midlatitude Zonal Wind Vascillation and its Interaction with the Ocean in GCM simulations, *Journal of Climate*, **13**, 562-578.
- WERDELL, P.J., et S.W. BAILEY, 2002: The SeaWiFS Bio-Optical Archive and Storage System (SeaBASS): Current Architecture and Implementation – Introduction, *NASA Tech. Memo 2002-211617*, G.S. Fargion and C.R. McClain, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland.
- WHITE, W.B., 2004: Comment on « Synchronous Variability in the Southern Hemisphere Atmosphere, Sea Ice, and Ocean Resulting from the Annular Mode », *Journal of Climate*, **17 (11)**, 2249-2254.



- WHITE, W.B. et R.G. PETERSON, 1996: An Antarctic circumpolar wave in surface pressure, wind, temperature and sea-ice extent, *Nature*, **380**, 600-702.
- WHITWORTH, T., 1983: Monitoring the net transport of the Antarctic Circumpolar Current at Drake Passage, *Journal of Physical Oceanography*, **13**, 2045-2057.
- WHITHWORTH, T., et R.G. PETERSON, 1985: The volume transport of the Antarctic Circumpolar Current from three year bottom pressure measurements, *Journal of Physical Oceanography*, **15**, 810-816.
- XIONG, J., G. TOLLER, V. CHIANG, J. SUN, J. ESPOSITO, et W. BARNES, 2005: MODIS Level 1B Algorithm Theoretical Basis Document, http://www.mcst.ssai.biz/mcstweb/info/online_doc.html.

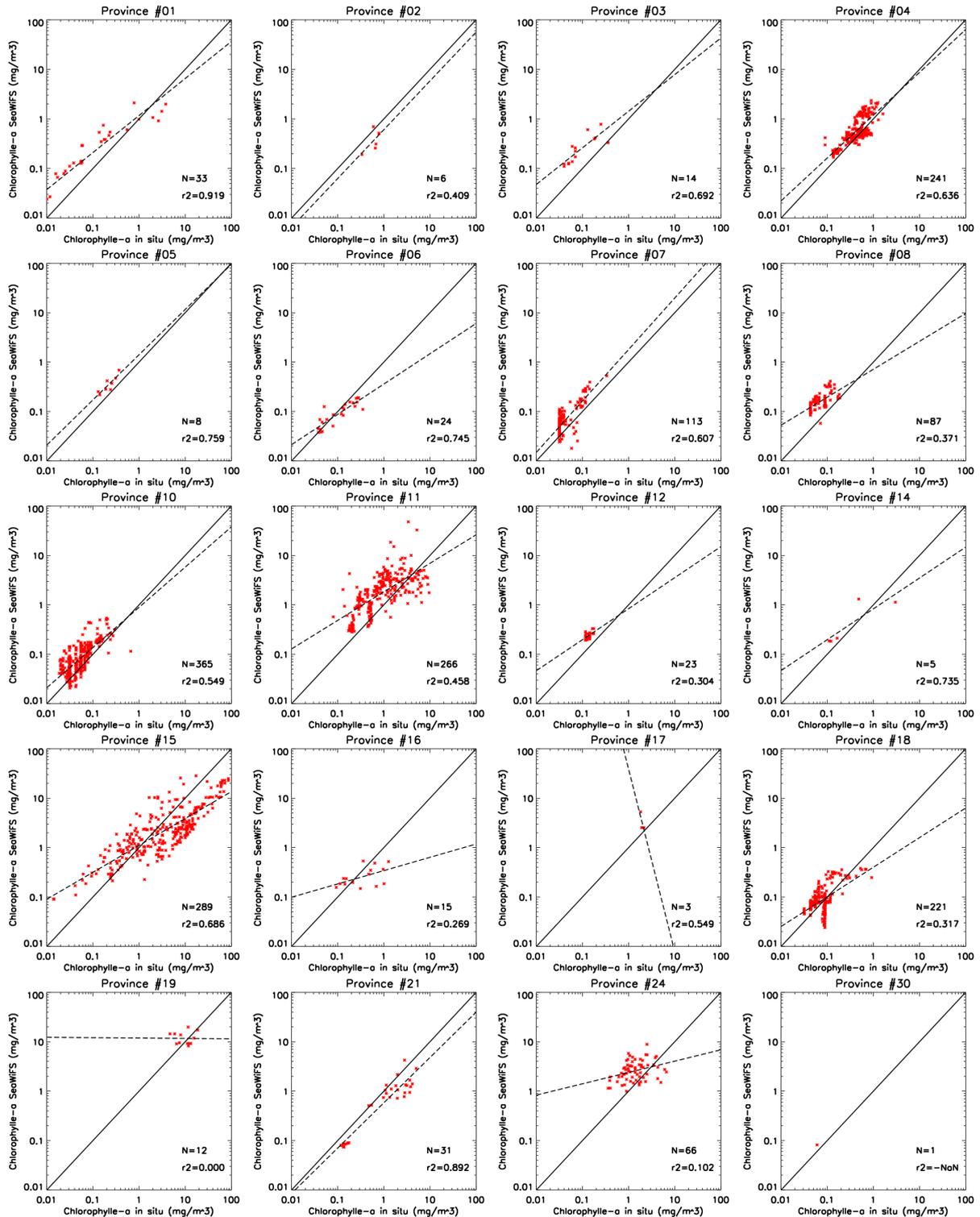
Annexes

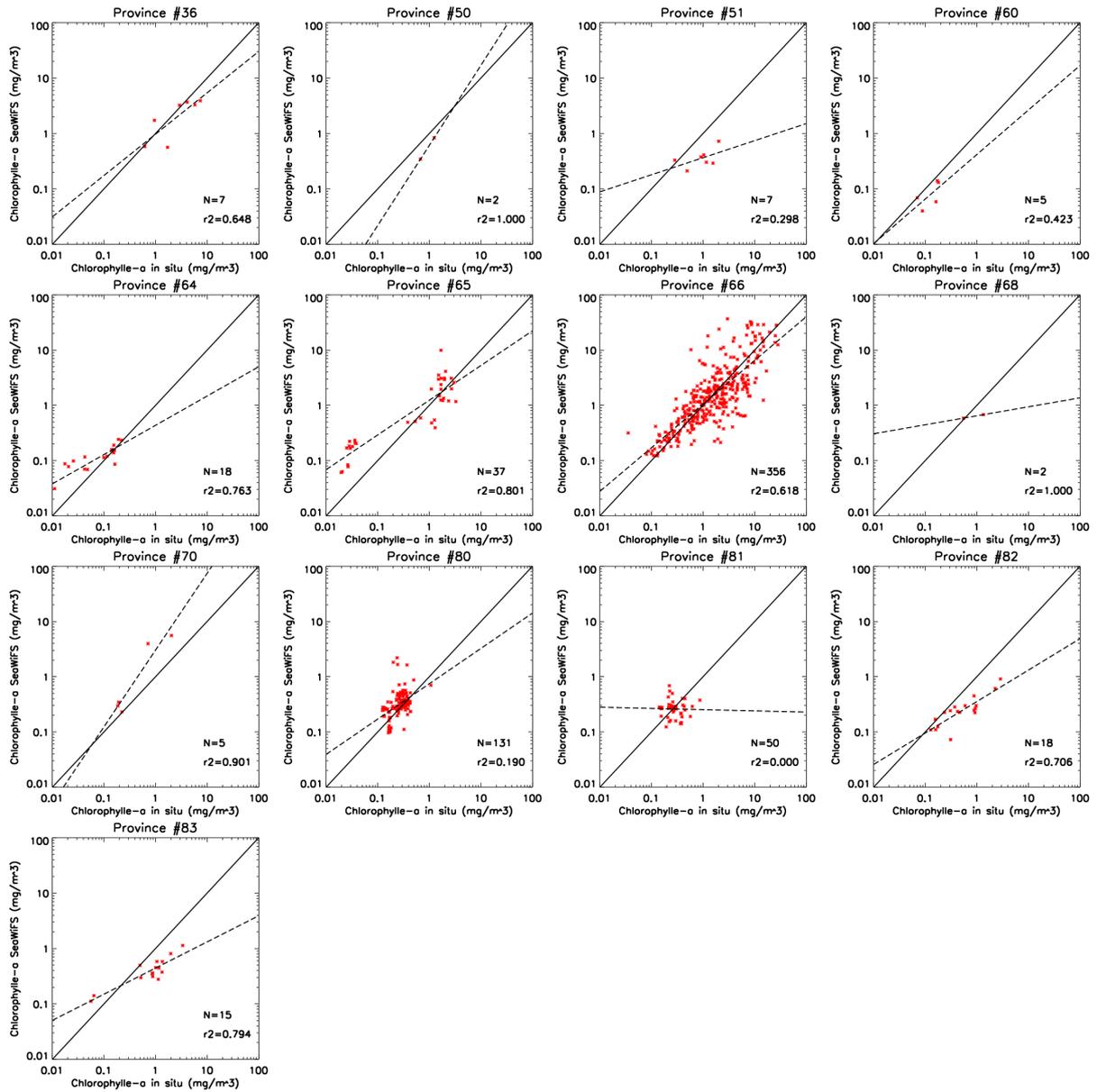
ANNEXE A

1 – MATCHUPS PAR PROVINCES.....	261
1.1 – SeaWiFS.....	261
1.2 – MODIS/Aqua.....	263
2 – ANALYSE OBJECTIVE	265
2.1 – Les cartes de climatologie mensuelle.....	265
2.2 – Tests de sensibilité.....	267
2.2.1 – Cartes pour le 25 mars 2003	267
2.2.2 – Rayon de corrélation temporel.....	267
2.2.3 – Centrage des données	268
2.2.4 – Nombre minimum d’observations dans la bulle d’influence.....	269
2.2.5 – Nombre maximum d’observations dans la bulle d’influence.....	269
2.2.6 – Sous-échantillonnage des positions d’étude	270
2.2.7 – Rayons de corrélation spatiaux.....	271
2.2.8 – Erreur de mesure.....	272
2.2.9 – Biais	273
2.2.10 – Variance	274
2.2.11 – Coefficient de la fonction de corrélation	275
3 – TRANSFORMEE EN ONDELETTES : TESTS.....	276
4 – MISSIONS SATELLITAIRES ALTIMETRIQUES.....	278
4.1 – Intérêt de la mesure	278
4.2 – Principe de la mesure.....	278
4.3 – La mission TOPEX-POSEIDON.....	281
4.3.1 – Introduction.....	281
4.3.2 – Le satellite	281
4.3.3 – L’orbite.....	283
4.3.4 – Le traitement des données	284
4.3.5 – Les produits TOPEX-POSEIDON	285
4.4 – Les missions ERS-1/2	285
4.5 – La mission Jason-1	287
4.5.1 – Introduction.....	287
4.5.2 – Le satellite	287
4.5.3 – Le segment sol	289
4.6 – La mission Envisat	289
5 – RESULTATS COMPLEMENTAIRES DES ANALYSES D’HAYASHI ET DE CERON	290
5.1 – Analyses entre septembre 1997 et mai 2006.....	290
5.1.1 – Les données utilisées	290
5.1.2 - Maxima des DSP obtenues par l’analyse d’Hayashi pour l’altimétrie	290
5.1.3 – Résultats de l’analyse d’Hayashi pour le couple 2/0	291
5.1.4 – Résultats de l’analyse de Céron appliquée aux données mensuelles SeaWiFS de concentration en chlorophylle pour le couple 2/0	292
5.1.5 – Résultats de l’analyse de Céron appliquée aux données altimétriques pour le couple 2/0.....	293

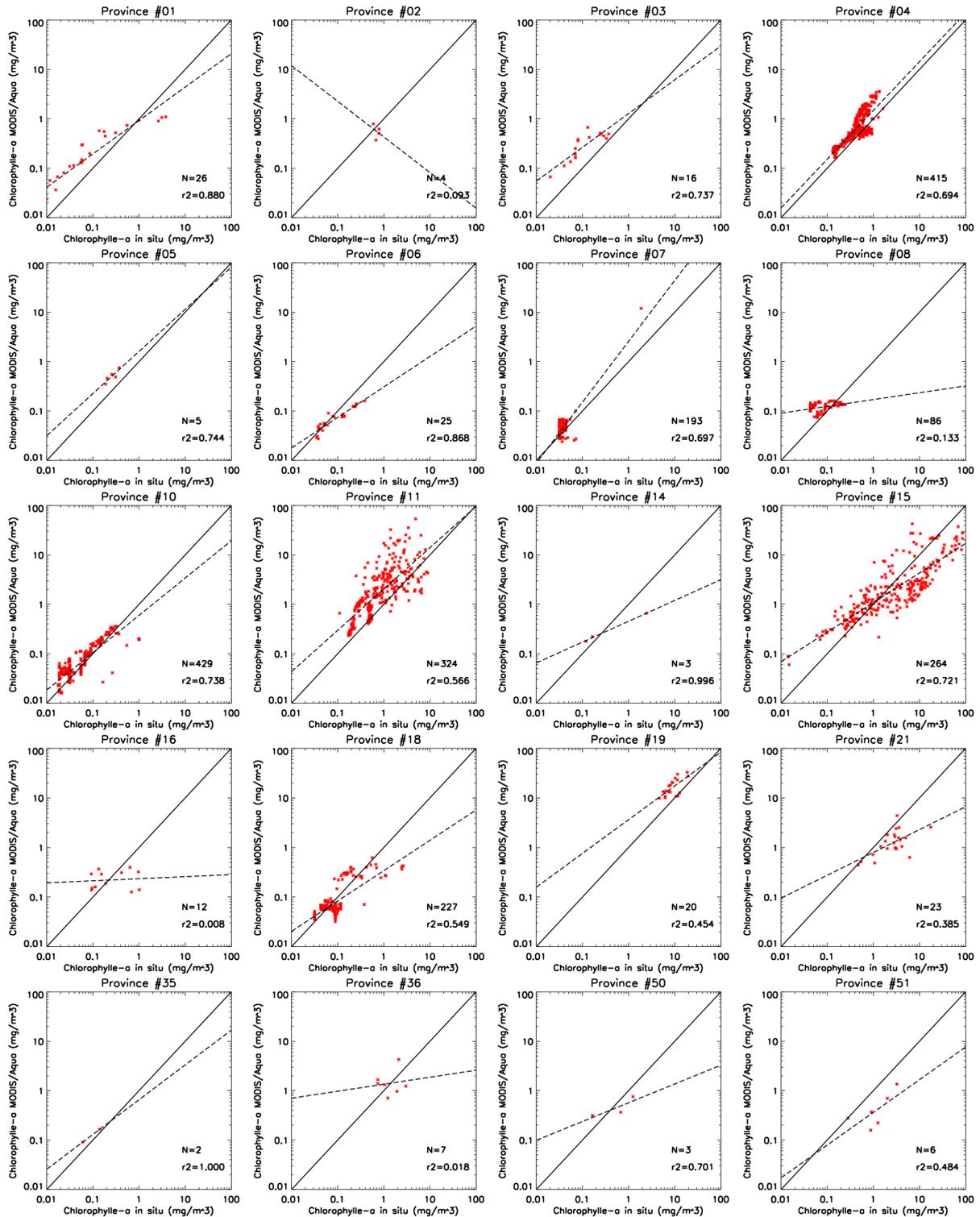
1 - Matchups par provinces

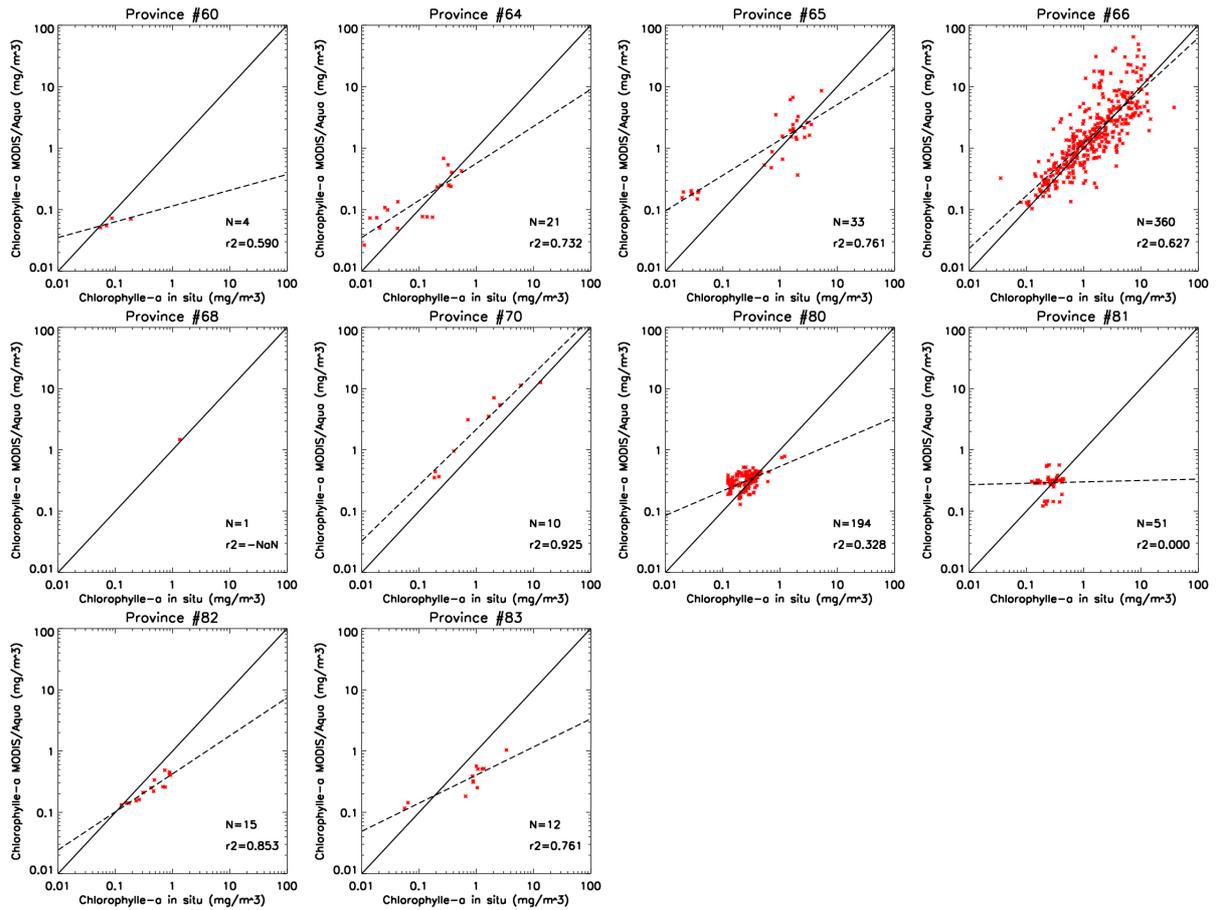
1.1 - SeaWiFS





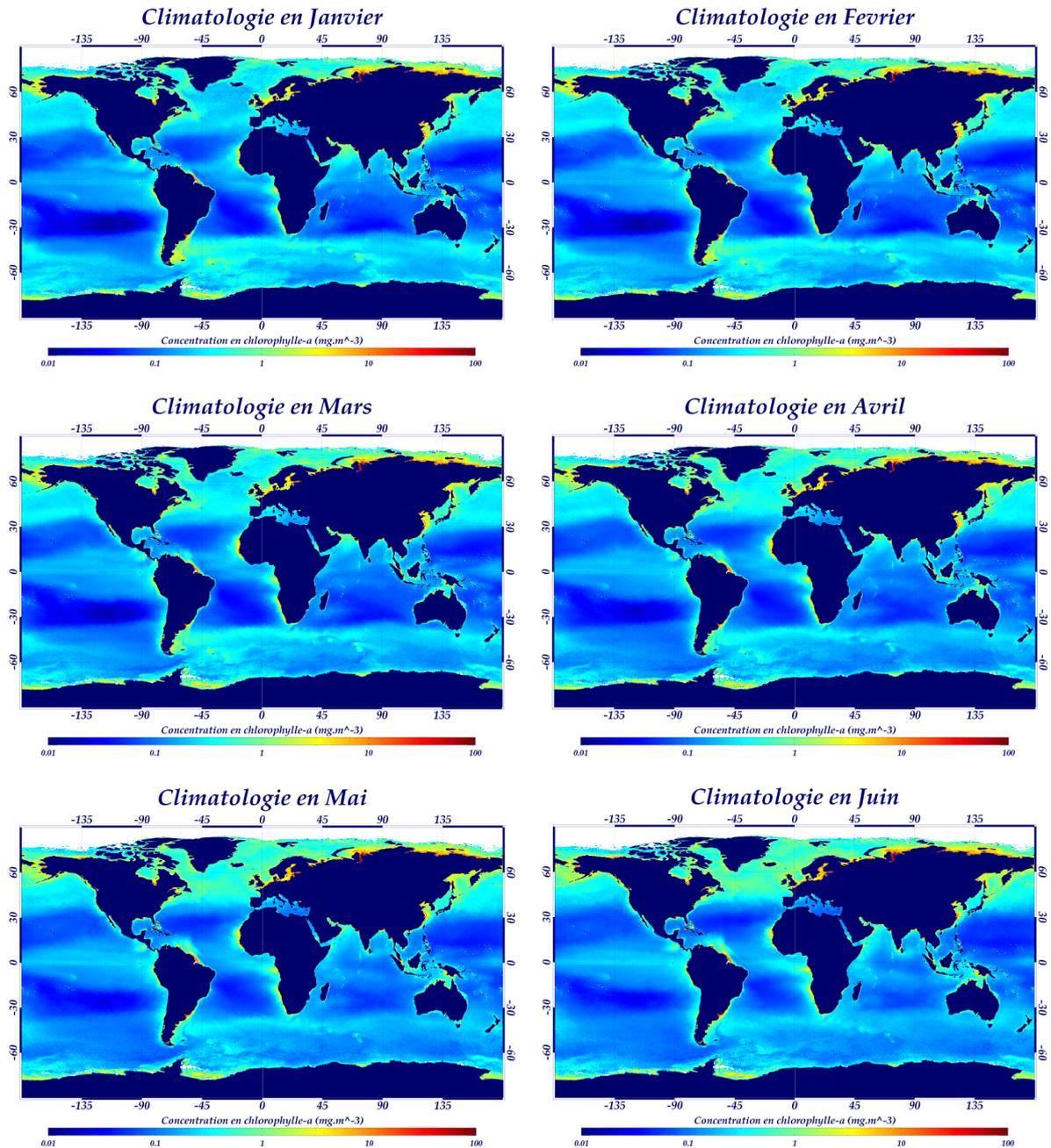
1.2 - MODIS/Aqua



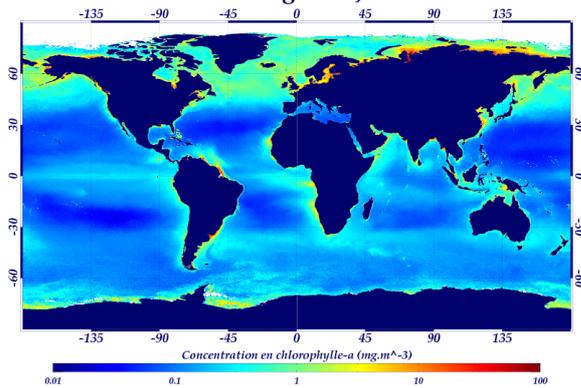


2 - Analyse objective

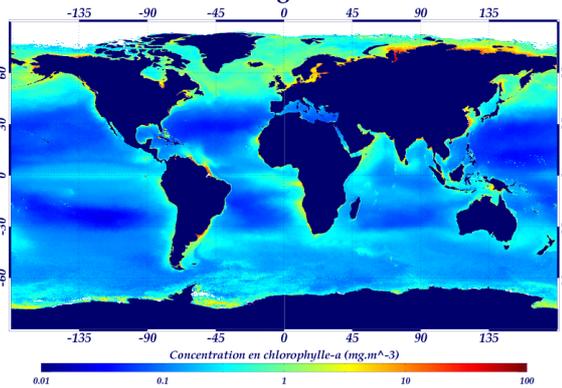
2.1 - Les cartes de climatologie mensuelle



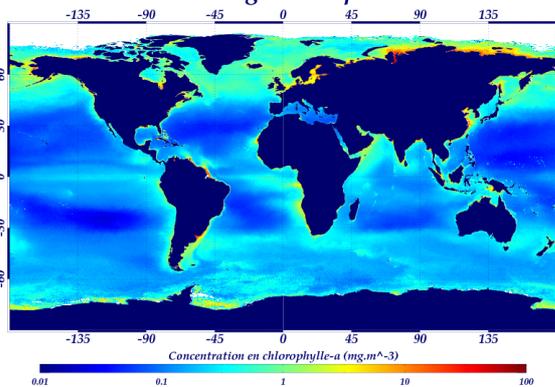
Climatologie en Juillet



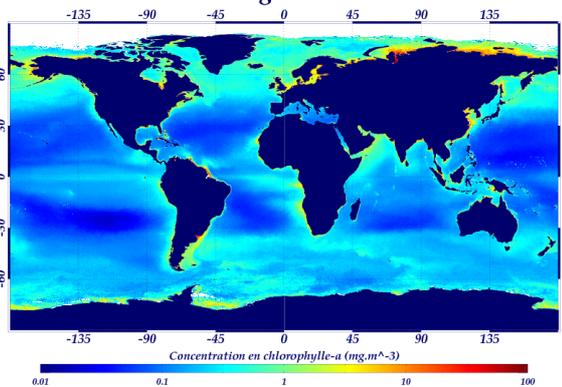
Climatologie en Aout



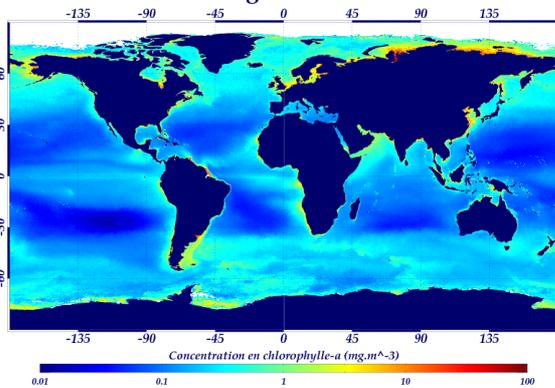
Climatologie en Septembre



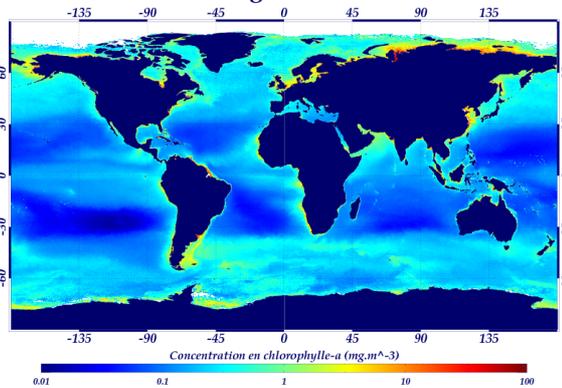
Climatologie en Octobre



Climatologie en Novembre

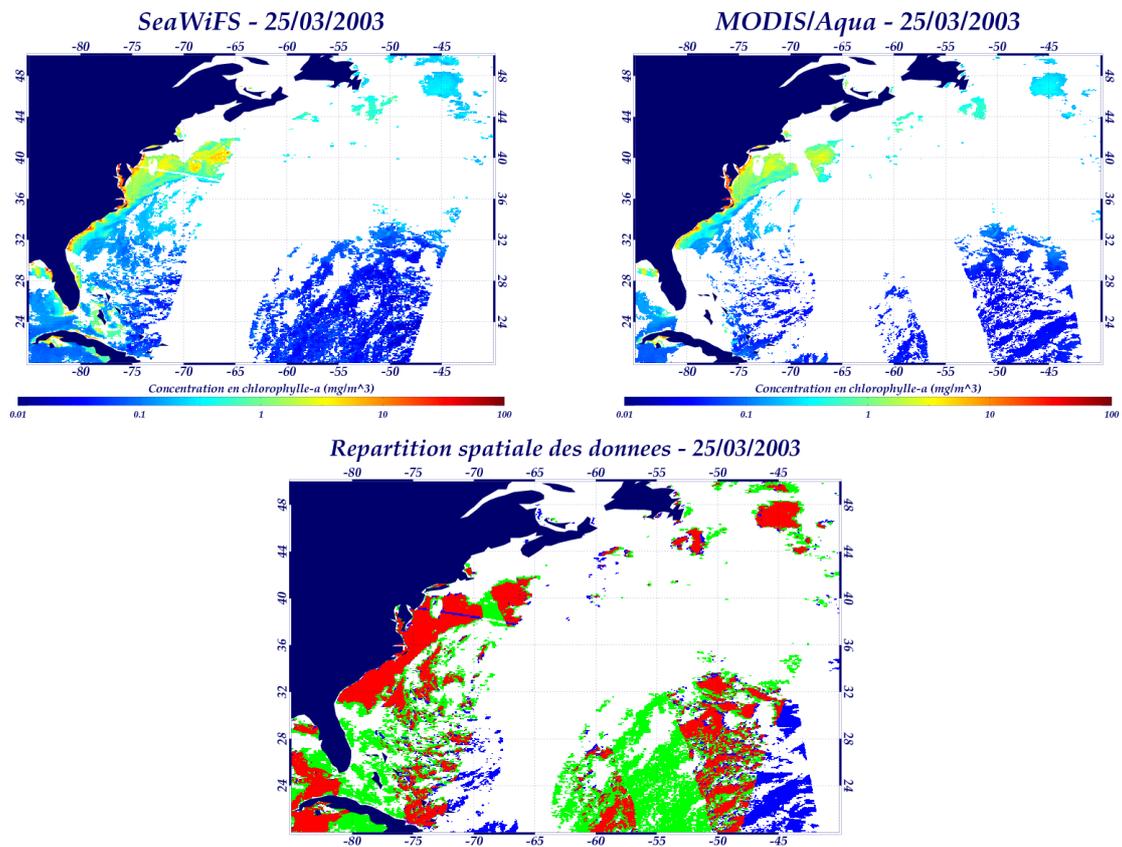


Climatologie en Decembre

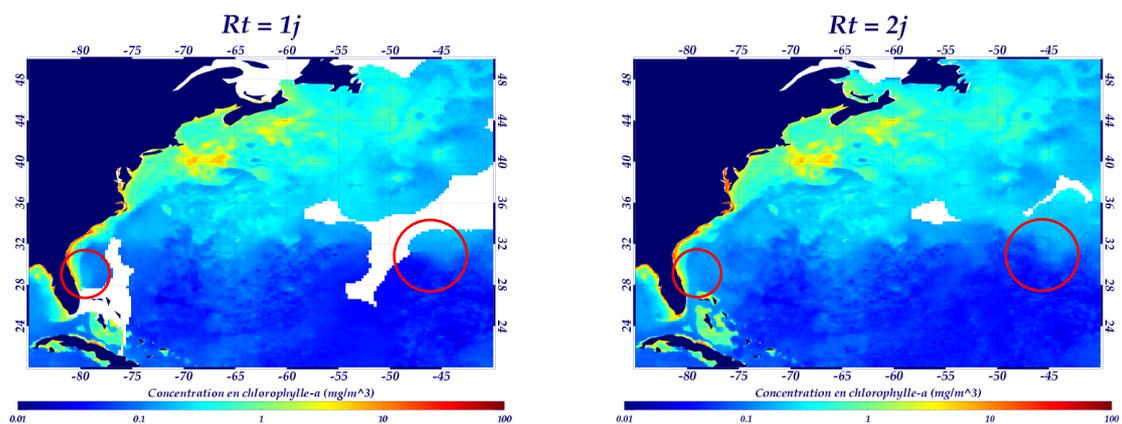


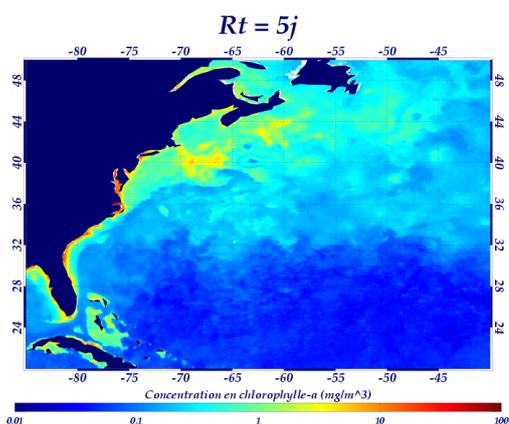
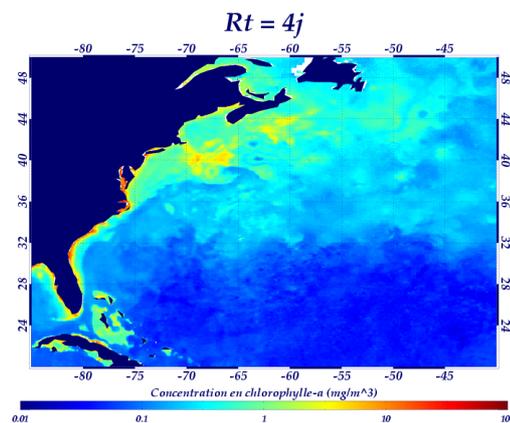
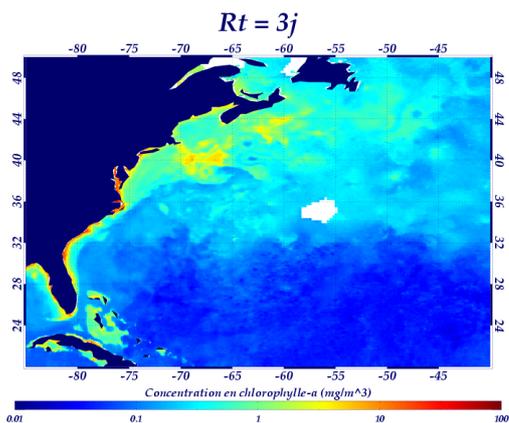
2.2 - Tests de sensibilité

2.2.1 - Cartes pour le 25 mars 2003

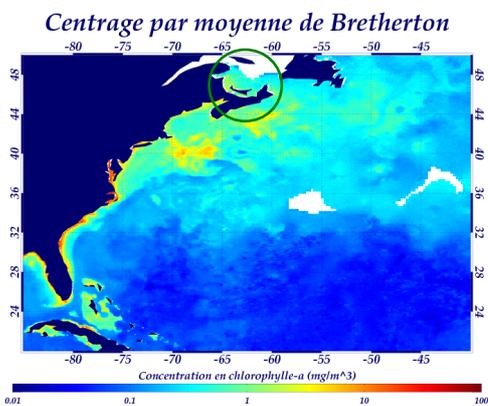
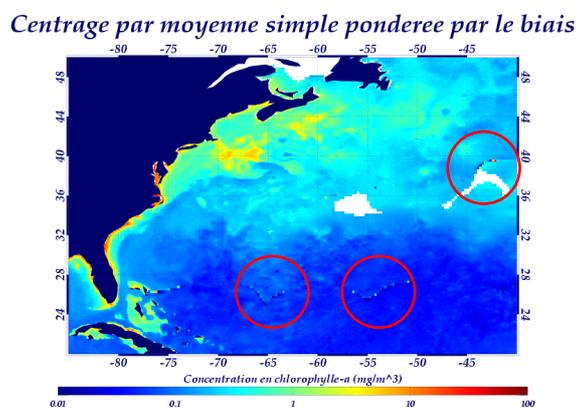
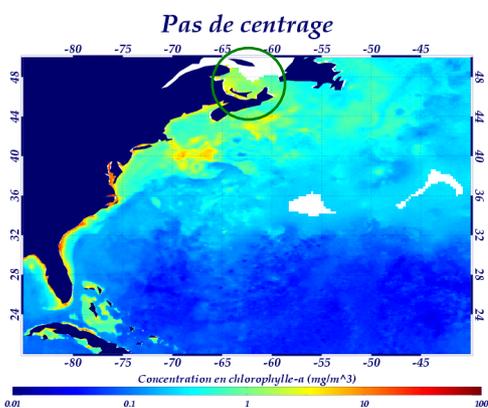


2.2.2 - Rayon de corrélation temporel

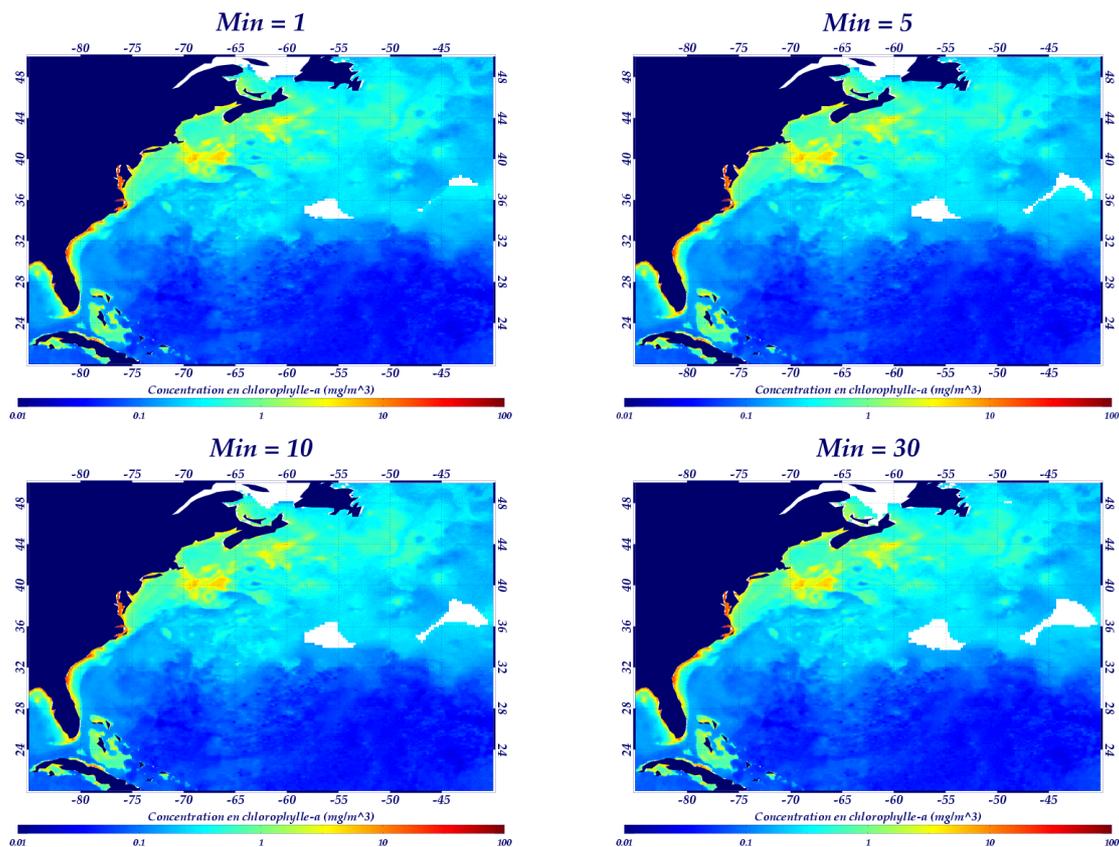




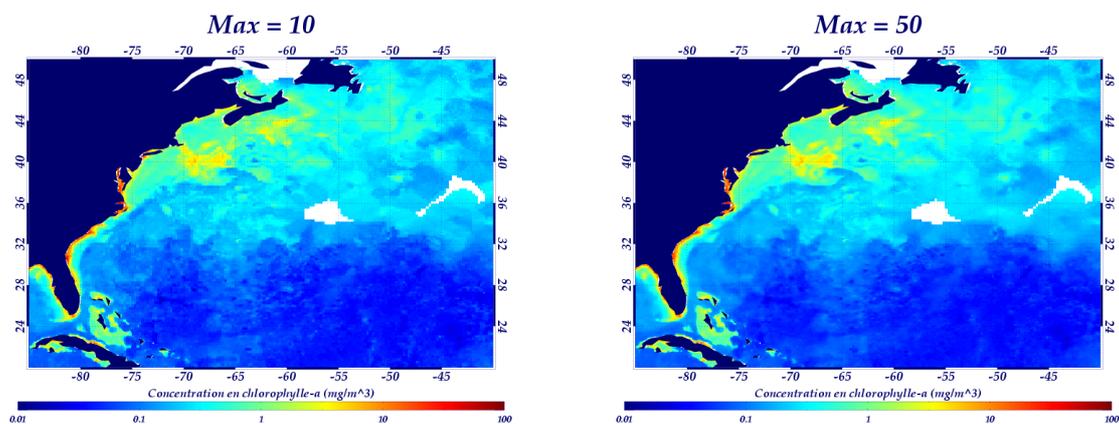
2.2.3 - Centrage des données

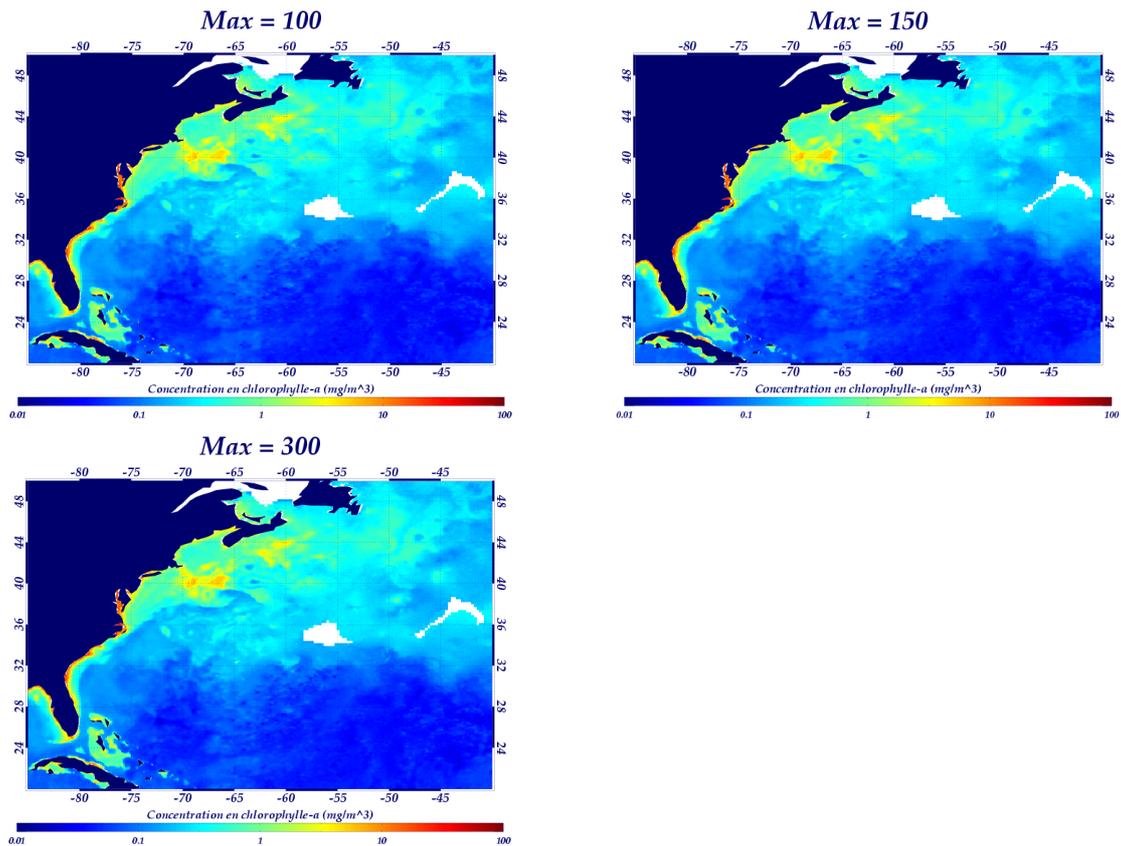


2.2.4 - Nombre minimum d'observations dans la bulle d'influence

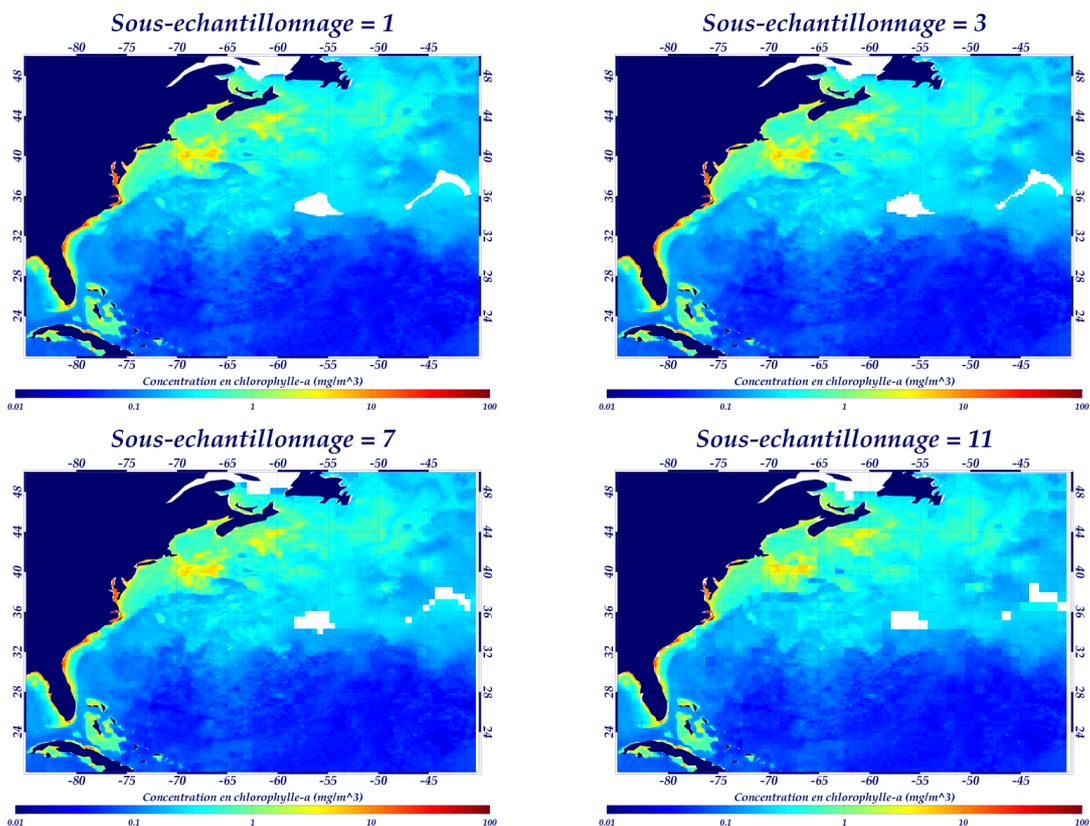


2.2.5 - Nombre maximum d'observations dans la bulle d'influence

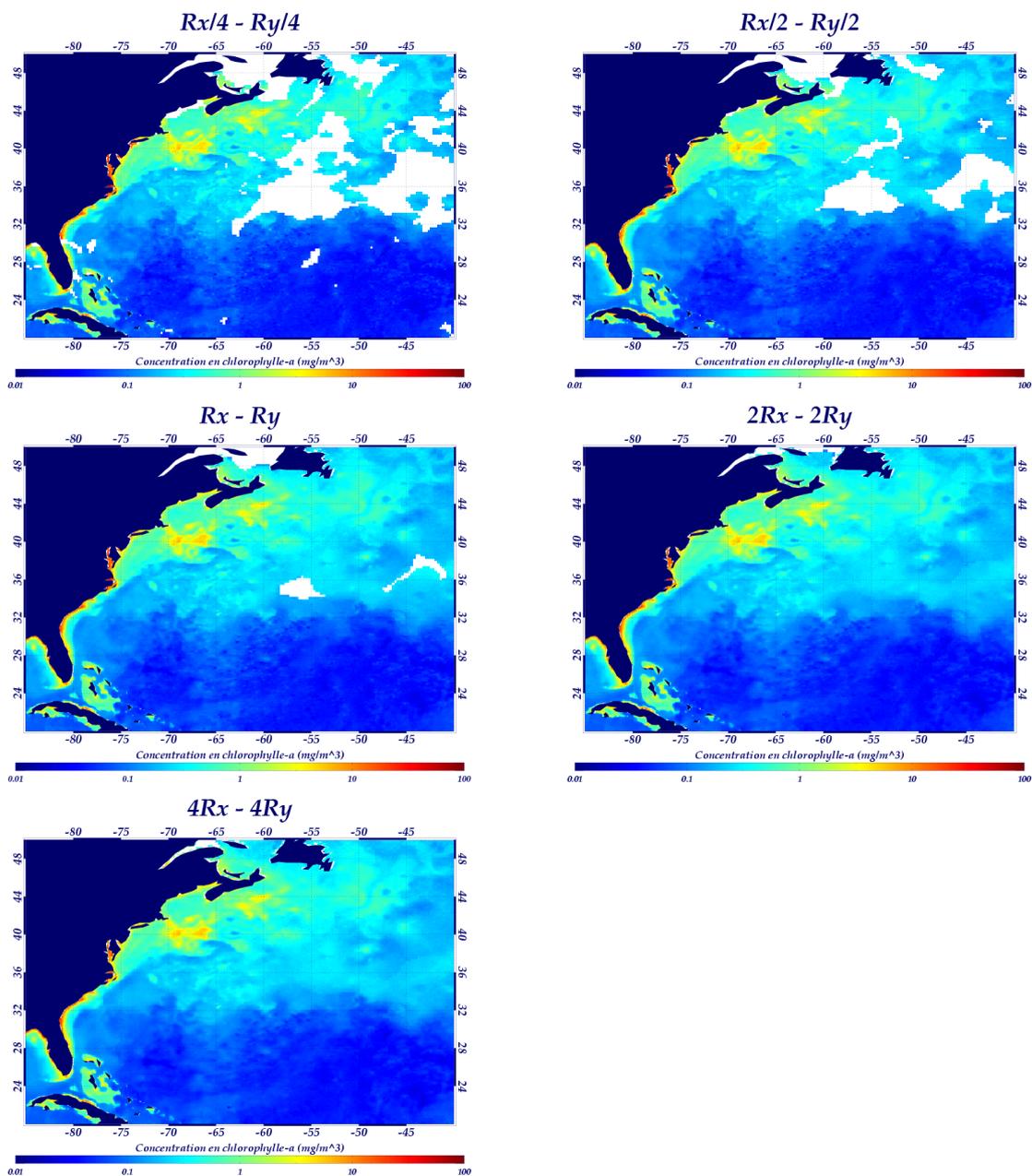




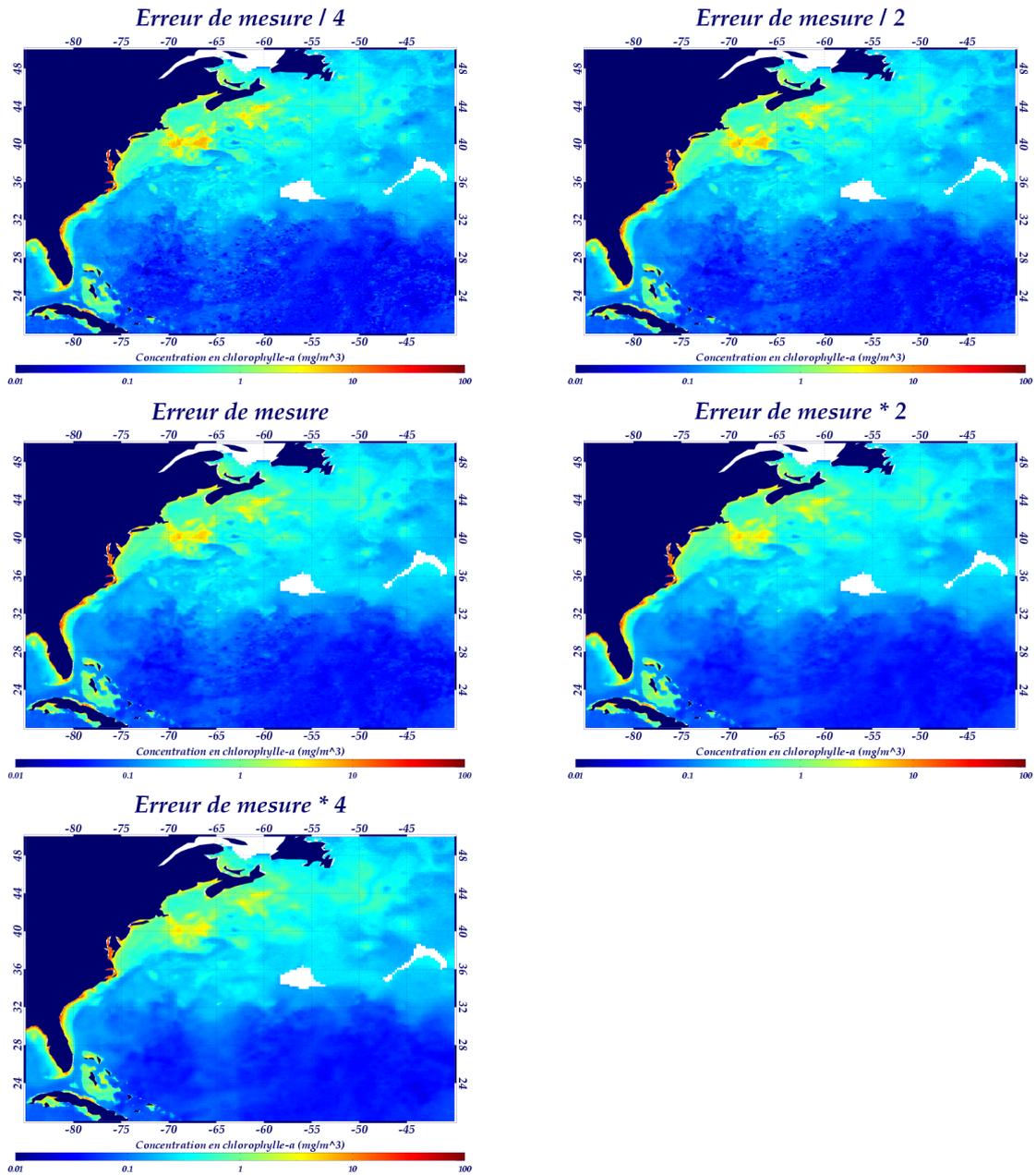
2.2.6 - Sous-échantillonnage des positions d'étude



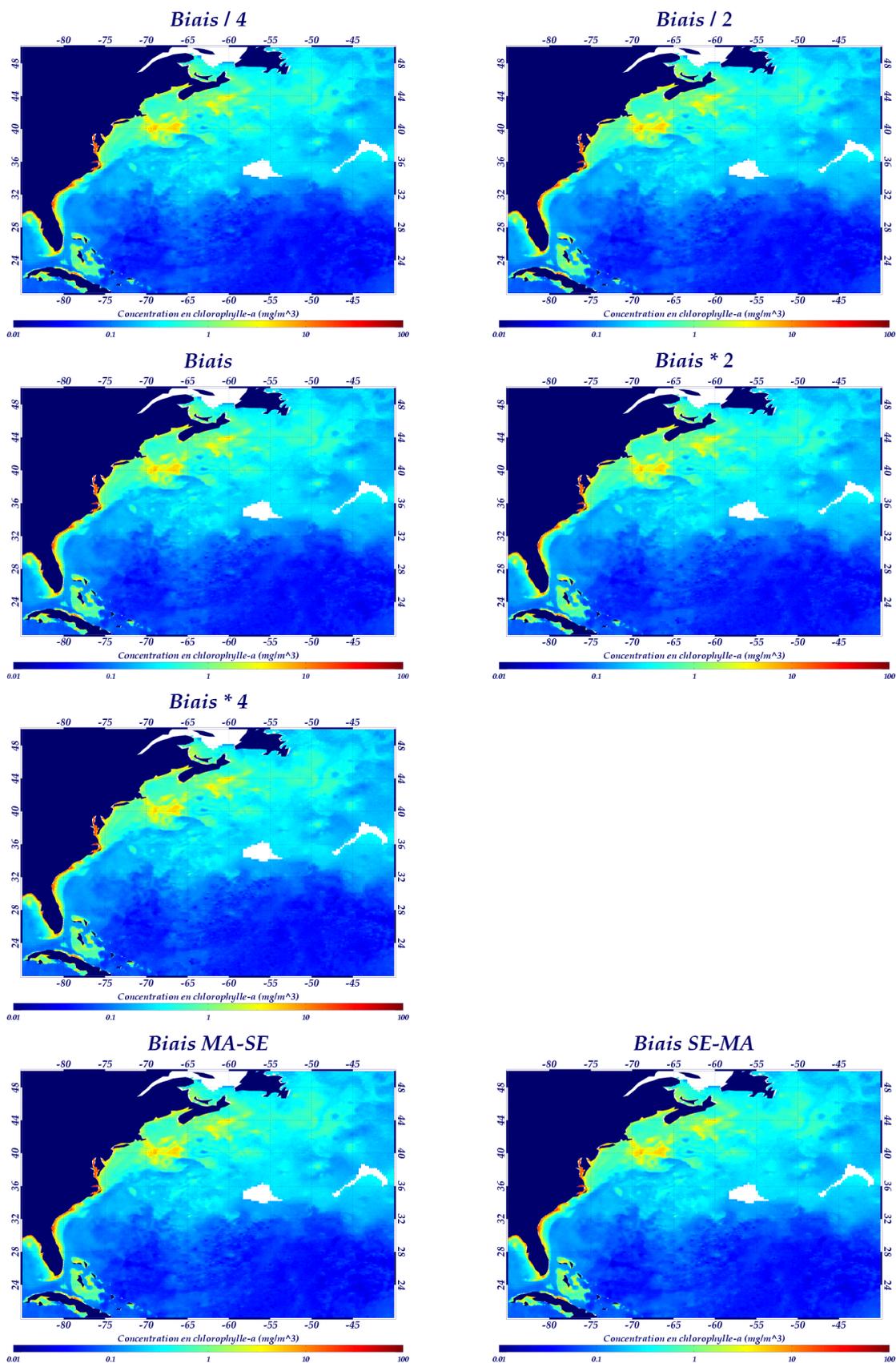
2.2.7 - Rayons de corrélation spatiaux



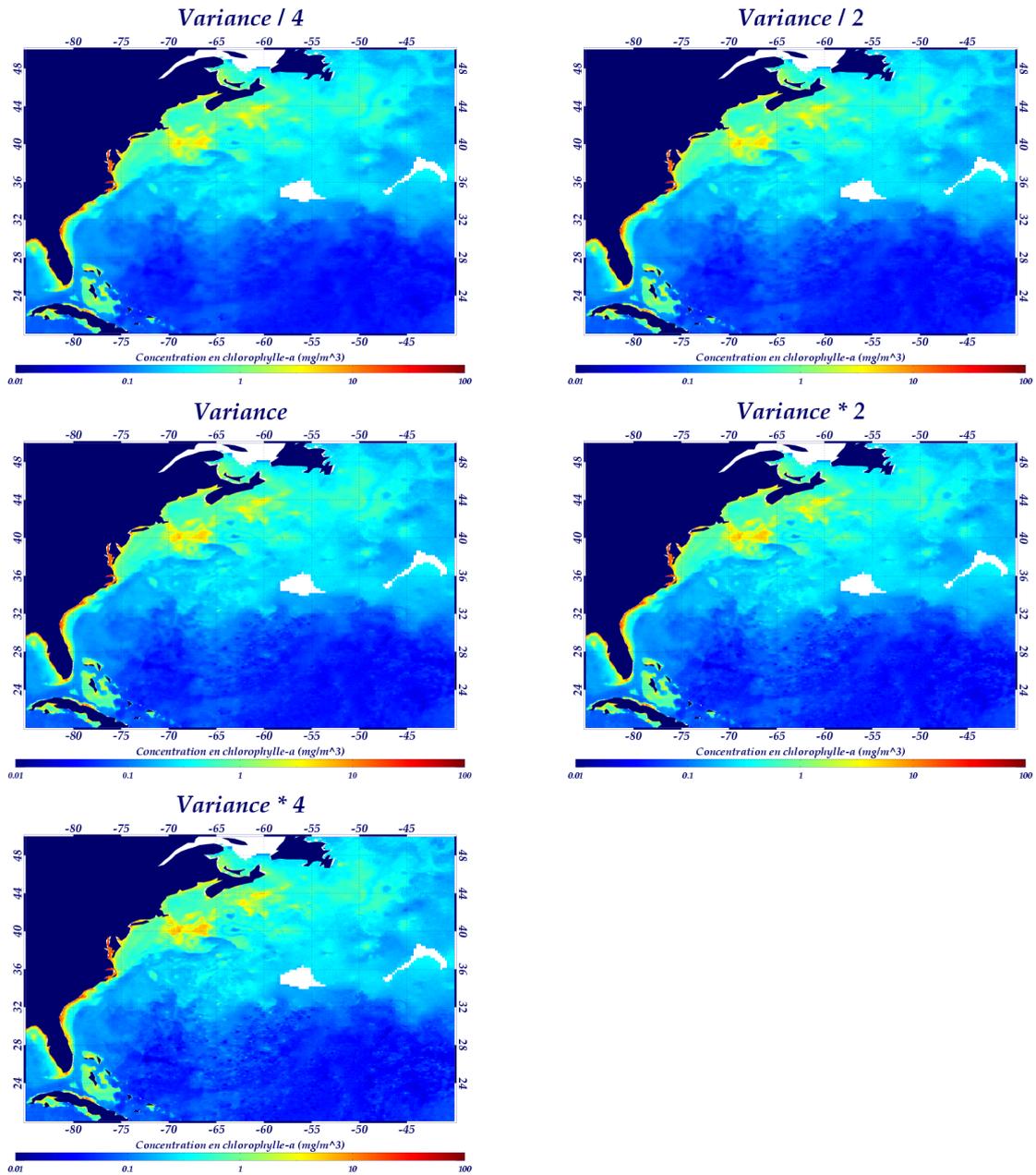
2.2.8 - Erreur de mesure



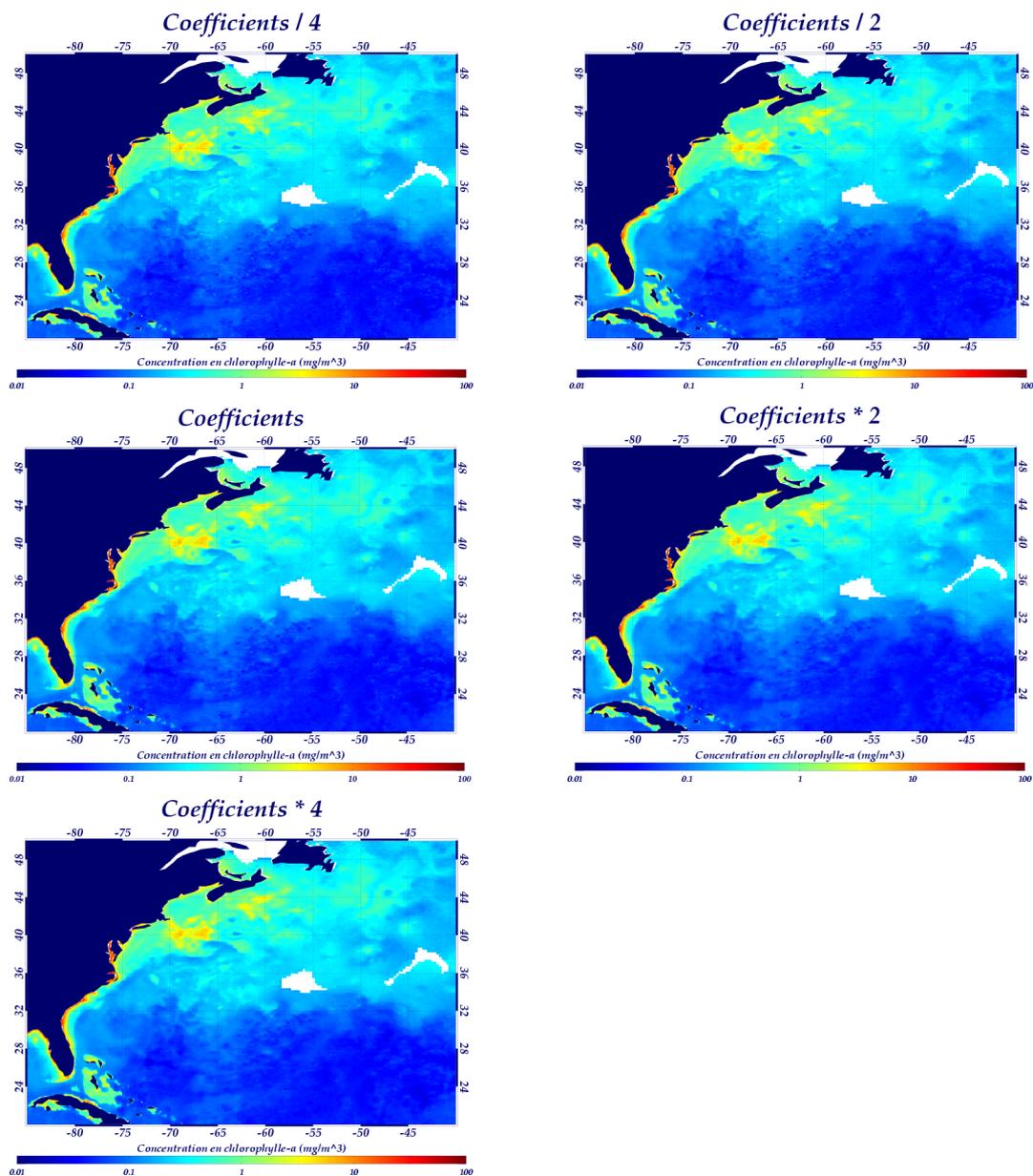
2.2.9 - Biais



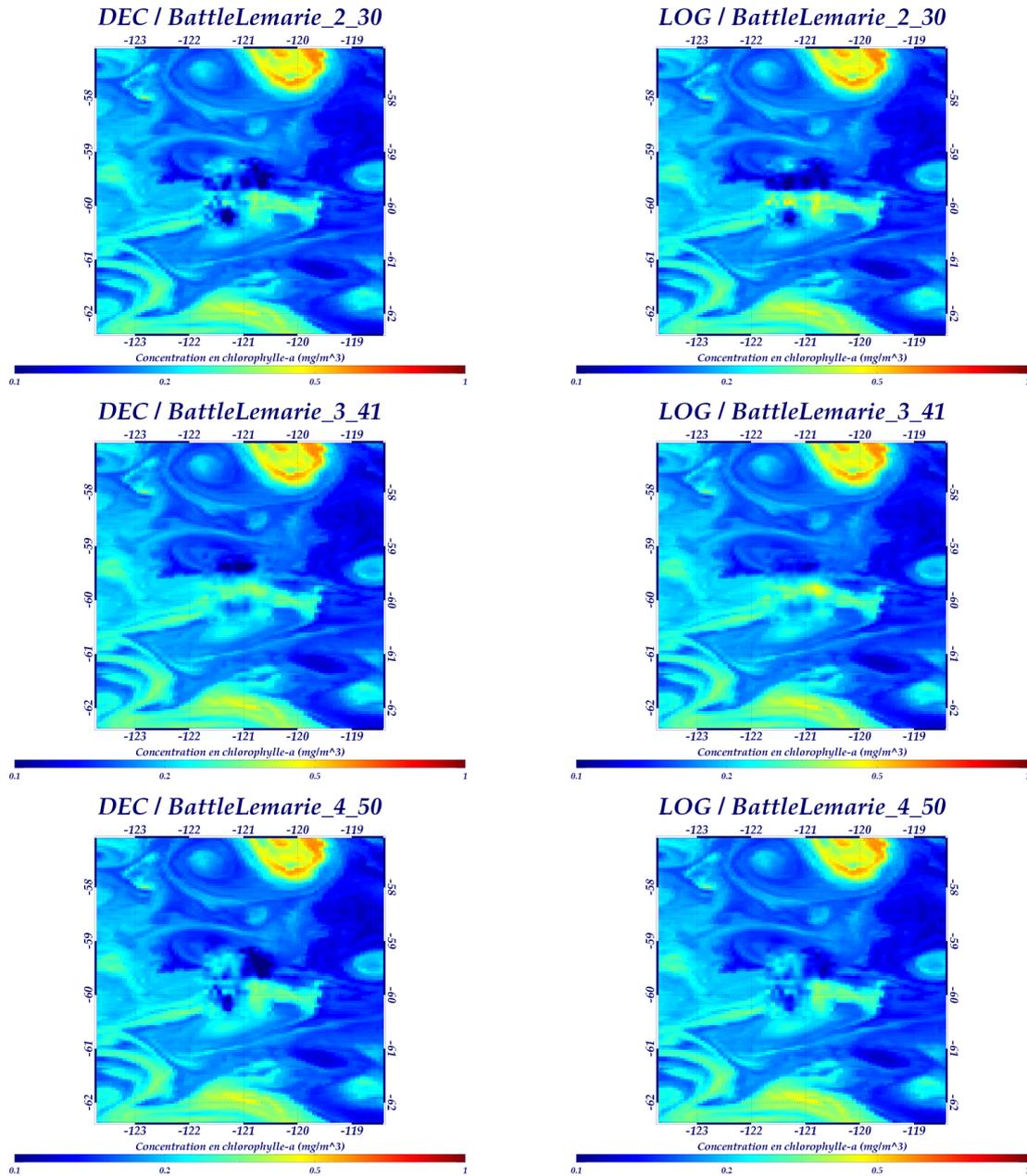
2.2.10 - Variance

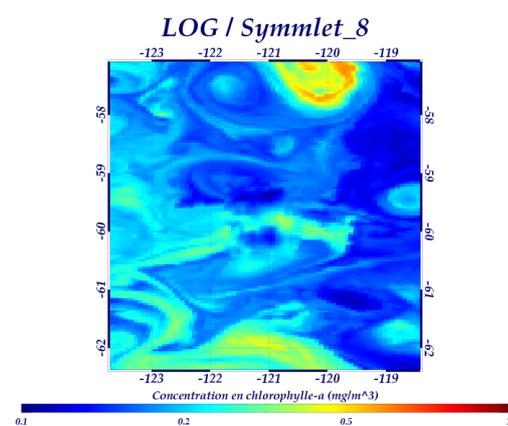
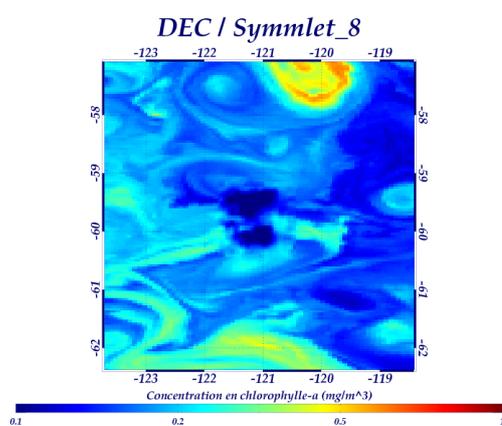
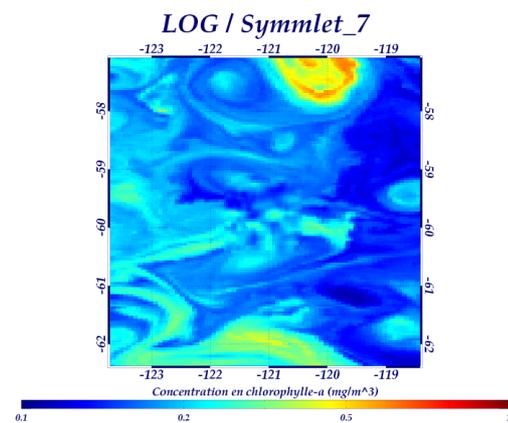
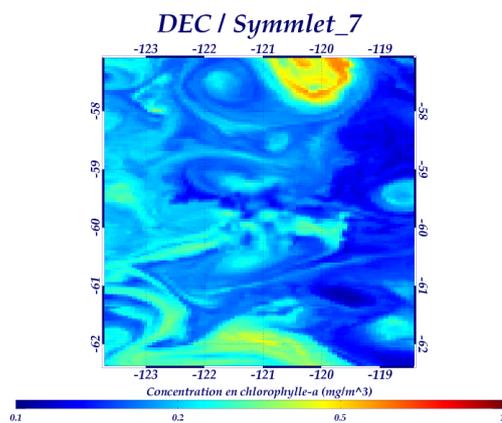
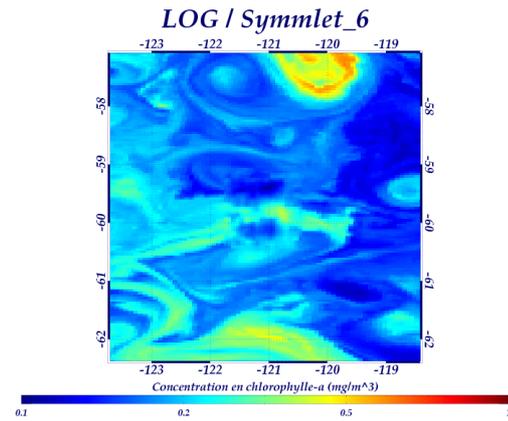
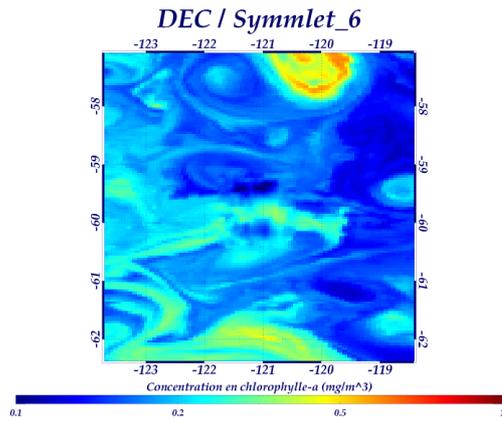


2.2.11 - Coefficient de la fonction de corrélation



3 - Transformée en ondelettes : tests





4 - Missions satellitaires altimétriques

4.1 - Intérêt de la mesure

L'altimétrie satellitaire est la mesure d'une hauteur par rapport à une référence qui permet, dans l'océan, de déduire entre autres les courants. Si l'océan était homogène et au repos, la surface de la mer coïnciderait avec le **géoïde**. Cependant, les forçages¹ entretiennent des écarts entre le niveau de la mer et le géoïde, que l'on appelle la **topographie dynamique**. La pente de celle-ci se traduit par des gradients horizontaux de pression auxquels l'océan (milieu en rotation) réagit par des courants (par exemple, géostrophiques). A l'inverse, si l'on connaît la latitude, la hauteur et la forme du niveau de la mer, on peut calculer la trajectoire et la vitesse des courants. La mesure altimétrique est utilisée pour étudier la circulation océanique globale et à mésoéchelle², les variations du niveau moyen de la mer pendant un événement tel qu'El Niño, la détection d'ondes de Rossby, le signal saisonnier du niveau de la mer, les marées océaniques, mais aussi dans d'autres domaines d'application comme par exemple, la topographie des calottes polaires.

La détermination précise des courants sur toute la colonne d'eau reste le but à atteindre pour les océanographes physiciens. Bien que les courants barotropes soient visibles par l'altimètre, la précision du calcul des courants profonds demeure insuffisante. Ce n'est, en revanche, pas le cas pour les courants instantanés qui sont beaucoup plus intenses.

Les données altimétriques sont également utilisées pour contraindre les modèles de circulation ou de marée puisqu'on dispose à la fois des données et des erreurs associées en chaque point de grille.

4.2 - Principe de la mesure

Un satellite altimétrique est un système de mesure composé d'au moins trois éléments : un altimètre bifréquence, un système d'orbitographie et un radiomètre micro-ondes. Le niveau de la mer est déterminé par la différence entre l'altitude du satellite au-dessus de l'ellipsoïde de référence et la distance du satellite à la surface de la mer, mesurée par le radar embarqué. La Figure A.1 résume le principe.

¹ Terme générique désignant les flux de chaleur et d'eau douce entre l'océan et l'atmosphère, la tension du vent, les anomalies de gravité, etc., qui entretiennent les courants océaniques.

² Echelle de la circulation océanique proche de l'échelle minimale à laquelle l'équilibre géostrophique est respecté, voisine de 200 km aux latitudes tempérées.

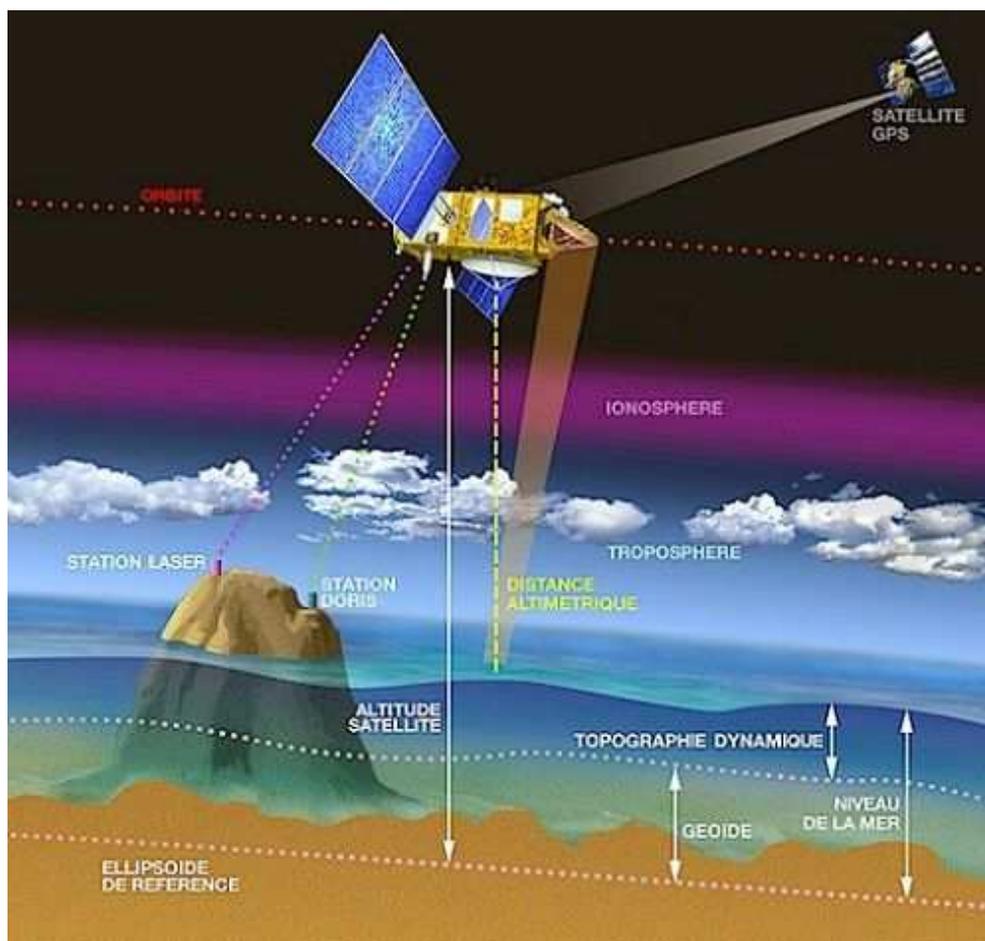


Figure A.1 – Principe de l'altimétrie.

La mesure altimétrique est dite « active », dans le sens où l'altimètre est un radar qui mesure la distance entre le satellite et la surface de la mer, à partir du trajet aller et retour de son signal. Quand cette hauteur est soustraite de la hauteur de l'orbite du satellite, la différence donne la hauteur relative de la surface de la mer par rapport au centre de masse de la Terre. L'allure de la surface est due à la répartition hétérogène des masses à l'intérieur de la Terre qui induisent des changements de gravité. Ces ondulations statiques du géoïde sont de l'ordre de $\pm 100\text{m}$. Des ondulations de moindre amplitude (quelques dizaines de cm) sont liées aux courants océaniques qui décrivent la topographie océanique, c'est-à-dire l'écart de la hauteur de mer par rapport au géoïde.

Le but est d'obtenir la meilleure précision possible alors que le satellite est en orbite vers 800 ou 1300 km d'altitude, selon les cas (1300 km pour T/P, altitude variable pour ERS-1/2). Pour que la hauteur de mesure soit précise, il faut d'une part que la direction de visée du radar soit connue – en pratique, on contrôle l'attitude du satellite à 0.1° près – d'autre



part, il faut corriger les perturbations de la ionosphère et de l'atmosphère sur la propagation des impulsions radar. La diffusion du signal radar par les électrons de la ionosphère allonge la distance apparente de 2 à 30 cm ; comme cette perturbation dépend de la fréquence du signal radar, il est possible de l'estimer à partir des mesures de hauteur faites à deux fréquences différentes. La vapeur d'eau de l'atmosphère retarde le signal radar : le satellite embarque aussi un radiomètre micro-ondes pour mesurer le contenu intégré en vapeur d'eau de l'atmosphère. La correction vaut de 10 à 40 cm et peut être connue à 2 cm près. L'oxygène et l'azote de l'air retardent aussi la propagation du signal radar : cette correction est égale à 2.3 m en moyenne et peut être déterminée avec une très bonne précision à partir de la pression atmosphérique.

Après corrections, le bruit de mesure est de l'ordre de 3 à 4 cm pour les altimètres des satellites ERS-1 et 2. Dans le cas de Topex-Poseidon, tout a été optimisé pour obtenir un bruit de mesure de 2 à 3 cm. Par ailleurs, toutes les composantes du système de mesure sont surveillées de façon à assurer une stabilité à long terme des données. La trace du radar à la surface de l'océan – le « footprint » – a un diamètre d'environ 7 km. Ceci permet de moyenniser l'effet des vagues. En fait, comme le signal radar provenant de la crête des vagues arrive un peu plus tôt que celui qui provient du creux de celles-ci, on peut aussi déterminer la hauteur caractéristique des vagues. Cependant, comme les crêtes des vagues diffusent plus le signal radar que les creux, la mesure du niveau de la mer est affectée d'un « biais d'état de mer » (ou biais électromagnétique), qui dépend de la hauteur des vagues et de la vitesse du vent. Comme la physique des états de mer est complexe, ce biais est estimé de façon empirique et constitue actuellement la principale incertitude des mesures altimétriques.

Le radar émet des impulsions de fréquence 13 GHz à un rythme de 1000 à 4000 Hz. Leurs échos sont moyennés par paquets, puis analysés pour en déduire la mesure de distance à la fréquence de 1 Hz. Comme le satellite se déplace le long de son orbite à une vitesse de 7 km/s, un altimètre fournit donc la hauteur de mer tous les 7 km le long de sa trace. Pour T/P, l'orbite se répète tous les 10 jours environ tandis que pour ERS, la répétitivité de l'orbite est de 35 jours. À l'équateur, la distance entre les traces est de ≈ 340 km pour le satellite T/P et de ≈ 80 km pour ERS. Cette distance inter-trace décroît avec la latitude, plus l'inclinaison du plan d'orbite est importante.

4.3 - La mission TOPEX-POSEIDON

4.3.1 - Introduction

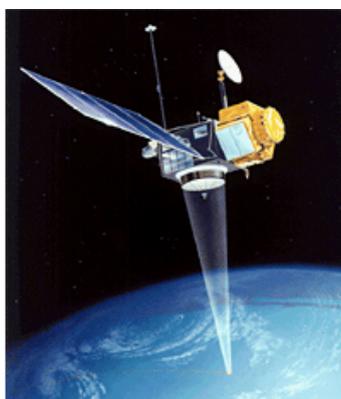


Figure A.2– Le satellite TOPEX-POSEIDON.

Le satellite TOPEX/POSEIDON (indiqué « T/P » par la suite) a été lancé le 10 août 1992 avec pour mission *observer et comprendre la circulation océanique*. Issu d'un partenariat entre la NASA (National Aeronautics and Space Administration) et le CNES, il a à son bord deux radars altimètres et différents systèmes de localisation précise, dont le système DORIS. L'énergie est fournie par un panneau solaire de 25 m².

Tous les 10 jours, le satellite fournit la topographie mondiale des océans mesurée avec une très bonne précision : la précision moyenne instantanée de l'estimation locale du niveau de l'océan est inférieure à 5 cm, et la précision moyenne sur 1 mois est inférieure à 2 cm.

Depuis le 15 septembre 2002, Topex/Poséidon est sur une nouvelle orbite, à mi-chemin entre ses anciennes traces (maintenant celles de Jason-1). La mission s'est terminée en octobre 2005, suite à un incident sur une des roues d'inertie du satellite.

4.3.2 - Le satellite

Le satellite T/P est une adaptation de la plate-forme Multimission Modular Spacecraft (MMS, Fairchild Space), qui a été utilisée avec succès pour les missions Solar Maximum Mission, Landsat 4 et Landsat 5. Les caractéristiques du MMS ont ainsi été modifiées pour emporter les 2400 kg de T/P, avec les six instruments scientifiques (quatre opérationnels et deux expérimentaux) décrits dans le Tableau A.1.

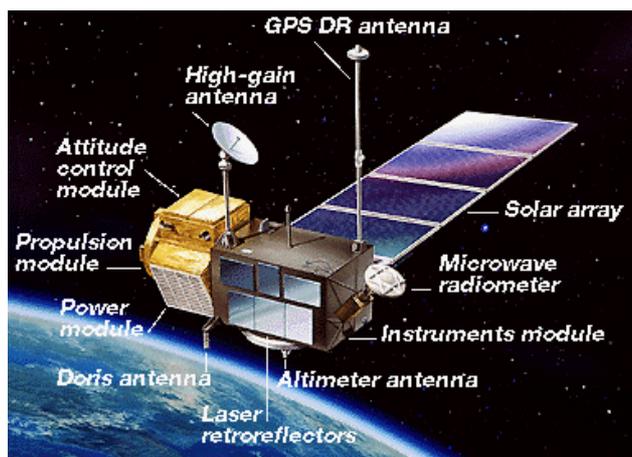


Figure A.3 – Composition du satellite T/P - De nombreuses améliorations ont été apportées sur T/P par rapport aux précédents systèmes (SeaSat, GeoSat), pour l'étude des océans à l'échelle des bassins. Ceci comprend, outre un satellite spécialement conçu, une série de capteurs, des systèmes de localisation, le choix de l'orbite, ainsi que le développement d'un modèle de gravité optimal pour la détermination précise de l'orbite, un segment sol dédié pour la mission.

Tableau A.1 – Liste des capteurs à bord de TOPEX-POSEIDON

<i>Nom</i>	<i>Type</i>	<i>Origine</i>	<i>Produits</i>	<i>Précision</i>
NRA ¹ (TOPEX)	Altimètre radar bifréquence	NASA opérationnel	Distance altimétrique Hauteur de vagues Module de la vitesse du vent Coefficient de rétrodiffusion Correction ionosphérique	2.4 cm sur l'altitude
SSALT ² (POSEIDON)	Altimètre à état solide monofréquence	CNES expérimental	Distance altimétrique Hauteur de vagues Module de la vitesse du vent Coefficient de rétrodiffusion	2.5 cm sur l'altitude
TMR ³	Radiomètre micro-ondes TOPEX	NASA opérationnel	Température de brillance Vapeur d'eau atmosphérique Eau liquide atmosphérique	0.2g / cm ² (densité de vapeur d'eau)
LRA ⁴	Réflecteurs laser	NASA opérationnel	Ephémérides ⁵ d'orbite précises ⁶	2 cm
DORIS ⁷	Récepteur d'orbitographie	CNES opérationnel	Ephémérides d'orbite précises ⁶ Localisation précise de balises au sol Correction ionosphérique	5 cm sur l'orbite
GPRDR ⁸	Récepteur GPS	NASA expérimental	Ephémérides d'orbite précises ⁶	10 cm sur l'orbite

¹ NASA Radar Altimeter

² single frequency Solid State radar ALTimeter

³ TOPEX Microwave Radiometer

⁴ Laser Retroreflector Array

⁵ Sorties des modèles qui permettent de connaître la trajectoire idéale d'un satellite.

⁶ Les mesures altimétriques sont repérées par rapport aux coordonnées géodésiques grâce au système de détermination d'orbite précise. Les deux équipes d'orbitographie précise utilisent la combinaison des données DORIS et LRA, l'un contrôlé par le CNES, l'autre par la NASA.

⁷ Détermination d'Orbite et Radiopositionnement Intégrés par Satellite

⁸ Global Positioning System Demonstration Receiver

4.3.3 - L'orbite

L'orbite de T/P est optimisée pour l'étude de la variabilité océanique à grande échelle, tout en permettant le survol de 90% des mers en 10 jours.

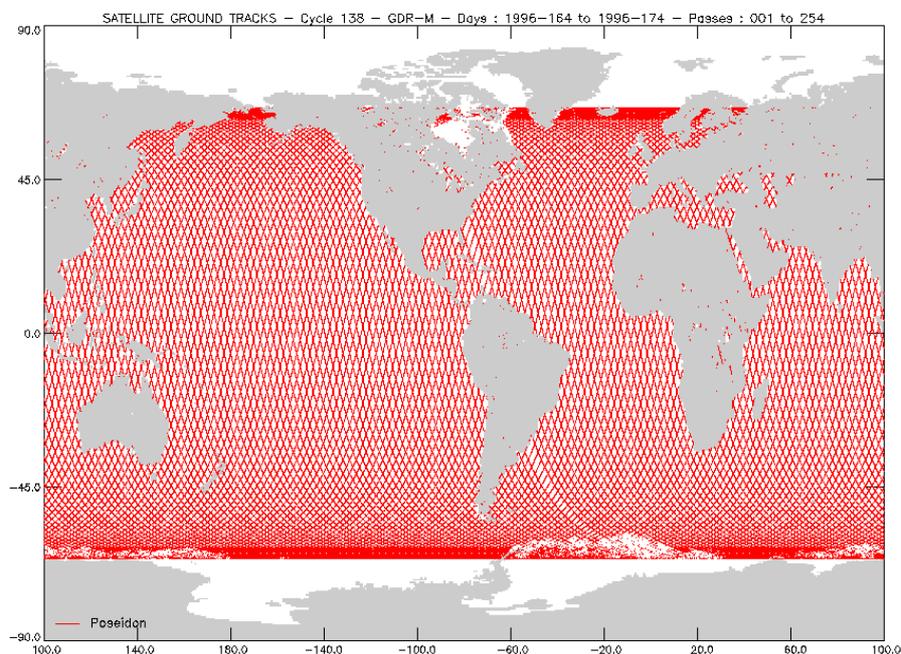


Figure A.4 – Traces au sol de TOPEX-POSEIDON.

Tableau A.2 – Caractéristiques de l'orbite de TOPEX-POSEIDON.

Caractéristiques principales	
Demi-grand axe	7 714.4278 km
Excentricité	0.000095
Inclinaison (non héliosynchrone ¹)	66.039°
Données auxiliaires	
Altitude de référence (équatoriale)	1336 km
Période nodale (durée d'une demi-révolution ou passe)	6 745.72 secondes (112 min 42 s ou 1 h 52min)
Répétitivité (durée du cycle)	9.9156 jours
Nombre de passes par cycle	254
Intertrace à l'équateur	315 km
Longitude à l'équateur de la passe 1	99.9242°
Vitesse orbitale	7.2 km/s
Vitesse de balayage au sol	5.8 km/s

L'altitude élevée du satellite (environ 1336 km) permet de minimiser les interactions avec l'atmosphère et le champ de gravité de la Terre, rendant ainsi la détermination de l'orbite plus facile et plus précise. L'inclinaison de l'orbite, qui amène le satellite jusqu'à 66° Nord et Sud, permet le survol de la plupart des mers non gelées. La période de revisite de l'orbite est d'un peu moins de 10 jours (9.9156 jours, soit à peu près 10 jours moins 2

¹ Quasi-polaire.

heures), c'est-à-dire que le satellite repasse au-dessus d'un même point (à 1 km près) tous les 10 jours, complétant ainsi un cycle. Ce choix est un compromis entre la résolution spatiale et la résolution temporelle, qui permet une étude de la variabilité océanique à grande échelle.

4.3.4 - Le traitement des données

Des données brutes aux données utilisées par les océanographes, de nombreuses étapes sont nécessaires. Les chaînes de traitement effectuent ce travail essentiel pour l'exploitation d'un satellite. SSALTO (Segment Sol multi-mission d'ALTimétrie, d'Orbitographie et de localisation précise) assure ces traitements pour T/P.

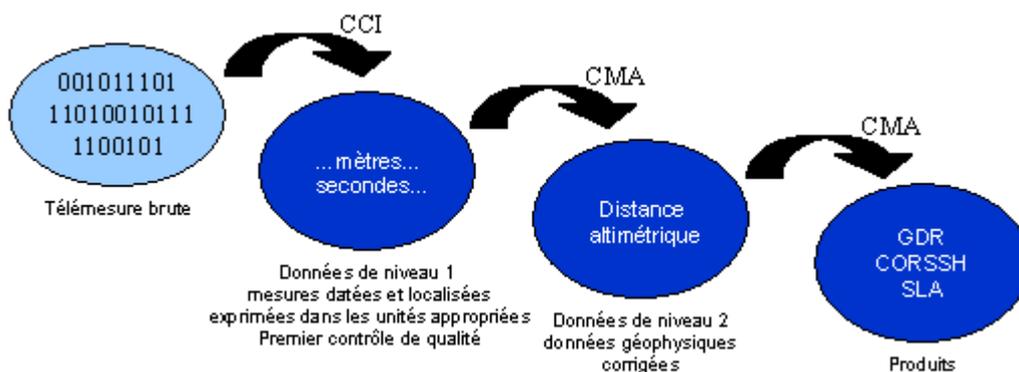


Figure A.5 - Chaîne de traitement des données.

De la télémessure brute aux données de niveau 1

La télémessure brute en provenance des instruments est reçue par les stations sol du CNES et de la NASA, qui la transmettent aux centres de contrôle et de traitement, localisés au JPL (Jet Propulsion Laboratory) et au CNES. La donnée brute est alors datée, localisée, exprimée en unité appropriée et enfin contrôlée pour en vérifier sa qualité. Ces opérations sont réalisées par le Centre de Contrôle des Instruments (CCI). Le produit final de cette première phase de traitement est appelé « données de niveau 1 ».

Des données de niveau 1 aux données géophysiques (niveau 2)

Les données de niveau 1 sont ensuite traitées pour les corriger des erreurs instrumentales, des erreurs de propagation à travers l'atmosphère ainsi que des perturbations dues à la réflexion sur la surface et des corrections géophysiques (marées terrestres, océaniques, polaires...).

La validation des données

Les données géophysiques sont ensuite validées par le Centre Multi-mission Altimétrique (CMA), afin d'en garantir la qualité. Cette validation permet le contrôle qualité précis des données ainsi que la surveillance d'éventuelles dérives des instruments. Cette étape est partie intégrante de la chaîne de traitement et donne le feu vert à la génération et à la diffusion des produits scientifiques.

4.3.5 - Les produits TOPEX-POSEIDON

Les produits T/P, distribués par AVISO¹, sont les suivants :

- 📄 **SLA (Sea Level Anomalies)** : hauteurs de mer résiduelles (ou anomalies de hauteur de mer) le long de la trace, calculées à partir d'une moyenne sur 3 ans,
- 📄 **MSLA (Maps of Sea Level Anomalies)** : hauteurs de mer résiduelles (ou anomalies de hauteur de mer), grillées (1/3°x1/3°),
- 📄 **CorSSH (Corrected Sea Surface Heights)** : hauteurs de mer corrigées le long de la trace,
- 📄 **Données vents-vagues** : données le long de la trace,
- 📄 **GDR-M (Geophysical Data Records-Merged)** : enregistrements géophysiques des données altimétriques le long de la trace de la mission T/P, homogénéisées entre les altimètres TOPEX et POSEIDON,
- 📄 **IGDR (Interim Geophysical Data Records)** : produits temps quasi-réel 48 heures (Ephémérides d'orbite ELFE DORIS, IGDR TOPEX et IGDR POSEIDON AVISO, comprenant les orbites précises CNES DORIS à 48 heures) le long de la trace.

4.4 - Les missions ERS-1/2

Les satellites ERS (European Remote Sensing) ont pour mission d'observer la Terre, en particulier l'atmosphère et les océans en utilisant des techniques radars. Ces satellites de l'ESA (European Space Agency) ont à leur bord plusieurs instruments, dont un radar altimètre.

¹ AVISO distribue les données des missions spatiales d'altimétrie (TOPEX-POSEIDON, JASON-1, ERS-1 et 2, Envisat) ainsi que les produits DORIS d'orbitographie et de localisation précises.



Figure A.6 - ERS-1.

ERS-1, lancé le 17 juillet 1991, a été mis en sommeil en juin 1996 et définitivement arrêté en mars 2000.

Pour répondre à ses missions, il a été placé sur différentes orbites répétitives:

- ☒ une orbite à 785 km d'altitude, répétitive à 3 jours pour l'observation des calottes polaires,
- ☒ une orbite répétitive à 35 jours pour les applications multidisciplinaires (dont l'observation des océans),
- ☒ une orbite répétitive à 168 jours pour les applications géodésiques ; ce type de cycle permet la mesure du niveau moyen de la mer et du géoïde océanique en raison de la très grande densité des traces de l'altimètre.

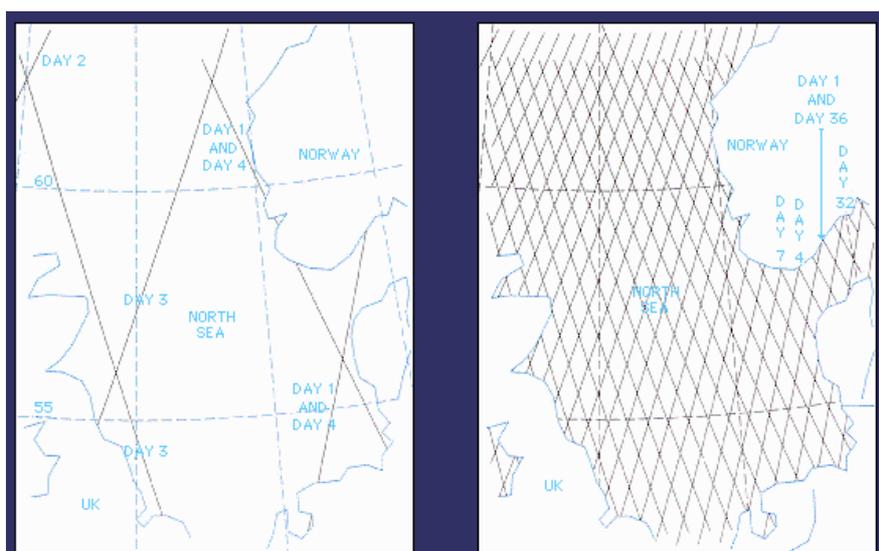


Figure A.7 – Comparaison de la densité des traces du radar altimètre en Mer du Nord dans les cycles de 3 jours et de 35 jours.

ERS-2, lancé le 21 avril 1995, est le successeur d'ERS-1 et a été utilisé en tandem avec celui-ci d'août 1995 à juin 1996 : mis sur des orbites identiques à 35 jours, les satellites se sont succédés à un jour d'intervalle.

4.5 - La mission Jason-1

4.5.1 - Introduction

Premier satellite d'une série prévue pour durer plusieurs décennies, Jason-1 est le successeur de Topex/Poséidon, dont il reprend les caractéristiques majeures (orbite, instruments, précision des mesures,...), toujours dans le cadre d'une coopération CNES-NASA. Contrôle du satellite et traitement des données sont effectués par un nouveau segment sol. Jason-1 est un mini-satellite, utilisant une plate-forme multimissions (Proteus). Le succès de Topex/Poséidon a en outre validé le concept, et a donc permis de continuer dans cette voie. Ainsi, tous les instruments embarqués à bord de Jason-1 dérivent de ceux de Topex/Poséidon, en particulier l'altimètre.

Jason-1 a été lancé le 7 décembre 2001, à 1300 km de la surface, sur une orbite répétitive à 10 jours, identique à celle de Topex/Poséidon.

4.5.2 - Le satellite

Cinq instruments sont embarqués sur Jason-1.

L'altimètre Poséidon-2 est l'instrument principal de la mission. Il est issu de l'altimètre expérimental Poséidon-1 de Topex/Poséidon. Poséidon-2 a bénéficié de plusieurs améliorations par rapport à Poséidon-1. En particulier, une deuxième fréquence a été rajoutée (bande C), pour permettre le calcul de la correction ionosphérique, l'échantillonnage a été amélioré, et, dans la mesure du possible, les circuits analogiques ont été remplacés par des composants numériques. Petit, léger et ne nécessitant pas une grande quantité d'énergie, il est aussi d'une grande fiabilité. C'est un radar qui émet des ondes à deux fréquences différentes (13.6 et 5.3 GHz, pour permettre la détermination du contenu en électrons de l'atmosphère) et analyse le signal réfléchi par la surface. Le temps de trajet aller-retour de l'onde est estimé très précisément afin de calculer la distance satellite / surface, moyennant quelques corrections. (Contribution CNES)

Le radiomètre JMR (Jason-1 Microwave Radiometer) collecte le rayonnement émis par la surface à trois fréquences différentes (18, 21, et 37 GHz). Le but est de déterminer le contenu en vapeur d'eau et en eau liquide de l'atmosphère, en combinant les mesures prises à chacune de ces fréquences. Une fois connu ce contenu en eau, on en déduit la correction

à appliquer à la mesure altimétrique, l'onde radar de l'altimètre étant ralentie par cette eau atmosphérique. (Contribution NASA)

Les systèmes de localisation à bord de Jason-1 se complètent pour permettre une mesure à 2 cm près de la position du satellite sur son orbite (distance radiale Terre-satellite). Le LRA (Laser Retroreflector Array) est très précis mais son utilisation est limitée par la nécessité d'avoir des stations au sol, leur difficulté d'utilisation, ainsi que par les conditions météorologiques. Il sert à calibrer les deux autres afin d'avoir l'orbite la plus précise possible. Le TRSR (Turbo Rogue Space Receiver) (GPS) fournit des données en complément de celles de Doris (Détermination d'Orbite et Radiopositionnement Intégrés par Satellite) pour la détermination temps réel et la détermination précise de l'orbite.

Le système Doris s'appuie sur un réseau de balises au sol (environ 50 sur tout le globe) qui émettent à deux fréquences en direction du satellite équipé d'un récepteur. Le mouvement relatif du satellite par rapport au sol provoque un décalage en fréquence (effet Doppler-Fizeau), analysé pour en déduire la vitesse relative de ce satellite. Ces données sont intégrées dans des modèles d'orbitographie, pour connaître en permanence la position précise (à 3 cm près) du satellite sur son orbite. (Contribution CNES)

TRSR : Comme sur Terre, la localisation GPS d'un satellite se fait par triangulation. L'objet étant repéré par au moins trois satellites GPS, on en déduit sa position exacte à un instant donné. Une fois intégrées dans un modèle d'orbitographie, ces données permettent de restituer en continu la trajectoire du satellite. (Contribution NASA)

LRA est un ensemble de miroirs destinés à réfléchir des tirs laser effectués depuis le sol. L'analyse du temps mis par le rayonnement laser pour faire le trajet aller-retour permet de localiser le satellite sur son orbite. (Contribution NASA)

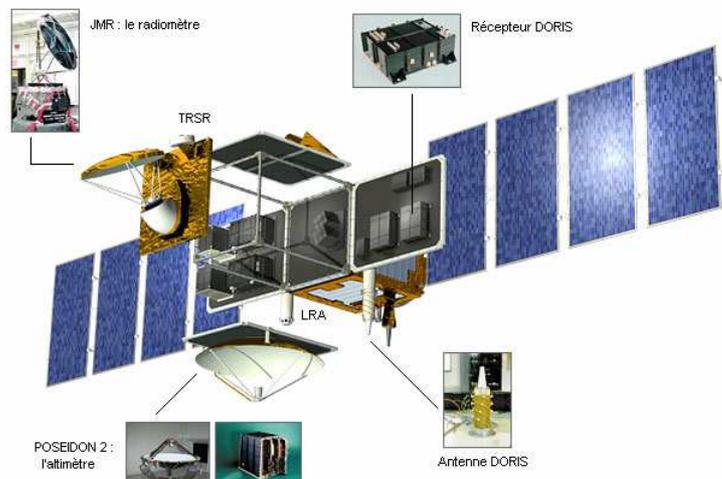


Figure A.8 – Positionnement des instruments à bords de Jason-1.

4.5.3 - Le segment sol

Le segment sol Jason-1 se compose de trois éléments, qui assurent la bonne marche de la mission, tant d'un point de vue humain (équipes) que matériel et informatique :

- ✎ le segment sol générique de Proteus (PGGS, Toulouse, France),
- ✎ le centre de contrôle projet (POCC, Pasadena, Ca, USA),
- ✎ le Segment Sol multimissions d'ALTimétrie, d'Orbitographie et de localisation précise (SSALTO, Toulouse, France).

Les opérations

Le contrôle du satellite est effectué d'une part par le Segment Sol Générique Proteus (PGGS : Proteus Generic Ground System, Toulouse, France), en début de mission, et par le Centre projet de contrôle des opérations (POCC : Project Operations Control Center, Pasadena, Ca), ainsi que par le Centre de Contrôle des Instruments (CCI : Centre de Contrôle des Instruments, Toulouse, France).

Le segment sol générique Proteus a la capacité d'opérer un satellite équipé d'une plateforme Proteus. Il comprend un terminal au sol, un centre de commande et de contrôle, ainsi qu'un réseau de transmission de données. Proteus fournit un tel équipement adapté à la mission Jason. Ce centre a été utilisé pour les premières phases de la mission (lancement, ...), et, quand le POCC a pris le relais pour les opérations de routine, pour des fonctions de navigation et d'analyse du satellite.

Le POCC assure les opérations pour Jason-1 durant sa phase de routine, y compris la télémesure et la commande temps réel.

Le traitement des données

Il est identique à celui de Topex-Poseidon (cf. partie 4.3.4).

4.6 - La mission Envisat

Le satellite Envisat (Environmental Satellite), successeur d'ERS-1 et -2, a été lancé le 1^{er} mars 2002 sur une orbite de 800 km à 35 jours. Dédié à l'étude de l'environnement, en particulier au suivi des changements climatiques, sa mission est d'observer l'atmosphère et la surface de la Terre. Construit par l'Esa, il a à bord une série d'instruments complémentaires (dont un altimètre radar et le système d'orbitographie et de localisation précise Doris) permettant d'observer le plus grand nombre de paramètres possibles.

Intégré aux nouveaux programmes internationaux d'étude du climat, comme GOOS et GODAE / Mercator, Envisat annonce l'entrée de l'océanographie spatiale dans l'ère opérationnelle, avec l'accès aux données en temps quasi-réel.

5 - Résultats complémentaires des analyses d'Hayashi et de Céron

5.1 - Analyses entre septembre 1997 et mai 2006

Cette partie annexe vient en complément des résultats montrés Chapitre 5-4.2.

5.1.1 - Les données utilisées

Les données analysées sont, d'une part les données mensuelles de concentration en chlorophylle issues du capteur SeaWiFS (dont les caractéristiques spatio-temporelles sont présentées Chapitre 5-4.2.1) et d'autre part, les données altimétriques d'anomalies de hauteur de mer issues de la combinaison de TOPEX-Poséidon, ERS-1/2, Jason-1 et Envisat [Ducet et al., 2000]. Les données altimétriques ont une résolution spatiale de **1/4°** et une résolution temporelle de **7 jours**. La durée temporelle de la série à analyser est d'un peu moins de 9 ans, plus exactement du 3 septembre 1997 au 30 mai 2006, soit **456 observations**. Une observation sur l'ensemble de la zone d'étude, le continent Antarctique mis à part, c'est-à-dire de 35°S à 50°S, amène une étude en latitude sur **61 points** avec une résolution de **27.8 km**. Une observation sur 360° en longitude amène une étude sur **1440 points** avec une résolution variant entre **22.8 km** (pour 35°S) et **17.9 km** (pour 50°S).

5.1.2 - Maxima des DSP obtenues par l'analyse d'Hayashi pour l'altimétrie

Tableau A.3 - Maxima des DSP (cm²/semaine) du signal entier et des parties stationnaire et propagative pour les dix paires les plus significatives obtenues par l'analyse d'Hayashi appliquées aux données altimétriques sur toute la zone 35°S-50°S. Les résultats obtenus sur les bandes 35°S-45°S et 45°S-50°S sont aussi indiqués.

Couples de nombres d'onde (zonal / méridional)	Bande 35°S-50°S			Bande 35°S-45°S			Bande 45°S-50°S		
	Total	Sta.	Prop.	Total	Sta.	Prop.	Total	Sta.	Prop.
2/0	0.0039	0.0009	0.0030	0.0051	0.0008	0.0041	0.0039	0.0009	0.0030
2/1	0.0053	0.0013	0.0036	x	x	x	0.0053	0.0013	0.0036
3/0	0.0196	0.0025	0.0161	0.0130	0.0026	0.0113	0.0196	0.0025	0.0161
3/1	0.0040	0.0009	0.0032	x	x	x	0.0040	0.0010	0.0032
4/0	0.0108	0.0030	0.0077	0.0154	0.0047	0.0107	0.0109	0.0030	0.0077
5/0	0.0129	0.0028	0.0098	0.0179	0.0040	0.0140	0.0129	0.0028	0.0098
6/0	0.0025	0.0004	0.0022	0.0026	0.0006	0.0023	0.0025	0.0004	0.0022
8/0	0.0043	0.0009	0.0033	x	x	x	0.0043	0.0009	0.0033
8/1	0.0022	0.0003	0.0020	x	x	x	0.0022	0.0003	0.0020
18/-1	0.0020	0.0004	0.0017	x	x	x	0.0020	0.0004	0.0017

5.1.3 - Résultats de l'analyse d'Hayashi pour le couple 2/0

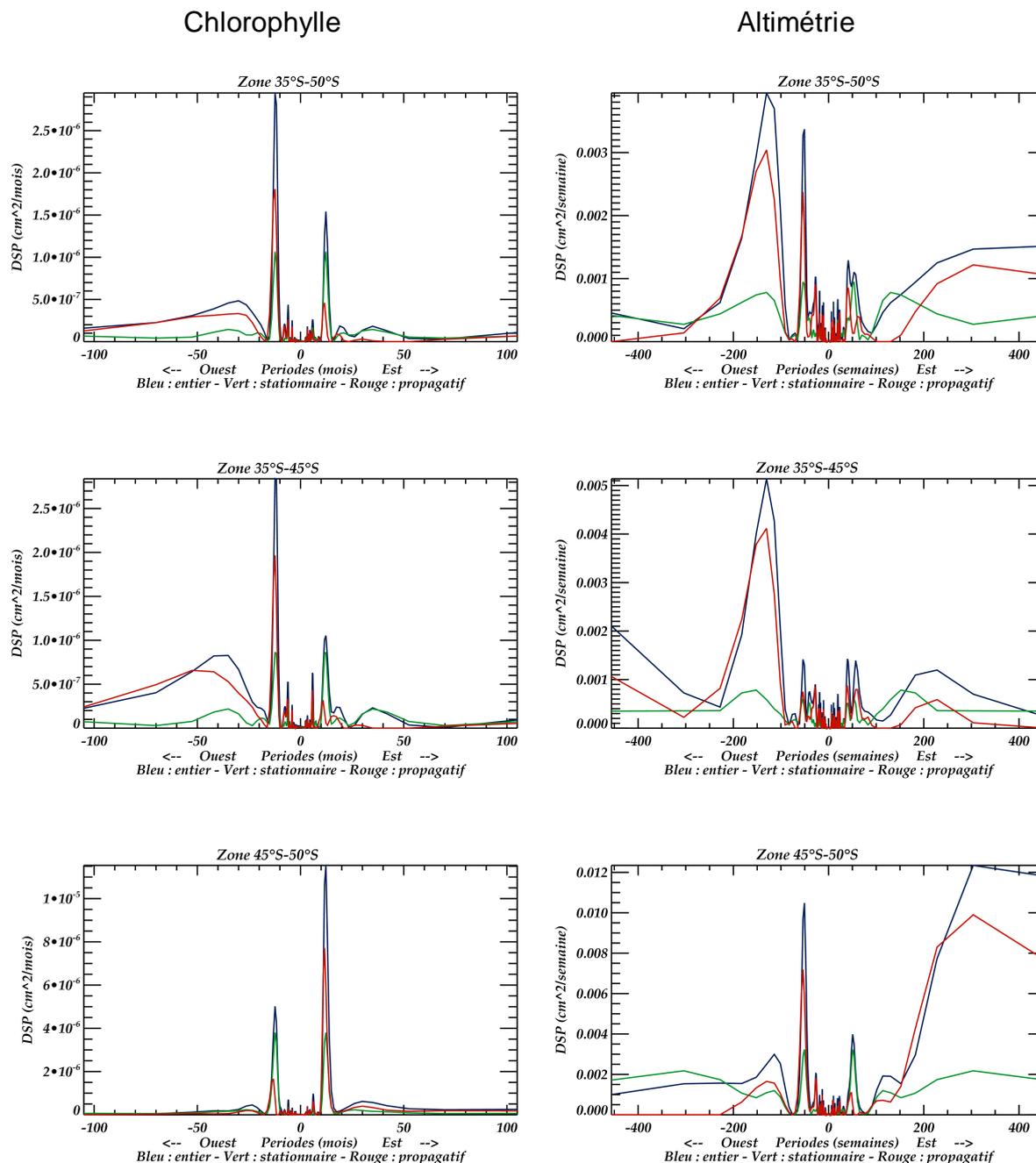


Figure A.9 – Variations de la DSP associée au couple 2/0 selon les périodes fournies par l'analyse d'Hayashi sur les domaines 35°S-50°S (en haut), 35°S-45°S (au milieu) et 45°S-50°S (en bas) pour le signal entier (bleu), la partie stationnaire (vert) et la partie propagative (rouge) et pour les données de chlorophylle mensuelle SeaWiFS (à gauche) et les données altimétriques (à droite) pour la période comprise entre septembre 1997 et mai 2006.

5.1.4 - Résultats de l'analyse de Céron appliquée aux données mensuelles SeaWiFS de concentration en chlorophylle pour le couple 2/0

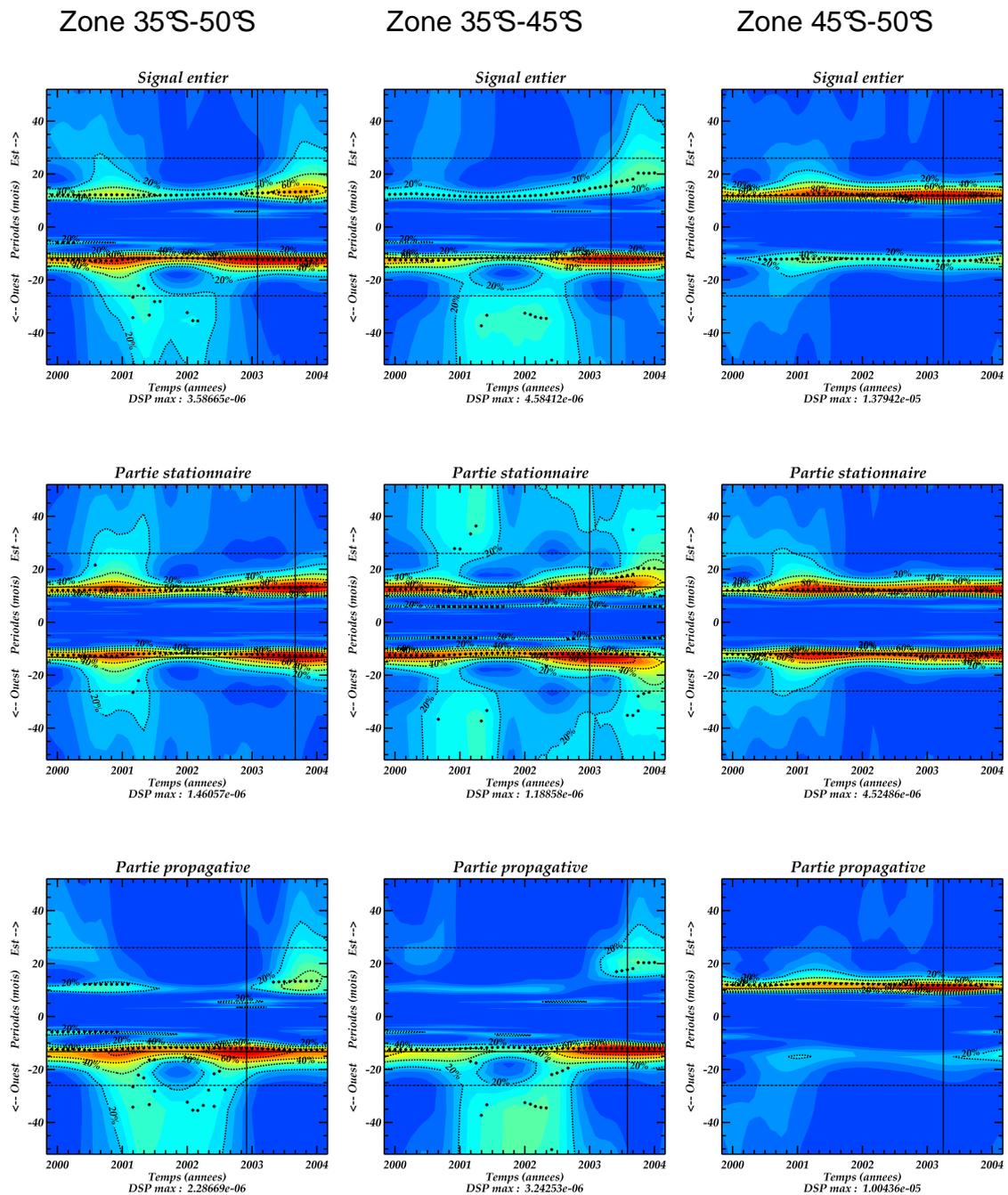


Figure A.10 – Variations en temps et en période de la DSP associée à la paire 2/0 du signal entier (en haut) et des parties stationnaire (au milieu) et propagative (en bas) pour les données de chlorophylle, sur les domaines 35°S-50°S (à gauche), 35°S-45°S (au milieu) et 45°S-50°S (à droite) et pour la période comprise entre septembre 1997 et mai 2006. La demi-largeur de la fenêtre analysante est de 26 mois = 2.2 ans.

5.1.5 - Résultats de l'analyse de Céron appliquée aux données altimétriques pour le couple 2/0

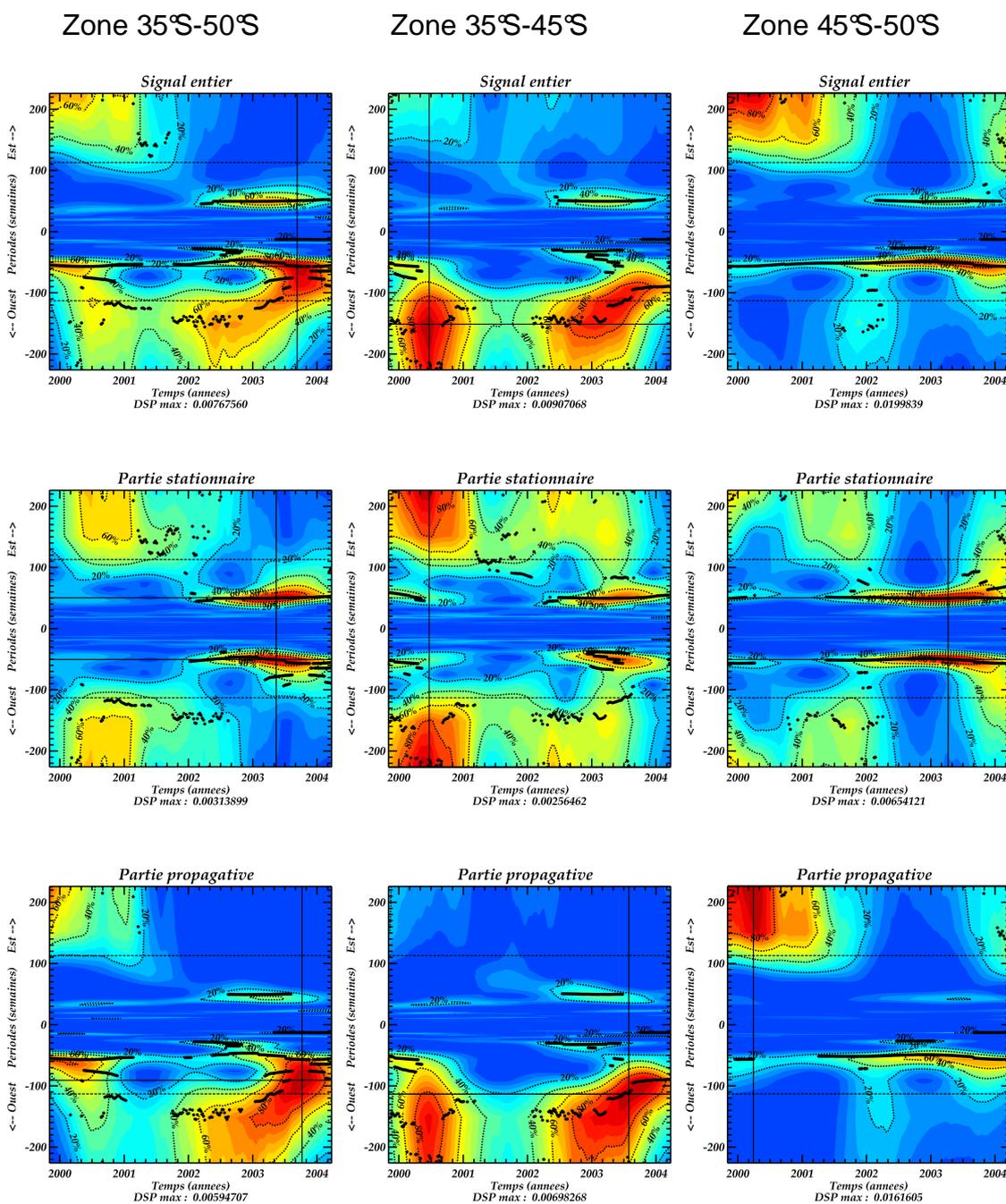


Figure A.11 – Variations en temps et en période de la DSP associée à la paire 2/0 du signal entier (en haut) et des parties stationnaire (au milieu) et propagative (en bas) pour les données altimétriques, sur les domaines 35°S-50°S (à gauche), 35°S-45°S (au milieu) et 45°S-50°S (à droite) et pour la période comprise entre septembre 1997 et mai 2006. La demi-largeur de la fenêtre analysante est de 113 semaines = 2.2 ans.

ANNEXE B

PUBLICATIONS

- 1 – C. POTTIER, J.-P. CÉRON, J. SUDRE, I. DADOU, S. BELAMARI ET V. GARÇON, 2004:
DOMINANT PROPAGATING SIGNALS IN SEA LEVEL ANOMALIES IN THE SOUTHERN OCEAN,
GEOPHYSICAL RESEARCH LETTERS, VOLUME 31, L11305, DOI:10.1029/2004GL019565..... 297
- 2 – C. POTTIER, V. GARÇON, G. LARNICOL, J. SUDRE, P. SCHAEFFER, ET P.Y. LE TRAON, 2006:
MERGING SEAWIFS AND MODIS/AQUA OCEAN COLOR DATA IN NORTH AND EQUATORIAL
ATLANTIC USING WEIGHTED AVERAGING AND OBJECTIVE ANALYSIS, *IEEE TRANSACTION ON
GEOSCIENCES AND REMOTE SENSING*, VOLUME 44, ISSUE 11, PART 2,
DOI:10.1109/TGRS.2006.878441, 3436-3451. 305

1 - C. Pottier, J.-P. Céron, J. Sudre, I. Dadou, S. Belamari et V. Garçon, 2004: Dominant propagating signals in sea level anomalies in the Southern Ocean, *Geophysical Research Letters*, Volume 31, L11305, doi:10.1029/2004GL019565.

Dominant propagating signals in sea level anomalies in the Southern Ocean

C. Pottier,^{1,2} J.-P. Céron,³ J. Sudre,¹ I. Dadou,¹ S. Belamari,³ and V. Garçon¹

Received 23 January 2004; revised 3 May 2004; accepted 12 May 2004; published 4 June 2004.

[1] We examine the time evolution over an 11 yr time series of the dominant modes of variability in sea level anomalies (SLA) from TOPEX/Poseidon and ERS-1/2 in the Southern Ocean (35–60°S). We use Hayashi's classical space-time spectral analysis and a newly developed method combining a Gabor analysis with Hayashi's method. Zonal wavenumbers 3 and 2 emerge unambiguously with null corresponding meridional wavenumber meaning that both associated modes have no meridional propagation. Their power spectrum density (PSD) is maximum within 50–55°S. The 3/0 mode displays a clear westward propagation associated with a 1.04 yr period. The 2/0 pair is characterized by a \sim 4.4 yr period and exhibits a clear eastward propagation with a maximal PSD just after the 1998 El Niño peak. Zonal wavenumber-2 mode is then probably part of the Antarctic Circumpolar Wave. *INDEX*

TERMS: 3339 Meteorology and Atmospheric Dynamics: Ocean/atmosphere interactions (0312, 4504); 4215 Oceanography: General: Climate and interannual variability (3309); 4560 Oceanography: Physical: Surface waves and tides (1255). **Citation:** Pottier, C., J.-P. Céron, J. Sudre, I. Dadou, S. Belamari, and V. Garçon (2004), Dominant propagating signals in sea level anomalies in the Southern Ocean, *Geophys. Res. Lett.*, 31, L11305, doi:10.1029/2004GL019565.

1. Introduction

[2] *White and Peterson* [1996, hereinafter referred to as WP96] reported anomalies in sea surface temperature (SST), atmospheric sea level pressure (SLP), meridional wind stress and sea ice extent propagating eastward around the Antarctic continent with a periodicity of 4–5 yr and a wave speed of \sim 8 cm/s. These concurrent anomalies, associated with a global zonal wavenumber-2 (ZWN2) pattern, were referred to as the Antarctic Circumpolar Wave (ACW). Simultaneously, *Jacobs and Mitchell* [1996, hereinafter referred to as JM96], using sea surface height, SST and wind stress curl data, reported a dominant mode characterized by a \sim 4 yr period and a 180° longitude wavelength. Using a 27-yr record of SST and SLP, *Carril and Navarra* [2001, hereinafter referred to as CN01] displayed a dominant ZWN2 structure on interannual time-scales, likely modulated by interdecadal variability. Concurrently, *Cai and Baines* [2001, hereinafter referred to as CB01] showed the presence of both ZWN 2 and 3 in the ACW structure during the last two decades of the

century, using historical SST. Their ACW major component was a ZWN2 structure forced by the Pacific-South America pattern, partially teleconnected with El Niño-Southern Oscillation (ENSO). Recently, *Venegas* [2003, hereinafter referred to as V03] confirmed that most of the variance of the ACW in the interannual band could be accounted for by a linear combination of two signals with different time and space characteristics. The first one has a 5 yr period and a dominant ZWN2 (ACW2) across the Southern Ocean. It is remotely forced by the tropical ENSO phenomenon with its strongest signature in the eastern Pacific sector. The second one is a ZWN3 mode with a 3.3 yr periodicity.

[3] V03 also noted that the SLP structure of the ACW2 signal resembles the dominant EOF mode of atmospheric variability referred to by *Thompson and Wallace* [2000] as the Southern Annular Mode (SAM). This atmospheric mode, first described by *Hartmann and Lo* [1998] and *Watterson* [2000] as a vacillation of the southern midlatitude zonal-mean westerly winds, is associated to the north-south displacement of the midlatitude westerly jet. It presents a zonally symmetric structure around the South Pole of opposite signs in the polar cap region and in the surrounding zonal ring centered near 45°S. The importance of SAM for the interannual variability in the Southern Hemisphere is still being discussed. *Hall and Visbeck* [2002, hereinafter referred to as HV02] suggested that “when adequate observations of all ocean variability in the Southern Ocean exist, most of the variability will ultimately be traced not to the ACW or ENSO but to the SAM”. In a comment paper responding to HV02's study, *White* [2003, hereinafter referred to as W03] conducted two independent analyses using a 20-yr record of SLP anomalies from the NCEP/NCAR reanalysis, and showed in contrast that the ACW and the teleconnected ENSO signals dominate the SAM in the real atmosphere's variability on interannual time scales.

[4] The aim of this article is to analyze the dominant oceanic signals in the Southern Ocean using combined TOPEX/Poseidon-ERS-1/2 sea level anomalies covering the 1992–2003 period, taking advantage of a newly developed method [*Céron et al.*, 2000, hereinafter referred to as C00].

2. Data and Method

[5] In this study, we analyze maps of sea level anomalies (SLA, cm) issued from the combined processing of TOPEX-Poseidon and ERS-1/2 data [*Ducet et al.*, 2000]. These data are available weekly on a 1/3° grid over the globe from October 1992 to September 2003 (569 cycles of 7 days). The study focuses on the 35–60°S region of the Southern Ocean; we thus need to deal with lands. The method used to estimate missing values is based on the

¹LEGOS/CNRS, Toulouse, France.

²CLS, Ramonville Saint Agne, France.

³Météo-France, Toulouse, France.

Least Mean Square (LMS) 1D method. Since the studied waves have almost a purely zonal propagation, we consider, for each latitude circle, each missing value as a linear combination of the 100 zonally closest data values, with a 10^{-5} step-size-convergence criterion in the LMS method.

[6] Two methods are used: the classical Hayashi's [1977, hereinafter referred to as H77; 1979] techniques and a newly developed one combining a Gabor analysis with the Hayashi method (C00). The bilinear tendency is removed from SLA data before performing the first step of both methods, which consists of a 2D Fourier analysis for each latitude-longitude map of SLA. This first step provides, for each observation, spatial spectral coefficients depending on the zonal and meridional wavenumbers. A selection of the most significant ones is then performed by comparing the modulus of each spatial spectral coefficient to a threshold depending on the white noise level: a coefficient is indeed kept as soon as its modulus exceeds once (over all observations) this threshold. The subsequent steps depend on the method to be used. In the formal H77 analysis, the second and last step of the method consists of the partition of the space-time power spectra into two parts, a standing one and a traveling one.

[7] We also use a new method that introduces a time-frequency analysis of the spatial spectral coefficients before Hayashi's partition of the space-time power spectra. Following C00, we chose a Gabor analysis, i.e., a Fourier transform with a Gaussian sliding window, which provides the localization of the periods in time. The values in the window correspond to weights allocated to the observations. The spectral estimation is improved by using a zero-padding allowing the calculation of the time Fourier transform over a double length sample. The largest width of the window to be used is $L_w = 569/2 \approx 284$ weeks, implying that only waves having a period less than $L_{max} = (L_w - 1)/2 \approx 142$ weeks (~ 3 yr) will be retrieved with confidence. With this new method, the second step consists of computing the Fourier transform of the windowed signal, while sliding continuously over the whole available sample. Each of the "local" space-time spectra computed through the Gabor analysis is thereby connected with the characteristics of the middle-window observations. Hayashi's partition into standing and traveling parts is then performed as a third step for all the space-time power spectra. The fourth and last step of C00's method consists of the skeleton extraction of the Gabor transform to get the instantaneous frequencies. The

Table 1. Maxima of the PSD (cm^2/week) (Complete Signal, Standing and Traveling Parts) for the 10 most significant pairs of wavenumbers obtained from C00 analysis

Pair of wavenumbers (zonal/ meridional)	PSD of the complete signal	PSD of the standing part	PSD of the traveling part
2/0	0.0285	0.0069	0.0215
2/1	0.0080	0.0018	0.0076
2/-1	0.0065	0.0015	0.0048
3/0	0.0341	0.0040	0.0319
3/1	0.0043	0.0008	0.0037
3/-1	0.0067	0.0016	0.0049
4/0	0.0051	0.0011	0.0049
5/0	0.0037	0.0014	0.0025
6/-1	0.0027	0.00084	0.0026
8/0	0.0036	0.0012	0.0023

The half-width of the sliding Gaussian window used is 142 weeks.

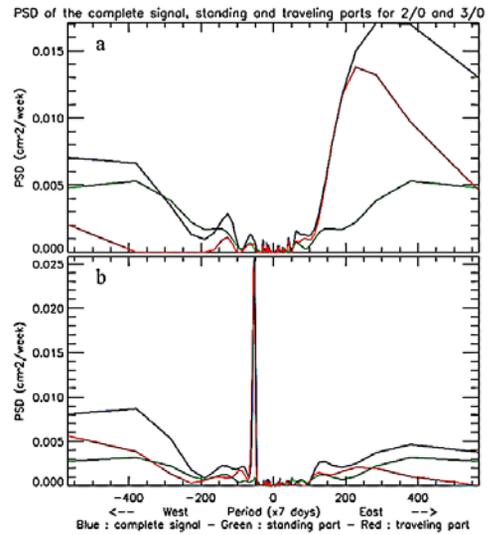


Figure 1. Period variations of the PSD provided by Hayashi's analysis over the $35\text{--}60^\circ\text{S}$ domain for the complete (blue line), standing (green line) and traveling (red line) signal associated with the 2/0 pair (a) and 3/0 pair (b).

algorithm is derived from Torrèsani [1995], and relies on the stationarity properties of the Gabor coefficients phase. This last point appears as a particularly useful issue since oceanic waves are time and amplitude modulated signals.

3. Results

[8] Table 1 shows the maxima of the power spectrum density (PSD) of the complete, standing and traveling parts of the signal analyzed over the $35\text{--}60^\circ\text{S}$ region. Only the 10 most significant pairs of wavenumbers are displayed. The importance of each mode is given by the PSD of the complete signal. Two modes emerge, given by pairs 3/0 and 2/0. As indicated by their meridional wavenumber equal to zero, the corresponding waves have no meridional propagation. These two modes (ZWN3 and ZWN2) are also characterized by maxima of the PSD higher for the traveling part than for the standing one: these modes are thus zonal propagating waves. A direct spatial localization of ZWN3 and ZWN2 modes is impossible, since the 2D Fourier analysis of the first step is global. To palliate this issue, we ran the Gabor analysis over several bands of latitude. The 3/0 and 2/0 pairs both display their maximum PSD in the same latitude band, namely between 50°S and 55°S (not shown). The 3/0 pair is however characterized by a higher meridional extent (from 35°S to 60°S) while the 2/0 pair is only significant south of 45°S .

[9] Figure 1 shows the results following H77 classical analysis, i.e., the variations in periods of the PSD, while Figure 2 exhibits the results of C00 analysis, i.e., the time and frequency variations of the PSD. Note that the latter (Figure 2) also displays the instantaneous periods (as bold dark lines) provided by the skeleton extraction of the Gabor transform. Positive (negative) periods indicate eastward (westward) propagation. The results for each wavenumber pair are analyzed as follows: first, the complete signal is studied to retrieve all its characteristics; then, the standing

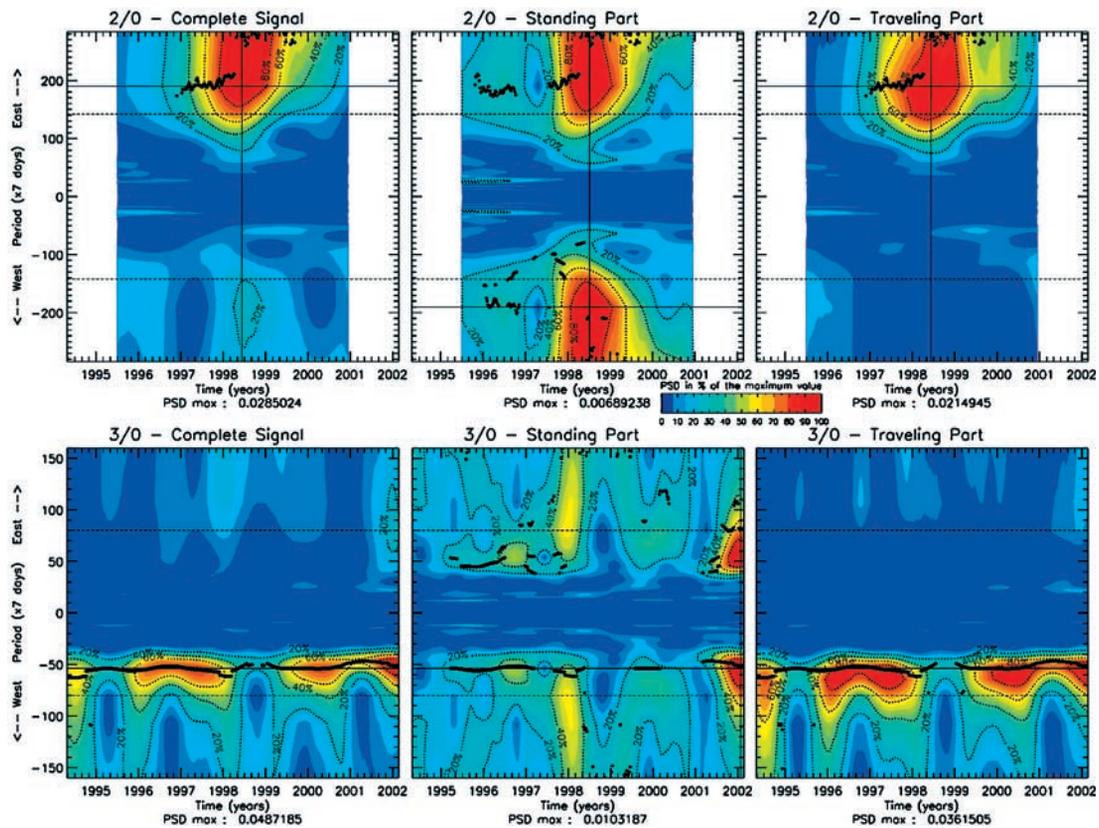


Figure 2. Time and frequency variations of the PSD of the Gabor transform applied over $35\text{--}60^\circ\text{S}$ to the complete signal (a), standing (b), and traveling (c) parts associated with the 2/0 pair (top) and 3/0 pair (bottom). The PSD is displayed as the percentage of its maximum value. Note that PSD_{max} (indicated below each x-axis) is different in each panel. The skeleton (instantaneous periods) is superimposed as bold dark lines. The white bands on the top panels are due to the use of a larger half-width of the Gaussian window for the 2/0 mode than for the 3/0 mode (142 against 80 weeks, see dotted lines). The intersection of the two thin continuous lines points out the maximum.

and traveling components are compared in order to sustain the first conclusions. Since H77 analysis is performed over the whole time series, it will display some information for periods higher (up to 284 weeks i.e., 5.5 yr) than those that may be detected by the Gabor analysis (no more than 142 weeks).

[10] Let's begin with the ZWN2 mode. The results of H77 analysis (Figure 1a) show that the major part of the complete signal is found for positive periods, which is supported by the partition in the standing and traveling parts. Indeed, the traveling part denotes clearly an eastward propagation with a period around 228 weeks (~ 4.4 yr), while the standing part is relatively weak. This eastward propagation is consistent with the results obtained following C00 analysis (Figure 2, top panels). The PSD highest values corresponding to an eastward propagating mode are observed between September 1997 and March 1999, with a pronounced maximum in June 1998 (Figure 2c). The associated period isolated by the skeleton extraction (190 weeks) is significantly lower than that provided by H77. It should in any case be considered with caution, since this period is much larger than the half-width of the sliding window ($L_{\text{max}} = 142$ weeks).

[11] For the ZWN3 mode, the results of H77 analysis (Figure 1b) point out an evident westward propagating

mode, with a period of ~ 54 weeks (~ 1.04 yr). This westward propagation is confirmed on Figure 2 (lower panels), that displays results obtained with C00 using a reduced half-width of the Gaussian window L_w (80 instead of 142 weeks), in order to increase the length of the studied sample while keeping a reasonable frequency resolution. The ZWN3 mode is partly filtered when the annual cycle is removed from the SLA dataset (not shown). It can however be interpreted with great confidence as a real westward propagating mode, since the continuity of the skeleton provided by the extraction algorithm denotes a structured phase around the 54 weeks period. Its time evolution is quite interesting (Figure 2c): the PSD has obviously two maxima (the first one between Dec. 95 and Nov. 97, the second one beginning in Sept. 99) with two nuclei each (resp. Feb. 96 and May 97, May 00 and Feb. 02). Note that the PSD exhibits also two minima (March 95 and Sept. 98), the latter corresponding to the PSD maximum of the 2/0 pair.

4. Discussion

[12] H77 and C00 analyses are used as two complementary methods to detect and characterize both standing and traveling modes in 1992–2003 SLA record of the Southern

Ocean. The main advantage of C00 analysis relies in the fact that this method provides the localization of the periods in time; its major shortcoming is related to the width of the analyzing window. The only modes that will be detected with accuracy with C00 method will indeed be those with a period smaller than the half-width of the largest window (142 weeks). For periods ranging between 142 and 284 weeks, the results may be questionable and should be confirmed through the H77 analysis. There is no ambiguity for the results concerning the ZWN3 westward propagating mode, which period is ~ 1 yr (54 weeks). In contrast, H77 analysis is essential to confirm the eastward propagation associated with the 2/0 mode which period (more than 4 yr) exceeds Gabor domain.

[13] The ZWN2 mode shows a clear eastward propagation that dominates all signals in the 50–60°S ring. This mode has a period of ~ 4.4 yr (228 weeks, following H77), and exhibits a maximal PSD in late autumn 1998. One can associate with confidence our 2/0 pair to the ACW, confirming previous studies with observed oceanic and atmospheric datasets (e.g., WP96, JM96, CN01, CB01, V03). The ACW path in SLP variability was shown to follow the subtropical front near 45°S in the eastern Atlantic, Indian, western and central Pacific sectors of the Southern Ocean before plunging to the South in the eastern Pacific to pass through Drake passage [White *et al.*, 2002]. The pronounced maximum found in the present study for the 2/0 pair within the 50–55°S region thus tends to suggest that the ACW amplitude could be stronger in the southeast Pacific and southwest Atlantic. Moreover, our results corroborate Peterson and White's [1998] scenario of some ACW source presumably related to the ENSO phenomenon. The study of the standardized SOI index from the NOAA Climate Prediction Center together with the PSD extracted for the traveling part of the 2/0 pair indicates that their time evolutions could be linked. The PSD starts increasing when the SOI drops to negative values marking the 1997–1998 El Niño onset. The minimal SOI value in March 1998 (–3.5) at the ENSO warm peak then precedes the PSD peak (in late autumn) by a few months (2–3). Last, the PSD decreases back again during the subsequent La Niña event.

[14] Let's consider now the 3/0 pair. H77 and C00 analyses reveal a global discernible ZWN3 pattern of variability in the SLA signal with a westward propagation and a 1.04 yr period. This result is consistent with JM96, who found in their Fourier decomposition a 1-yr period peak in the squared amplitude spectra of sea surface height, characterized by a ZWN3 and a westward propagation. A ZWN3 ACW component with a uniform phase over Antarctica was indeed found by CB01, who suggested a link with the atmospheric SAM. More recently, W03, using a Multi Taper Method Singular Value Decomposition analysis on SLP data, identified a signal near the 1.1 yr period, SAM in character but mixed with a global ZWN3 pattern of variability. Hughes *et al.* [2003] showed clearly that variability in subsurface pressure from Antarctic tide-gauges is related to atmospheric forcing in the form of the SAM. We plotted time series of the SAM index defined as the leading principal component of 850 hPa geopotential height anomalies (<http://tao.atmos.washington.edu/data/aa0>) together with the time series of

PSD obtained over the 35–60°S region for the ZWN3 traveling mode filtered at 54 weeks. The westward propagating mode associated with the 3/0 pair is maximum when SAM values are negative, and weakens when the SAM reaches its highest values in 1998 (not shown), thus allowing the standing mode to predominate over the traveling part (Figures 2b and 2c, lower panels). Following HV02, who concluded that a more intense Antarctic Circumpolar Current (ACC) is closely associated with a positive SAM, one might guess that our westward propagating ZWN3 mode may, due to Doppler effect, weaken and even become a minor component compared to the standing one when the ACC transport increases. The partition between traveling and standing components of our ZWN3 mode may then be linked to a strong coupling between the ACC and the atmosphere.

[15] In conclusion, we demonstrate the efficiency of H77 and C00 analyses to identify the dominant waves in SLA data in the Southern Ocean: the eastward propagating ACW (2/0 pair) and a westward propagating 3/0 pair. The significance of our results for ocean biogeochemistry is currently being investigated using the now available satellite chlorophyll long time series.

[16] **Acknowledgments.** This study was supported through CNES funding to LEGOS and by a CIFRE/ANRT fellowship to C.P. The altimeter products were produced by the CLS Space Oceanography Division as part of the Environment and Climate EU ENACT project and with support from CNES. We particularly thank Marie Drévillon for fruitful discussions, Ian Watterson and an anonymous reviewer for very constructive comments which improved our manuscript.

References

- Cai, W., and P. G. Baines (2001), Forcing of the Antarctic Circumpolar Wave by El Niño-Southern Oscillation teleconnections, *J. Geophys. Res.*, *106*(C5), 9019–9038.
- Carril, A. F., and A. Navarra (2001), Low-frequency variability of the Antarctic Circumpolar Wave, *Geophys. Res. Lett.*, *28*(24), 4623–4626.
- Céron, J.-P., S. Belamari, and L. Auger (2000), Gabor analysis and its application to the characterization of the time evolution of a 2D field, EGS Meeting, Nice, France.
- Ducet, N., P.-Y. Le Traon, and G. Reverdin (2000), Global high resolution mapping of ocean circulation from Topex/Poseidon and ERS-1 and -2, *J. Geophys. Res.*, *105*(C8), 19,477–19,498.
- Hall, A., and M. Visbeck (2002), Synchronous Variability in the Southern Hemisphere Atmosphere, Sea Ice, and Ocean Resulting from the Annular Mode, *J. Clim.*, *15*, 3043–3057.
- Hartmann, D. L., and F. Lo (1998), Wave-Driven Zonal Flow Vacillation in the Southern Hemisphere, *J. Atmos. Sci.*, *55*, 1303–1315.
- Hayashi, Y. (1977), On the Coherence between Progressive and Retrogressive Waves and a Partition of Space-Time Power Spectra into Standing and Traveling Parts, *J. Applied Meteorology*, *16*, 368–373.
- Hayashi, Y. (1979), A Generalized Method of Resolving Transient Disturbances into Standing and Traveling Waves by Space-Time Spectral Analysis, *J. Atmos. Sci.*, *36*, 1017–1029.
- Hughes, C. W., P. L. Woodworth, M. P. Meredith, V. Stepanov, T. Whitworth, and A. R. Pyne (2003), Coherence of Antarctic sea levels, Southern Hemisphere Annular Mode, and flow through Drake Passage, *Geophys. Res. Lett.*, *30*(9), 1464, 10.1029/2003GL017240.
- Jacobs, G. A., and J. L. Mitchell (1996), Ocean circulation variations associated with the Antarctic Circumpolar Wave, *Geophys. Res. Lett.*, *23*(21), 2947–2950.
- Peterson, R. G., and W. B. White (1998), Slow oceanic teleconnections linking the Antarctic Circumpolar Wave with tropical ENSO, *J. Geophys. Res.*, *103*(C11), 24,573–24,583.
- Thompson, D. W. J., and J. M. Wallace (2000), Annular Modes in the Extratropical Circulation. Part I: Month-to-Month Variability, *J. Clim.*, *13*, 1000–1016.
- Torrésani, B. (1995), *Analyse continue par ondelettes*, Savoirs actuels, InterEditions/CNRS Editions, pp. 239.
- Venegas, S. A. (2003), The Antarctic Circumpolar Wave: A Combination of Two Signals?, *J. Clim.*, *16*(15), 2509–2525.

- Watterson, I. G. (2000), Southern Midlatitude Zonal Wind Vacillation and its Interaction with the Ocean in GCM simulations, *J. Clim.*, *13*, 562–578.
- White, W. B. (2003), Comment on “Synchronous variability in the Southern Hemisphere atmosphere, sea ice, and ocean resulting from the annual mode,” *J. Clim.*, in press.
- White, W. B., S.-C. Chen, R. J. Allen, and R. C. Stone (2002), Positive feedbacks between the Antarctic Circumpolar Wave and the global El Niño-Southern Oscillation Wave, *J. Geophys. Res.*, *107*(C10), 3165, doi:10.1029/2000JC000581.
- White, W. B., and R. G. Peterson (1996), An Antarctic circumpolar wave in surface pressure, wind, temperature and sea-ice extent, *Nature*, *380*, 699–702.
-
- S. Belamari and J.-P. Céron, Météo-France, 42 Avenue Gaspard Coriolis, 31057 Toulouse Cedex, France.
- I. Dadou, V. Garçon, C. Pottier, and J. Sudre, LEGOS/CNRS, 18 Avenue Edouard Belin, 31401 Toulouse Cedex 9, France. (veronique.garcon@cnes.fr)

2 - C. Pottier, V. Garçon, G. Larnicol, J. Sudre, P. Schaeffer, et P.Y. Le Traon, 2006: Merging SeaWiFS and MODIS/Aqua Ocean Color Data in North and Equatorial Atlantic using Weighted Averaging and Objective Analysis, *IEEE Transaction on Geosciences and Remote Sensing*, Volume 44, Issue 11, Part 2, doi:10.1109/TGRS.2006.878441, 3436-3451.

IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING

A PUBLICATION OF THE IEEE GEOSCIENCE AND REMOTE SENSING SOCIETY

NOVEMBER 2006

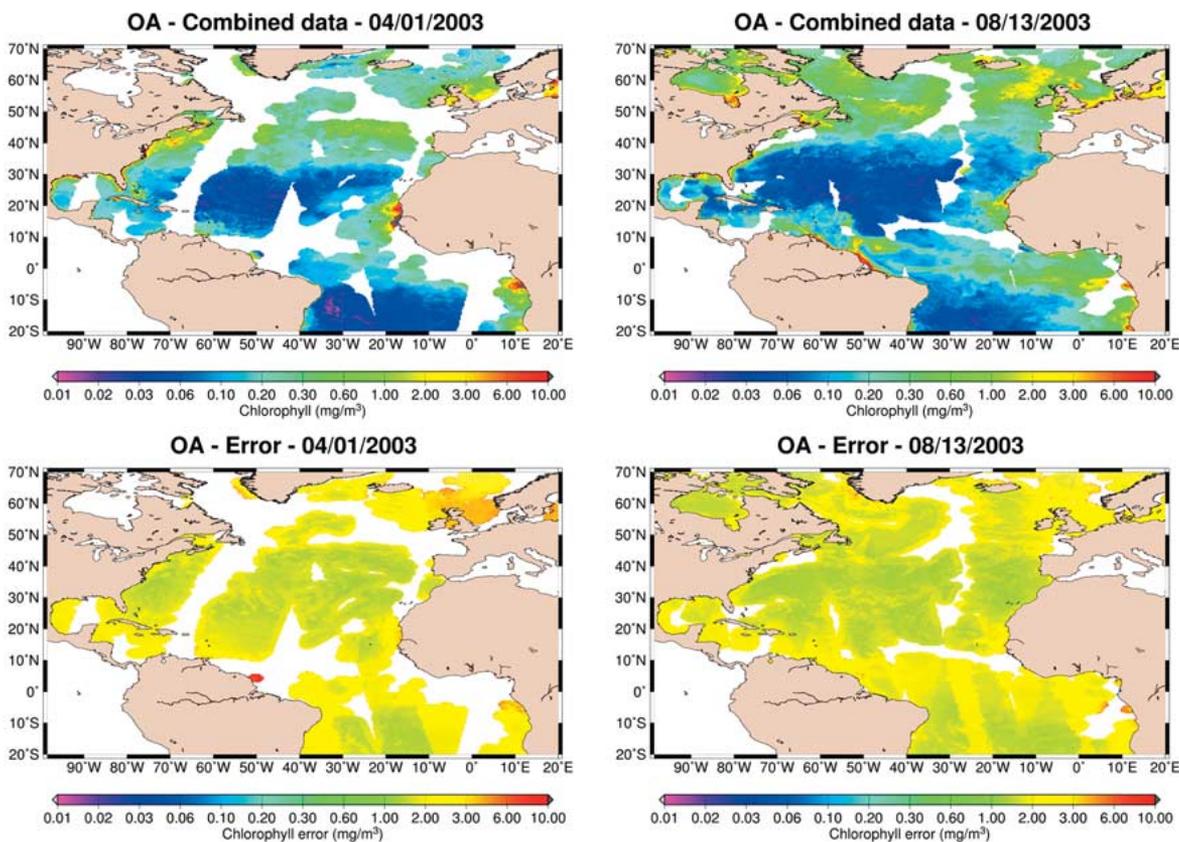
VOLUME 44

NUMBER 11

IGRSD2

(ISSN 0196-2892)

PART II OF TWO PARTS



Merging of daily maps of chlorophyll-*a* concentrations from SeaWiFS and MODIS/Aqua by objective analysis. Results of the application to logtransformed values for (left) April 1, 2003 and (right) August 13, 2003. (Top) Map of the combined data. (Bottom) Map of the associated error.

Merging SeaWiFS and MODIS/Aqua Ocean Color Data in North and Equatorial Atlantic Using Weighted Averaging and Objective Analysis

Claire Pottier, Véronique Garçon, Gilles Larnicol, Joël Sudre, Philippe Schaeffer, and Pierre-Yves Le Traon

Abstract—Two approaches of ocean color data merging were tested and compared in the North and Equatorial Atlantic Basin: the weighted averaging and the objective analysis. The datasets used were the daily level-3 binned data of chlorophyll-*a* from the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) and Moderate Resolution Imaging Spectroradiometer on the Aqua satellite over the year 2003, which is the first common full year of operation. Since they represent input for both approaches, matchups between the satellite and the *in situ* data from the SeaWiFS Bio-optical Archive and Storage System and the Atlantic Meridional Transect were first studied to compute a spatial map of the root mean-square error and of the bias. Because of the log distribution of the chlorophyll fields, each approach was applied to untransformed and log-transformed values. The application of the weighted averaging to log-transformed values does not show significant differences in comparison to its application to untransformed values. This is not the case, however, for the objective analysis that gives better results when applied to log-transformed values. Both approaches give combined chlorophyll data of equivalent quality, although the objective analysis could be improved with a better statistical characterization of noise and signal covariance. The main advantage of the objective analysis is its ability to interpolate in space (and time) by taking into account the characteristic scales of chlorophyll-*a*. As a result, the spatial coverage of the combined data is at least twice as large in the case of objective analysis than weighted averaging.

Index Terms—Biology, marine vegetation, sea surface, signal analysis.

I. INTRODUCTION

THE USE of satellite image data to investigate oceanic processes has become essential for oceanographic research and monitoring. The synoptic and global data play a fundamental role, since conventional platforms cannot adequately cover the vast and rapidly varying ocean at proper time and

space scales. Since the first ocean color mission, which is the Coastal Zone Color Scanner (CZCS), demonstrated the ability to observe globally the abundance and distribution of oceanic phytoplankton from space, the data have been used in many studies to better understand the role of the ocean in biogeochemical cycles, particularly as the biological carbon pump.

Chlorophyll-*a* (the main pigment in phytoplankton) distribution and variability can be determined from ocean color remote sensing. First, the normalized water-leaving radiance spectrum is derived from the calibrated top-of-atmosphere radiance measured by the satellite sensor at visible and near-infrared wavelengths by removing the contribution from the atmosphere through a process known as atmospheric correction [1]. Normalization refers to the correction for bidirectional effects (viewing angle dependence and effects of seawater anisotropy [2]). The spectrum of the normalized water-leaving radiance can subsequently be converted to chlorophyll-*a* concentration with biooptical algorithms, such as empirical band ratio formulations used for operational products [3].

Ten satellites having an ocean color sensor on board are now on orbit for regional or global scale applications. The main characteristics of the two studied here, which are the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) on OrbView-2 (NASA, USA) since August 01, 1997 and the Moderate Resolution Imaging Spectroradiometer (MODIS) on Aqua (NASA, USA) since May 04, 2002, are presented in Table I. Any individual ocean color mission is limited in ocean coverage by its swath width and gaps caused by clouds or sun glint, which disable the extraction of ocean color [4]. As an example, SeaWiFS can provide about 15% coverage of the global ocean at 0.5° resolution in one day under climatological cloudiness and sun glint [4]. This daily coverage would increase to 25% by merging SeaWiFS, MODIS/Aqua, and MODIS/Terra (MODIS sensor on the Terra satellite, since December 18, 1999) datasets. In this way, besides the increase of the statistical confidence in the data, the identification and study of biological and physical phenomena (such as primary production and algae blooms for instance) and their variability are made easier by filled maps.

There are many difficulties with ocean color data merging because the sensors have varying designs and characteristics. Let us have a look at those from SeaWiFS and MODIS/Aqua. SeaWiFS is in descending sun-synchronous near-polar circular orbit with a 12:20-noon local equator crossing time, whereas MODIS/Aqua is in ascending orbit with a 1:30 PM equator crossing time (Table I). The MODIS scan-mirror assembly uses a continuously rotating double-sided scan mirror to scan

Manuscript received July 4, 2005; revised April 11, 2006. This work was supported in part by the FP6 European IP Marine Environment and Security for the European Area (MERSEA) funding and in part by the CIFRE/ANRT fellowship to C. Pottier.

C. Pottier is with the Collecte Localisation Satellites (CLS) Space Oceanography Division, 31520 Ramonville-Saint-Agne, France, and also with the LEGOS/Centre National de la Recherche Scientifique (CNRS), 31401 Toulouse Cedex 9, France (e-mail: claire.pottier@cls.fr; claire.pottier@notos.cst.cnes.fr).

V. Garçon and J. Sudre are with the LEGOS/CNRS, 31401 Toulouse Cedex 9, France (e-mail: veronique.garcon@cnes.fr; joel.sudre@cnes.fr).

G. Larnicol and P. Schaeffer are with CLS Space Oceanography Division, 31520 Ramonville-Saint-Agne, France (e-mail: gilles.larnicol@cls.fr; philippe.schaeffer@cls.fr).

P.-Y. Le Traon is with Institut Français de Recherche pour l'Exploitation de la Mer (IFREMER), 29280 Plouzané, France (e-mail: pierre.yves.le.traon@ifremer.fr).

Digital Object Identifier 10.1109/TGRS.2006.878441

TABLE I
CHARACTERISTICS OF SEAWIFS AND MODIS/AQUA SENSORS

Sensor	SeaWiFS	MODIS/Aqua
Satellite	Orbview-2	Aqua
Launch date	08/01/1997	05/04/2002
Equator crossing time	12:20 noon	1:30 pm
Spectral bands center (nm)	8 bands: 413 443 490 510 555 670 765 865	36 bands, with 412 443 488 531 551 667 & 678 748 869
Data available since...	09/04/1997	11/29/2002
Swath width	2,806 km	2,330 km
Spatial resolution	1100 m (LAC) 4500 m (GAC)	1000 m
Tilt	-20°, 0°, +20°	No
Orbit	Descending	Ascending

LAC stands for Local-Area Coverage. GAC stands for Global-Area Coverage.

$\pm 55^\circ$ [5]; its viewing swath is 10 km along track at nadir and 2330 km cross track at $\pm 55^\circ$. SeaWiFS is equipped with an off-axis folded telescope coupled with a rotating half-angle mirror (which rotates at half speed of the folded telescope), and that is phase-synchronized (<http://oceancolor.gsfc.nasa.gov/SeaWiFS/SEASTAR/SPACECRAFT.html>). In addition to the polarization scrambler, this arrangement permits a minimum level of polarization to be achieved: SeaWiFS nominal polarization sensitivity is therefore 20 times less than the maximum polarization sensitivity of MODIS/Aqua, and a polarization correction is thus needed for MODIS/Aqua to improve the agreement between both sensors' water-leaving radiances [6]. Although the maximum scan angle of SeaWiFS is $\pm 58.3^\circ$ at MODIS-like altitude, the SeaWiFS global-area coverage (GAC) swath is limited to 1502 km. Contrary to MODIS/Aqua, SeaWiFS tilts during each orbit in order to reduce the effects of sun glint from the sea (Table I; <http://oceancolor.gsfc.nasa.gov/SeaWiFS/SEASTAR/SPACECRAFT.html>). As a result, SeaWiFS' products show a zonal band without any data. Both sensors provide a global coverage of the earth in two days. MODIS has 36 spectral bands from which nine are used for ocean color (bands #8 to #16, the others are used for measuring land/cloud/aerosols properties, atmospheric, and cirrus-cloud water vapor, surface/cloud/atmospheric temperature, cloud top altitude, and ozone; <http://modis.gsfc.nasa.gov/about/specifications.php>). SeaWiFS has eight spectral bands: all are used for ocean color measurement. MODIS bands 1–2 have a nominal nadir resolution of 250 m; bands 3–7 have a nadir resolution of 500 m; and bands 8–36 have a nadir resolution of 1 km [5]. SeaWiFS resolution for all eight bands is 1.1 km at nadir (Table I). Since GAC data are subsampled every fourth pixel and every fourth line, SeaWiFS' resolution is 4.5 km. MODIS spectral-band centers dedicated to ocean color are not exactly the same as SeaWiFS' (Table I). Therefore, the empirical algorithm to compute chlorophyll-*a* concentration from the normalized water-leaving radiances is not exactly the same for both sensors. However, the MODIS OC3M algorithm is similar to the SeaWiFS OC4v4 algorithm ([\[nasa.gov/eval/oc.cgi\]\(http://seabass.gsfc.nasa.gov/eval/oc.cgi\)\). Since merger activities depend on the calibration and validation quality of data products, differences in standard products among missions need to be evaluated: It is not the object of this study, but the interested reader can refer to \[7\]–\[9\].](http://seabass.gsfc.</p>
</div>
<div data-bbox=)

Data-merging methods can start at the level of the water-leaving radiance or at the level of the derived products such as chlorophyll. Maritorena and Siegel [10] used normalized water-leaving radiances from SeaWiFS and MODIS/Aqua together in a semianalytical biooptical model (GSM01) to produce global retrievals of three biogeochemically relevant variables (chlorophyll, combined dissolved and detrital absorption coefficient, and particulate backscattering coefficient). Gregg and Conkright [11] used a blended analysis to combine *in situ* data and satellite CZCS (NASA) chlorophyll in order to construct a climatological seasonal representation of the global chlorophyll distribution. Within the NASA SIMBIOS program, several methods were tested from the simple binning of the daily chlorophyll from different sources mapped to a common grid to more sophisticated methods [12]–[14].

Our objective here within the Global Monitoring for Environment and Security (GMES) integrated project Marine Environment and Security for the European Area (MERSEA) is to provide an accurate and consistent stream of ocean color data at a resolution and extent of coverage compatible with operational forecasting of the marine environment. Because we want to obtain daily maps of chlorophyll, we chose here to directly combine the chlorophyll-concentration products of both SeaWiFS and MODIS/Aqua. In the next section, we will present the datasets used and the results of the comparison to the comprehensive archives of *in situ* chlorophyll data. We then present and compare two approaches over the North and Equatorial Atlantic region for the year 2003: the weighted averaging and the objective analysis first introduced in oceanography by Bretherton *et al.* [15] and frequently used in altimetry [16]. A performance assessment of each approach will finally be provided at the North Atlantic basin scale as well as at the biogeochemical-province scale.

II. DATA

We used SeaWiFS and MODIS/Aqua daily level-3 binned data of chlorophyll concentration obtained from the NASA/Goddard Earth Science (GES)/Distributed Active Archive Center (DAAC). The datasets are issued from the reprocessing version 4 (July 2002, <http://oceancolor.gsfc.nasa.gov/REPROCESSING/SeaWiFS/R4/>) for SeaWiFS and version 1 (February 2005, <http://oceancolor.gsfc.nasa.gov/REPROCESSING/Aqua/R1/>) for MODIS/Aqua. The data were all considered whatever their flags are. These data are both on an equal area grid, in which the resolution is $1/12^\circ$ and $1/24^\circ$ at the Equator, respectively. For practical considerations, we resampled them, for this paper, on a regular grid of $0.1^\circ \times 0.1^\circ$ (i.e., 11.1×11.1 km at the Equator) for SeaWiFS and $0.05^\circ \times 0.05^\circ$ (i.e., 5.6×5.6 km at the Equator) for MODIS/Aqua. Fig. 1 shows an example of the location of a SeaWiFS and a MODIS/Aqua chlorophyll pixel at longitude *i* and latitude *j*). These resolutions were selected because they stand as a compromise between not being far from the original

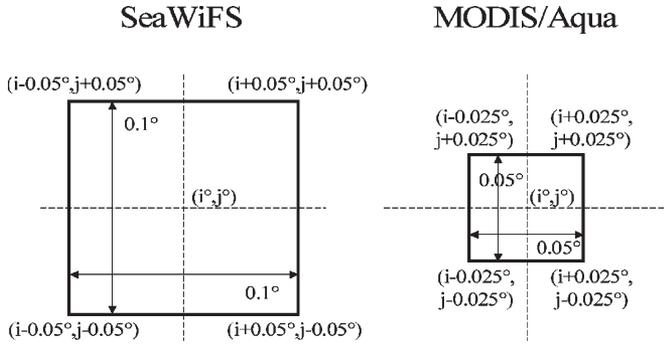


Fig. 1. Example of the location of SeaWiFS' and MODIS/Aqua's chlorophyll pixel at longitude i and latitude j .

level-3 resolutions and saving disk space. In addition, they conserve the ratio between the resolutions of level-3 binned data of both sensors. We will focus our study on the year 2003, which is the first common full year of operation of both SeaWiFS and MODIS/Aqua sensors. Our area of interest is the North and Equatorial Atlantic region, comprised between 98.5° W– 20° E and 20° S– 70° N. The distribution of untransformed and log-transformed SeaWiFS chlorophyll daily data (example shown: August 13, 2003) in the global ocean and in the North Atlantic basin (Fig. 2) confirms the lognormal distribution of ocean chlorophyll found by Campbell [17]: the logarithm of chlorophyll concentrations has a normal distribution. We will thus apply the merging algorithms on untransformed and log-transformed values.

In situ chlorophyll measurements collected from ships or oceanic platforms are usually used to validate original SeaWiFS, MODIS/Aqua, or even the combined chlorophyll datasets: These coincident in time and coverage observations between *in situ* and satellite measurements are called matchups. This allows computing errors and bias between each dataset. Two *in situ* datasets are used here. The first one was obtained from the NASA SeaWiFS Bio-optical Archive and Storage System (SeaBASS) [18] (Gregg, personal communication). Data are collected using a number of instrument packages (such as profilers and handheld instruments) on a variety of platforms (including ships and moorings). In the studied region, it includes 704 fluorometrically/spectrophotometrically derived chlorophyll-concentration values (milligrams per cubic meter) at depths of 0–5 m for the selected 2003-year period. The second set contains the Atlantic Meridional Transect (AMT) [19] underway data, which are obtained through the British Oceanographic Data Centre (BODC). The AMT Program undertakes biological, chemical, and physical oceanographic research during both annual passages of the R/V James Clark Ross between Plymouth (U.K.) and the Falkland Islands. Underway instruments continuously take measurements of the sea surface data (e.g., salinity, temperature, attenuation, chlorophyll, and nutrients) by measuring continuously pumped surface seawater. The database includes 62 613 fluorometrically derived chlorophyll-concentration values (milligrams per cubic meter) from through flow over two one-month periods: from May 12, 2003 to June 17, 2003 (AMT12 cruise) and from September 11, 2003 to October 13, 2003 (AMT13 cruise). This led to 63 317 *in situ* measurements. *In situ* data that were coincident (occurring

within the same day) and collocated (occurring within a single SeaWiFS or MODIS/Aqua level-3 pixel of chlorophyll) were averaged, following [20]. This led to 2754 *in situ* values on a 0.1° grid and 5144 on a 0.05° grid. After considering coincident, collocated averages, and cloud-free SeaWiFS and MODIS/Aqua data, the final result was 893 comparison matchup points for SeaWiFS and 1476 for MODIS/Aqua over 2003 in the North Atlantic basin. The locations of the matchups for both sensors are shown in Fig. 3. To perform a regional analysis, we subdivided the basin into 19 ocean domains, which are the biogeochemical provinces defined in [21] (Fig. 3). Statistical analyses were performed both globally and within these provinces.

The coefficient of determination r^2 from the correlation analysis indicates the covariance between the satellite dataset and the *in situ* observations. For both sensors, the open ocean has higher r^2 than the coastal provinces, as shown in Fig. 4. Actually, satellite chlorophyll concentrations less than 1 mg/m^3 , which are generally found in the open ocean, have a better agreement with *in situ* data than chlorophyll concentrations higher than 1 mg/m^3 (generally found in the coastal provinces).

For untransformed values, the error measurement is estimated by the root-mean-square (rms) error, which is defined as

$$\text{rms}_d(\text{mg/m}^3) = \sqrt{\frac{1}{N} \sum (S - I)^2} \quad (1)$$

and the bias is estimated by the averaged difference (AD) defined as

$$\text{AD}_d(\text{mg/m}^3) = \frac{1}{N} \sum (S - I) \quad (2)$$

where S indicates satellite (SeaWiFS or MODIS/Aqua) chlorophyll concentration, I indicates *in situ* chlorophyll concentration, and N is the number of samples.

For log-transformed values, rms_l and AD_l are computed by replacing S and I by $\log(S)$ and $\log(I)$, respectively (“log” is the logarithm to base 10). Fig. 5 shows the values found per biogeochemical province and for the Atlantic basin. The rms error and the bias computed on chlorophyll are relatively weak in the open ocean (the rms error is less than 0.09 mg/m^3 in the subtropical gyres and less than 0.7 mg/m^3 everywhere else; the bias is less than 0.04 mg/m^3 in the subtropical gyres and less than 0.4 mg/m^3 everywhere else) and higher in the coastal provinces (the rms error and the bias are greater than 1 mg/m^3). SeaWiFS is performing slightly better than MODIS/Aqua at the Atlantic basin scale. There are differences between the atmospheric correction schemes and biooptical algorithms applied to MODIS and SeaWiFS data. The much higher uncertainties affecting MODIS/Aqua's chlorophyll- a with respect to SeaWiFS' in eastern U.S. coastal waters suggest that these differences might be highlighted in the presence of some optically complex waters and/or continental aerosols. In general, MODIS/Aqua's water-leaving radiances in the blue channels are lower than SeaWiFS' equivalents and field measurements [22]. This has been confirmed in coastal waters [23]. Eventually, it is likely that the determination of chlorophyll- a is affected in different ways by the uncertainties associated with the respective radiometric products.

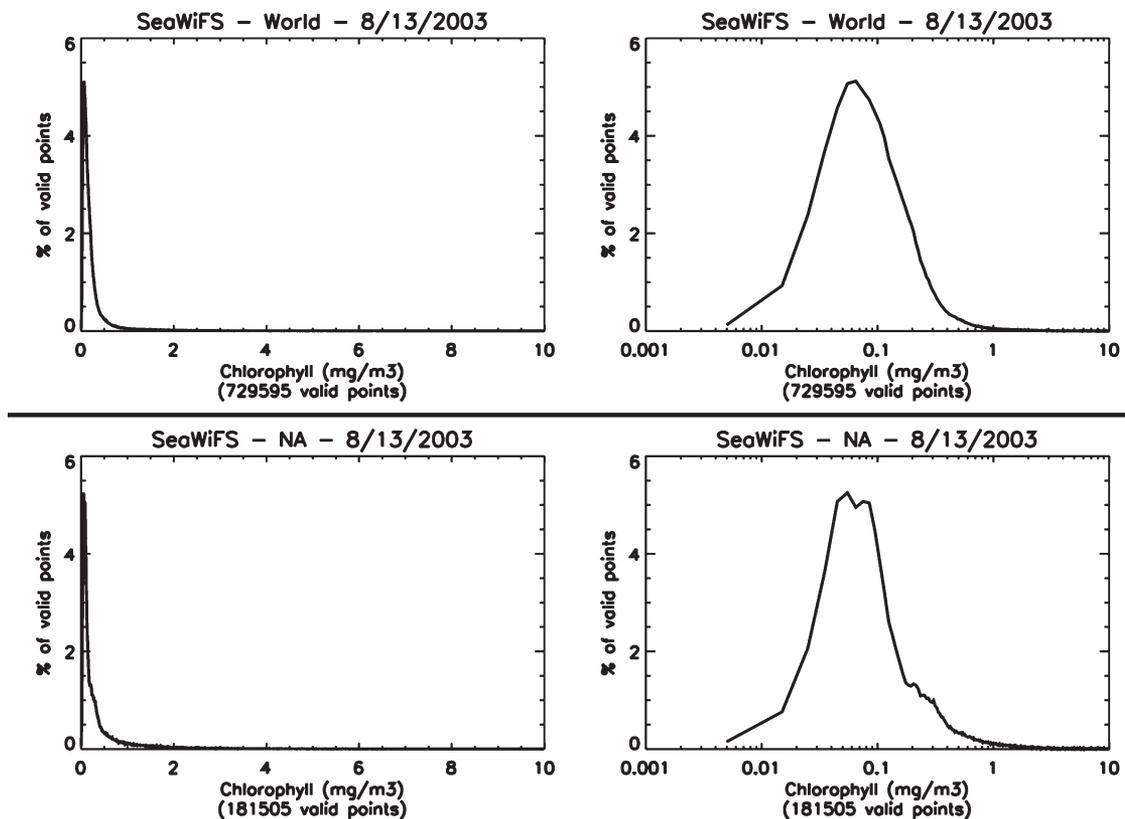


Fig. 2. Distribution of (left) untransformed and (right) log-transformed SeaWiFS chlorophyll daily data for August 13, 2003 in (top) the global ocean and (bottom) in the North and Equatorial Atlantic.

For our merging methods, a map of the measurement error and/or a map of the bias between *in situ* and satellite data are needed. We thus used the values given in Fig. 5. For the biogeochemical provinces for which there was no matchup, the rms error and the bias given were those from the nearest province. Nearest province means the closest geographically and the closest biogeochemically. These provinces associations are thus made following the biome (polar, westerlies, trade winds, and coastal; [21]) to which they belong (cf. Table II). In order to avoid sharp gradients on rms error and bias at boundaries between the biogeochemical provinces, the rms error and bias fields were smoothed at 100 km.

Results will be presented here for two particular days. First, August 13, 2003, for which the spatial coverage is remarkably good for both sensors (Fig. 6, right panel). Indeed, it is 23% for SeaWiFS and 20% for MODIS/Aqua. The second day chosen, which is April 1, 2003, is quite particular: In addition to the sparse spatial coverage of SeaWiFS (12%), MODIS/Aqua presents only one track (4%; Fig. 6, left panel).

III. WEIGHTED AVERAGING

A. Description

The first approach used is the weighted averaging. From the map of chlorophyll concentrations for a given day and the map of the measurement errors for each sensor [i.e., the rms error computed between *in situ* measurements and each sensor data (see Section II)], we obtain the map of the combined chlorophyll for this day and the map of the associated errors.

The value of a combined chlorophyll pixel equals the weighted averaging of the chlorophyll pixels of each sensor and at the same location. The weights are the confidence we have in each sensor in comparison to the other, which are computed from the rms errors. Since SeaWiFS and MODIS/Aqua have not the same spatial resolution, two choices are available: The resolution of the combined data is either the lowest one (0.1°: Algorithm WA1) or the highest one (0.05°: Algorithm WA2).

1) *Algorithm WA1*: Let us begin with the untransformed values. The first step of Algorithm WA1 [see Fig. 7(a)] consists of oversampling MODIS/Aqua's chlorophyll pixels to bring them to SeaWiFS' resolution. Considering the location of the chlorophyll pixels of each sensor (Fig. 1), MODIS/Aqua's candidate M is obtained from nine original chlorophyll pixels with the following formula:

$$\begin{aligned}
 M &= \frac{1}{4}M_5 + \frac{1}{8}(M_2 + M_4 + M_6 + M_8) \\
 &\quad + \frac{1}{16}(M_1 + M_3 + M_7 + M_9) \\
 &= \sum_{i=1}^9 \lambda_i M_i
 \end{aligned} \tag{3}$$

with $\sum_{i=1}^9 \lambda_i = 1$.

The λ_i coefficients represent the fraction area of each MODIS/Aqua's original chlorophyll pixel covered by the MODIS/Aqua's candidate. Of course, only the existing data points are considered.

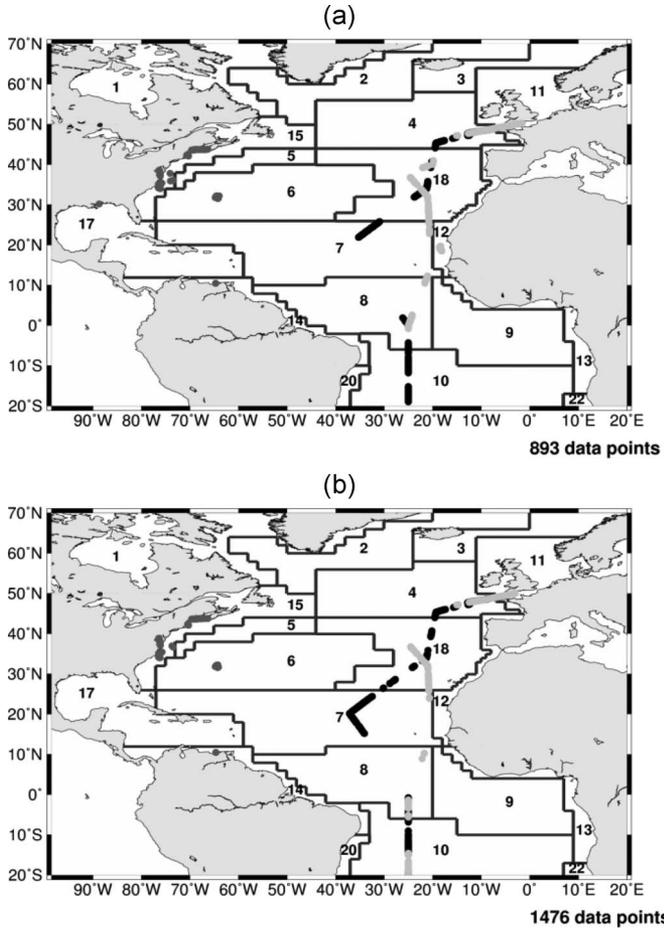


Fig. 3. Location of the matchups for (a) SeaWiFS and (b) MODIS/Aqua with SeaBASS's (in midgray), AMT12 cruise's (in black), and AMT13 cruise's (in pale gray) data. The biogeochemical provinces defined and numbered following [21] are delimited by gray lines (cf. Table II).

To compute the associated error EM, we can consider the distribution with an M mean and an EM standard deviation as a linear combination of the nine distributions of an M_i mean and an EM_i standard deviation (i.e., the error associated with the M_i chlorophyll pixel, which is the rms error associated with the point and is computed using matchups with *in situ* data in Section II). Since no specific rule exists to compute the standard deviation of a linear combination of lognormally distributed variables, the error (i.e., standard deviation) EM of M must be computed in a basic way as follows:

$$EM = \sqrt{\sum_{n=1}^N X_n^2} \quad \text{with } X_n = \sum_{i=1}^9 \lambda_i Y_{i,n}.$$

First, we build nine lognormal distributions Y_i ($i = \{1, \dots, 9\}$) of N points (after several tests, we chose $N = 1000$), with an M_i mean and an EM_i standard deviation; then, we compute their linear combination X_n (the weights are the λ_i); the error EM of M is the standard deviation of X_n . The lack of several values out of the nine brings so little difference that we considered it had no real impact on the EM value.

The second and last step of Algorithm WA1 is the combination of SeaWiFS and MODIS/Aqua's candidates. The com-

bined chlorophyll pixel C is computed using the following:

$$C = \left(1 - \frac{\%ES}{\%EM + \%ES}\right) S + \left(1 - \frac{\%EM}{\%EM + \%ES}\right) M \quad (4)$$

where $\%EX = EX/X$. Actually, the weights of S and M represent the percentage of confidence we have in the corresponding chlorophyll pixel in comparison to the other. The associated error EC is computed in the same way as in step 1, i.e., it is considered as the standard deviation of a linear combination of two lognormal distributions of means S and M and standard deviations ES and EM.

To obtain MODIS/Aqua's candidate for log-transformed values, the “ X ” values are replaced by the “log X ” values. The geometric mean is used. Therefore, (3) is replaced by

$$M = 10^m \quad \text{with } m = \sum_{i=1}^9 \lambda_i \log M_i. \quad (5)$$

This time, let EM_i be the error associated with $\log(M_i)$. Since $\log(M_i)$ has a normal distribution, we have

$$EM = \sqrt{\sum_{i=1}^9 (\lambda_i EM_i)^2}. \quad (6)$$

To compute the combined chlorophyll pixel, (4) is replaced by the following:

$$C = 10^{m'} \quad (7)$$

with

$$m' = \left(1 - \frac{\%ES}{\%EM + \%ES}\right) \log S + \left(1 - \frac{\%EM}{\%EM + \%ES}\right) \log M.$$

As in the first step, the associated error EC is the linear combination of the errors of $\log(S)$ and $\log(M)$.

2) *Algorithm WA2*: Let us begin with the untransformed values. The first step of Algorithm WA2 [see Fig. 7(b)] consists of bringing down SeaWiFS' chlorophyll pixels to the resolution of MODIS/Aqua's. Considering the location of the chlorophyll pixels of each sensor (Fig. 1), SeaWiFS' candidates are obtained as the following:

$$\begin{aligned} S1' &= \frac{1}{4}(S1 + S2 + S4 + S5) \\ S2' &= \frac{1}{2}(S2 + S5) \\ S3' &= \frac{1}{4}(S2 + S3 + S5 + S6) \\ S4' &= \frac{1}{2}(S4 + S5) \\ S5' &= S5 \\ S6' &= \frac{1}{2}(S5 + S6) \\ S7' &= \frac{1}{4}(S4 + S5 + S7 + S8) \\ S8' &= \frac{1}{2}(S5 + S8) \\ S9' &= \frac{1}{4}(S5 + S6 + S8 + S9). \end{aligned} \quad (8)$$

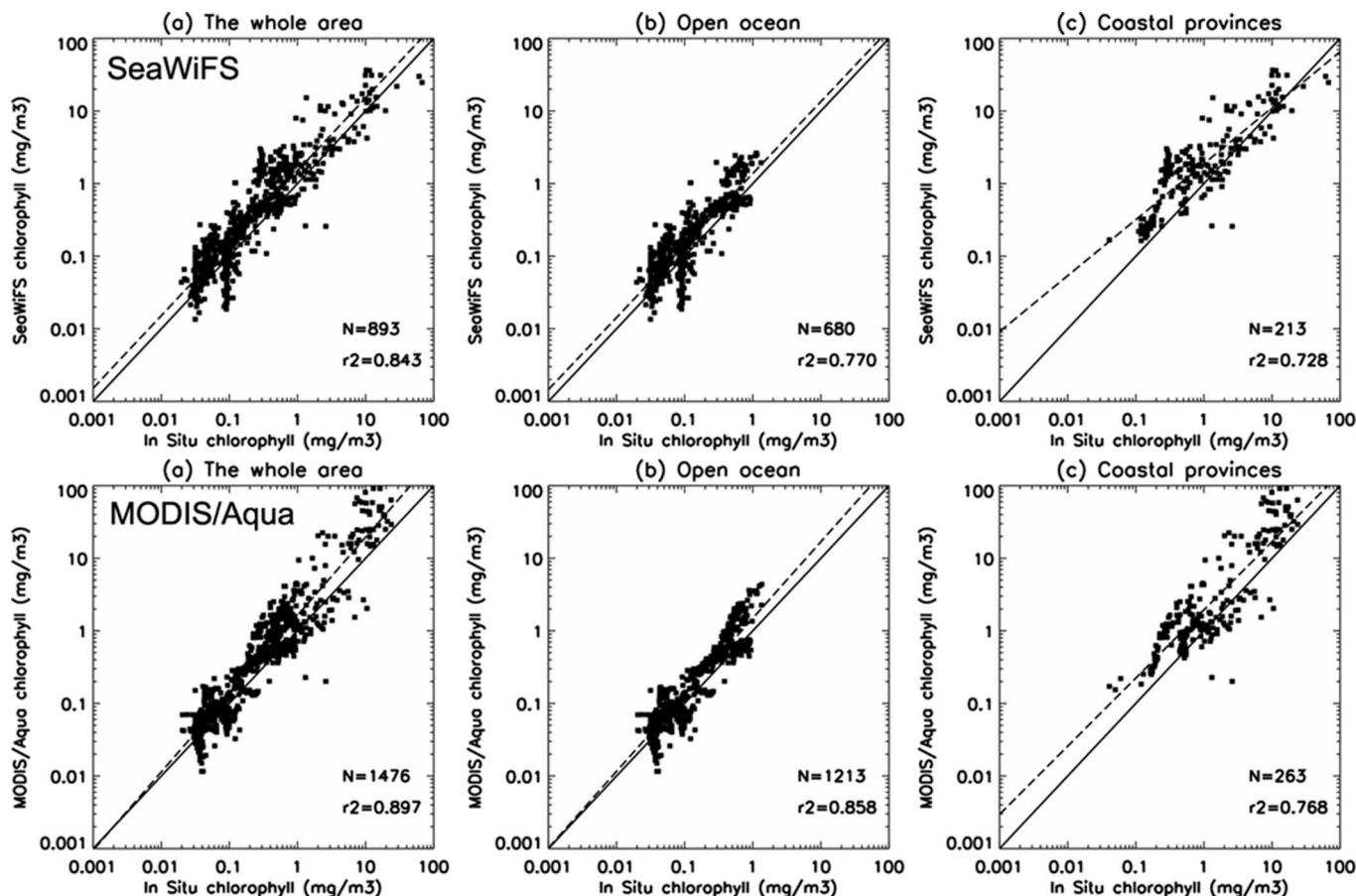


Fig. 4. Scatterplots of (top) *in situ*/SeaWiFS and (bottom) *in situ*/MODIS/Aqua chlorophyll data into (left) global ocean, (middle) open ocean (provinces No. 4, 5, 6, 7, 8, 10, and 18), and (right) coastal area (provinces No. 11, 12, 14, and 15). N is the number of points used for the computation; r^2 is the coefficient of determination. The dark line is the ideal regression line; the dotted one is the real regression line.

The associated errors are obtained by the same way (just replace “S” by “ES”). The combination of SeaWiFS’ and MODIS/Aqua’s candidates is the same as for Algorithm WA1.

As far as log-transformed values are concerned, (8) are derived using the same techniques as those used in Algorithm WA1; the combination is then similar.

B. Results

Algorithms WA1 and WA2 were applied on the North Atlantic basin for the whole year 2003. Both algorithms were used on untransformed and log-transformed values. For each of the four cases, matchups between the combined and *in situ* data were examined, and rms errors and biases were computed on untransformed and log-transformed values. In each case, the values of the rms error and of the bias range between those of SeaWiFS and MODIS/Aqua used alone. RMS errors and biases computed on chlorophyll and on log chlorophyll show that for Algorithm WA1 (combined data at SeaWiFS’ resolution) and log-transformed chlorophyll, the results are better than when using untransformed data. The results are opposite for Algorithm WA2 (combined data at MODIS/Aqua’s resolution) (Table III). They also indicate that Algorithm WA2 gives better results than Algorithm WA1. Therefore, the application of Algorithm WA2 on untransformed values gives the lowest rms errors and biases.

The application on untransformed values comprises computations of many lognormal distributions, therefore needing a large amount of CPU time. Therefore, the application on log-transformed values is the fastest one, with duration up to 10 min with a 700-MHz processor for both algorithms for one daily map on the North Atlantic basin. The application on untransformed values is quite slower: 10 to 20 min (Algorithm WA1) and 20 to 30 min (Algorithm WA2).

As an example, Fig. 8 shows the result of the application of Algorithm WA2 on the untransformed values for April 1, 2003 (day for which the spatial coverage is the worst, left panel) and August 13, 2003 (day for which the spatial coverage is the best, right panel). The spatial coverage of the combined data is 15% and 34%, which is an improvement of 3% and 11%, respectively, compared to the initial SeaWiFS coverage.

IV. OBJECTIVE ANALYSIS

The second approach studied is based on optimal interpolation used in oceanography by Bretherton *et al.* [15]. This methodology has been already successfully applied to satellite altimetric data in various applications [16].

Its aim is to determine the value of a field θ at a point in space and time given various measurements of the field unevenly spread over space and time $\Phi_{\text{obs},i}$ (with $i \in [1, N]$). It computes an interpolated grid-point value as a weighted-linear

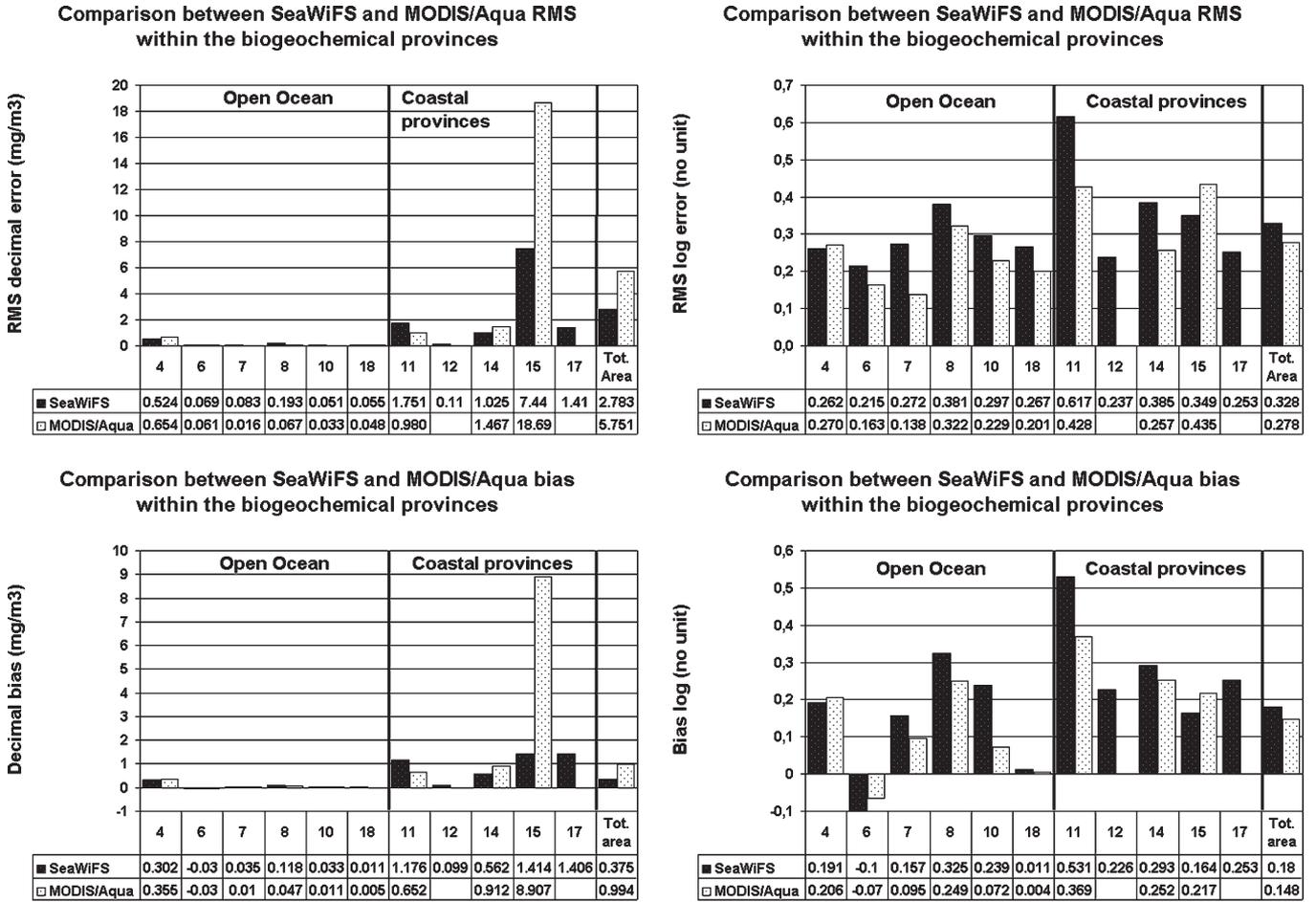


Fig. 5. (Top) RMS error computed on (left) chlorophyll and (right) log chlorophyll between *in situ* data and SeaWiFS (black) and MODIS/Aqua (dotted) chlorophyll datasets for the 11 provinces and the whole basin. (Bottom) Biases computed on (left) chlorophyll and (right) log chlorophyll.

TABLE II
BIOGEOCHEMICAL PROVINCES IN THE NORTH AND EQUATORIAL ATLANTIC BASIN [21]. IN THE CLASSIFICATION OF THE ECOLOGICAL GEOGRAPHY OF THE SEA SUGGESTED BY LONGHURST [21], FOUR BIOMES APPEAR: THE COASTAL BIOME, THE POLAR BIOME, THE WESTERLIES BIOME, AND THE TRADE-WIND BIOME

Province #	Biome	Province	Association with province #
1	Polar	Boreal Polar Province	15
2	Polar	Atlantic Arctic Province	4
3	Polar	Atlantic Subarctic Province	4
4	Westerlies	North Atlantic Drift Province	x
5	Westerlies	Gulf Stream Province	x
6	Westerlies	North Atlantic Subtropical Gyral Province (West)	x
7	Trades	North Atlantic Tropical Gyral Province	x
8	Trades	Western Tropical Atlantic Province	x
9	Trades	Eastern Tropical Atlantic Province	8
10	Trades	South Atlantic Gyral Province	x
11	Coastal	NE Atlantic Shelves Province	x
12	Coastal	Canary Coastal Province	11 (MODIS/Aqua) – x (SeaWiFS)
13	Coastal	Guinea Current Coastal Province	11 (MODIS/Aqua) – 12 (SeaWiFS)
14	Coastal	Guianas Coastal Province	x
15	Coastal	NW Atlantic Shelves Province	x
17	Trades	Caribbean Province	14
18	Westerlies	North Atlantic Subtropical Gyral Province (East)	x
20	Coastal	Brazil Current Coastal Province	14
22	Coastal	Benguela Current Coastal Province	11 (MODIS/Aqua) – 12 (SeaWiFS)

The last column shows the association chosen between biogeochemical provinces where there are matchups (indicated by "x") and where there is none.

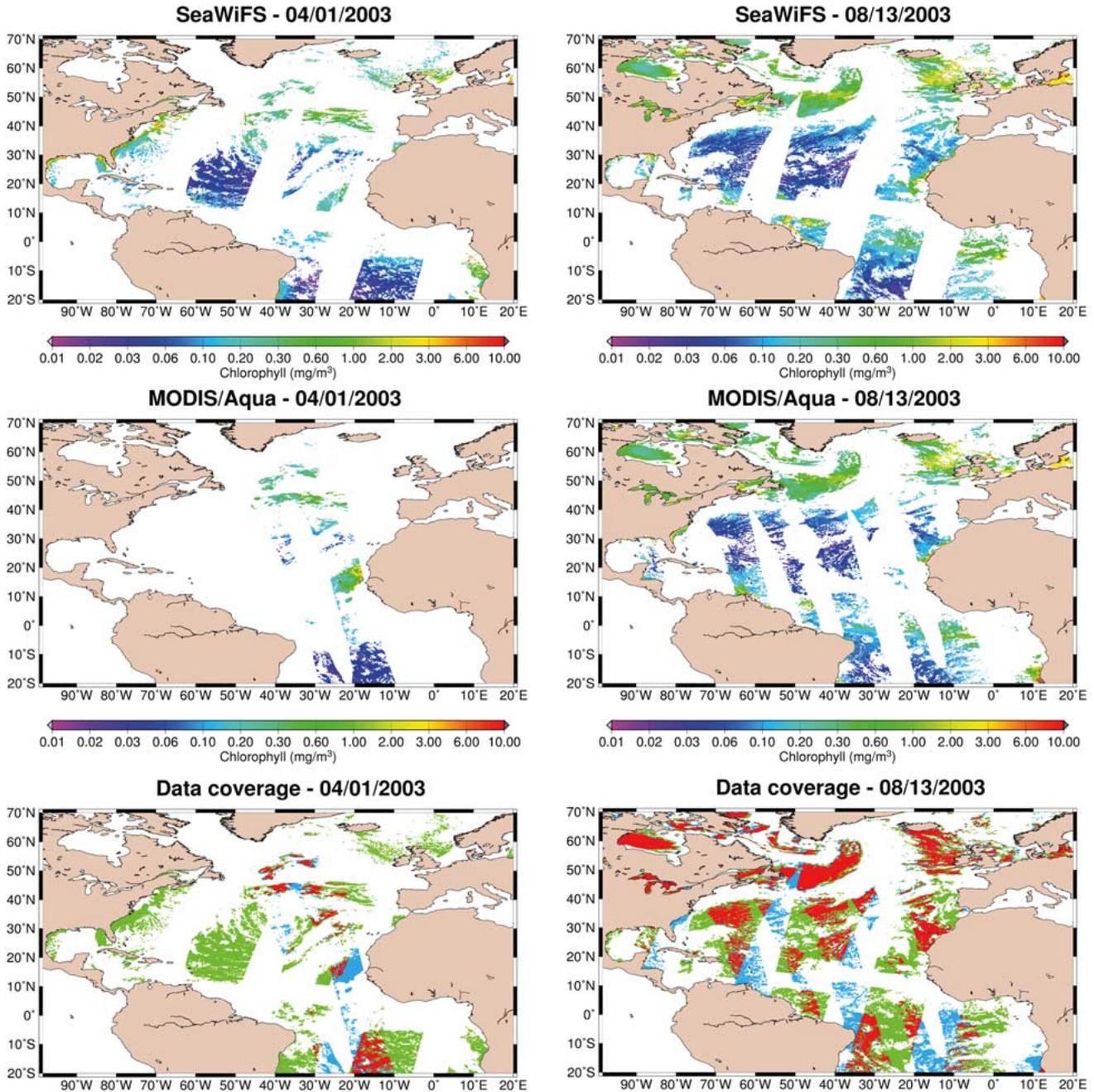


Fig. 6. (Top) SeaWiFS, (middle) MODIS/Aqua, and (bottom) a map of the data coverage of the two sensors for (left) April 1, 2003 and (right) August 13, 2003. Clear gray is for values lower than 0.01 mg/m^3 , and dark gray is for values greater than 10 mg/m^3 . On the maps of the data coverage, white is for no sensor, blue is for MODIS/Aqua only, green is for SeaWiFS only, and red is for both sensors.

combination of observations. The weights are optimal and take into account the spatial distribution of observations relative to one another (spatial correlation), the instrument errors and biases (that we have computed from the matchups in Section II), the variance of the field, etc. Two observed values at corresponding space/time coordinates do not have to be identical, and an interpolated variable of a field does not have to be identical with an observed value at corresponding space/time coordinates. Therefore, this method compensates for sensor-to-sensor differences in instrument design and characteristics, calibration peculiarities, and data processing.

A. Description

The estimated chlorophyll field is computed as follows:

$$\theta_{\text{est}}(x) = \sum_{i=1}^N \sum_{j=1}^N A_{ij}^{-1} C_{xj} \Phi_{\text{obs}^i} \quad (9)$$

with $\Phi_{\text{obs}^i} = \Phi_i + \varepsilon_i$, which is the chlorophyll value measured by the sensor, Φ_i is the *true* value of chlorophyll, ε_i is the measurement error, and A_{ij} is the covariance matrix between the observations:

$$A_{ij} = \langle \Phi_{\text{obs}^i} \Phi_{\text{obs}^j} \rangle = \langle \Phi_i \Phi_j \rangle + \langle \varepsilon_i \varepsilon_j \rangle \quad (10)$$

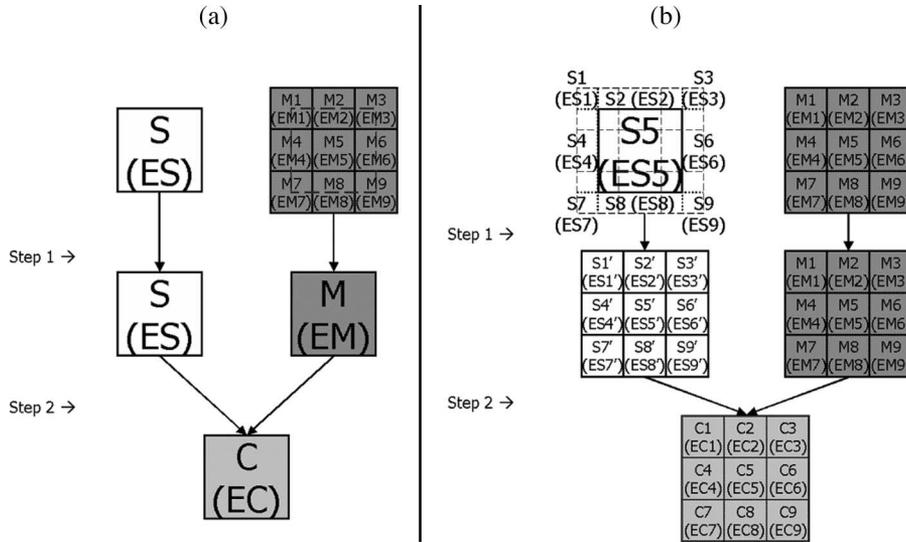


Fig. 7. (a) Description of the first algorithm of weighted averaging. X and EX are the pixel's chlorophyll concentration and the associated error, respectively (in milligrams per cubic meter with untransformed values or in no unit with log-transformed values). X is "S" for SeaWiFS, "M" for MODIS/Aqua, and "C" for the combined data. (b) Description of the second algorithm of weighted averaging: Same notations as for (a).

TABLE III
COEFFICIENTS OF DETERMINATION r^2 , RMS ERRORS, AND BIASES BETWEEN *In Situ* AND COMBINED CHLOROPHYLL OBTAINED BY WEIGHTED AVERAGING (WA1 AND WA2), AND BY OBJECTIVE ANALYSIS IN THE NORTH ATLANTIC BASIN

	N	Stat. on chlorophyll (mg/m ³)			Stat. on log chlorophyll (no unit)		
		r^2	RMS	Bias	r^2	RMS	Bias
SeaWiFS	893	0.53	2.78	0.38	0.84	0.33	0.18
MODIS/Aqua	2249	0.63	5.75	0.99	0.88	0.28	0.15
Algorithm Chlorophyll	1288	0.47	5.48	0.83	0.88	0.29	0.16
WA1 Log chlorophyll	1288	0.47	5.22	0.76	0.88	0.29	0.16
Algorithm Chlorophyll	2268	0.47	3.76	0.53	0.86	0.30	0.18
WA2 Log chlorophyll	2268	0.47	3.96	0.55	0.86	0.30	0.17
Objective Chlorophyll	2261	0.58	3.58	0.60	0.85	0.34	0.21
analysis Log chlorophyll	2283	0.63	3.70	0.57	0.86	0.32	0.20

Statistics for SeaWiFS and MODIS/Aqua alone are reported in top two lines. N is the number of points used for the computation.

and C_{xj} is the covariance vector between the observations and the point to be estimated as follows:

$$C_{xj} = \langle \theta(x) \Phi_{obsj} \rangle = \langle \theta(x) \Phi_j \rangle. \quad (11)$$

The variance of the error associated with the estimated chlorophyll field is given by

$$e^2 = C_{xx} - \sum_{i=1}^N \sum_{j=1}^N C_{xi} C_{xj} A_{ij}^{-1}. \quad (12)$$

The objective analysis used in [16] decomposes the errors into white noise and biases by expressing $\langle \varepsilon_i, \varepsilon_j \rangle$ in the following forms:

- 1) $\langle \varepsilon_i, \varepsilon_j \rangle = \delta_{ij} b^2$ for points (i, j) from different sensors;
- 2) $\langle \varepsilon_i, \varepsilon_j \rangle = \delta_{ij} b^2 + E$ for points (i, j) from the same sensor.

b^2 is the variance of the measurement noise (i.e., the square of the rms errors computed in Section II), and E is the variance of the bias (i.e., the square of the mean ADs in Section II).

Since there are a high number of measurements, the interpolation is done with observations included in an "influence

bubble," whose size is defined by the zonal R_x and meridional R_y correlation scales.

B. Pretreatment

1) *Data:* The practical requirement for the use of objective analysis is that a "first guess" field of the signal should exist, which the algorithm corrects by interpolating the input signal. The input data of the algorithm Φ_{obs^i} are chlorophyll anomalies, i.e., the chlorophyll field minus this "first-guess." Different options are possible for this background field: it can be the previous-day chlorophyll map, the mean of the last-week daily chlorophyll maps, or the monthly climatology. Reynolds and Smith [24] have already addressed this problem of the choice of the background field for the application of objective analysis on sea-surface temperature (SST). A monthly climatology and the previous-day map for SST are both solutions with advantages and drawbacks. The preceding analysis has the advantage to be more accurate than the climatology. However, the climatology is more homogeneous in terms of statistics. This led us to choose the climatology solution. First, because the procedure to estimate the correlation function is more robust and easier to

develop, and second, because it is wiser to use a first guess with a coverage as extended as possible.

This monthly mean field is computed by using a one-year Fourier filtering on the 1997–2003 SeaWiFS monthly data. Indeed, this method only takes into account the general yearly trend of the signal and not the values of outliers that might occur once. The climatology for April and August is shown in Fig. 9.

All the following treatments have been computed on untransformed anomalies (chlorophyll minus monthly climatology) and log-transformed anomalies (logarithm to the base 10 of the chlorophyll minus logarithm to the base 10 of the monthly climatology).

2) *Variance*: The temporal variance was computed on untransformed and log-transformed chlorophyll values, as shown in Fig. 10. To avoid marked differences between neighboring points, which are inadequate for the objective analysis, a 50-km median filter was applied to both fields. Fig. 10 shows that the variance is relatively weak in the low-productivity oligotrophic regions such as the subtropical gyres and high everywhere else on the continental margins, in the subpolar gyre, near the Amazon Plume, near the upwelling of Mauritania, and in the equatorial zone. The variance of the untransformed values [Fig. 10(a)] exhibits sharper gradients than the variance of the log-transformed values [Fig. 10(b)].

3) *Correlation Function*: The empirical space–time correlation function has been first computed on untransformed anomalies in each point of a $2^\circ \times 2^\circ$ grid in the North Atlantic basin using (13). Only the values included in a subdomain centered around each location studied are considered. The computation is done only if at least 50% of the subdomain is filled.

$$\text{CORR}_{\text{emp}}(\Delta\text{long}, \Delta\text{lat}, \Delta t) = \frac{\sum_{\text{map}=1}^{\text{NumberOfMaps}} \frac{\sum_{\text{pair}=1}^{\text{NumberOfPairs}} \text{PROD}(\text{map}, \text{pair}, \Delta\text{long}, \Delta\text{lat}, \Delta t)}{\text{VAR}(\text{subdomain}) \times \text{NumberOfPairs}}}{\text{NumberOfMaps}} \quad (13)$$

where $\Delta\text{long} = \text{long}2 - \text{long}1$, $\Delta\text{lat} = \text{lat}2 - \text{lat}1$, $\text{pair} = (\text{long}1, \text{lat}1, 0) \leftrightarrow (\text{long}2, \text{lat}2, \Delta t)$; *NumberOfPairs* is the number of pairs of values included in the subdomain; and *NumberOfMaps* is the number of maps of chlorophyll in the time series used. The PROD product is given by

$$\begin{aligned} \text{PROD}(\text{map}, \text{pair}, \Delta\text{long}, \Delta\text{lat}, \Delta t) &= [\text{chl}(\text{map}, \text{long}1, \text{lat}1, 0) - \text{mean}(\text{subdomain})] \\ &\times [\text{chl}(\text{map}, \text{long}2, \text{lat}2, \Delta t) - \text{mean}(\text{subdomain})] \end{aligned}$$

where chl stands for chlorophyll concentration and the variance of the subdomain, shown at the bottom of the page, where *NLong*, *NLat*, and *NTps* are the numbers of points in longi-

tude, latitude, and time of the subdomain, respectively, and *NumberOfPoints* is the number of the points included in the subdomain.

The choice of the size of the subdomain for such computation is important. A small subdomain gives a very local information. A larger area gives a smoother information but a more accurate one because the computation is performed with more pairs. To have a first idea of the subdomain size, we use space–time correlation scales found in previous studies. Using the semivariogram approach from geostatistics on Level 3 daily standard mapped images (reprocessing 2) for the year 1998, Doney *et al.* [25] found for the global ocean a maximum zonal correlation scale of 250 km and a maximum meridional correlation scale of 200 km. The dimensions 250×200 km of this subdomain are kept constant all over the North Atlantic, and we compute our empirical correlation scales over this subdomain size. Uz and Yoder [26] found low correlation between pairs of images as short as one-day apart. This allows us to work only in two dimensions (longitude, latitude) and then obtain a space correlation function. We also computed the correlation function with a time scale of five days but the results were not convincing. The correlation coefficients have been computed per month and for the whole year to examine a possible seasonal variation of the correlation function.

To model the empirical correlation function, two functions were studied.

First, let us examine the exponential model used by Kwiatkowska and Fargion [12]. The basis function is the following:

$$\text{CORR}_{\text{exp}}(r) = A_1 \exp(A_2 r) + A_3 \quad (14)$$

where $r = \sqrt{(x/R_x)^2 + (y/R_y)^2}$; R_x and R_y are the zonal and meridional correlation radii (first zero crossing of CORR_{exp} with the zonal and meridional axes, respectively), and A_1 , A_2 , and A_3 are the coefficients to determine. The function must complete the following conditions: strictly decreasing, $\text{CORR}_{\text{exp}}(0) = 1$, and $\text{CORR}_{\text{exp}}(1) = 0$ (i.e., null correlation at the first zero crossing). Therefore, we obtain

$$\text{CORR}_{\text{exp}}(r) = (1 - \alpha) \left(\frac{\alpha}{\alpha - 1} \right)^r + \alpha \quad (15)$$

with $\alpha < 0$. The smaller is the α , the more sharp pointed is the function.

The second model is based on an inverse function (A_1 , A_2 , A_3 , and A_4 are coefficients to determine)

$$\text{CORR}_{\text{exp}}(r) = \frac{A_1}{A_2 r + A_3} + A_4 \quad (16)$$

$$\text{VAR}(\text{subdomain}) = \frac{\sum_{\text{long}=1}^{\text{NLong}} \sum_{\text{lat}=1}^{\text{NLat}} \sum_{\text{tps}=1}^{\text{NTps}} [\text{chl}(\text{map}, \text{long}, \text{lat}, \text{tps}) - \text{mean}(\text{subdomain})]^2}{\text{NumberOfPts}}$$

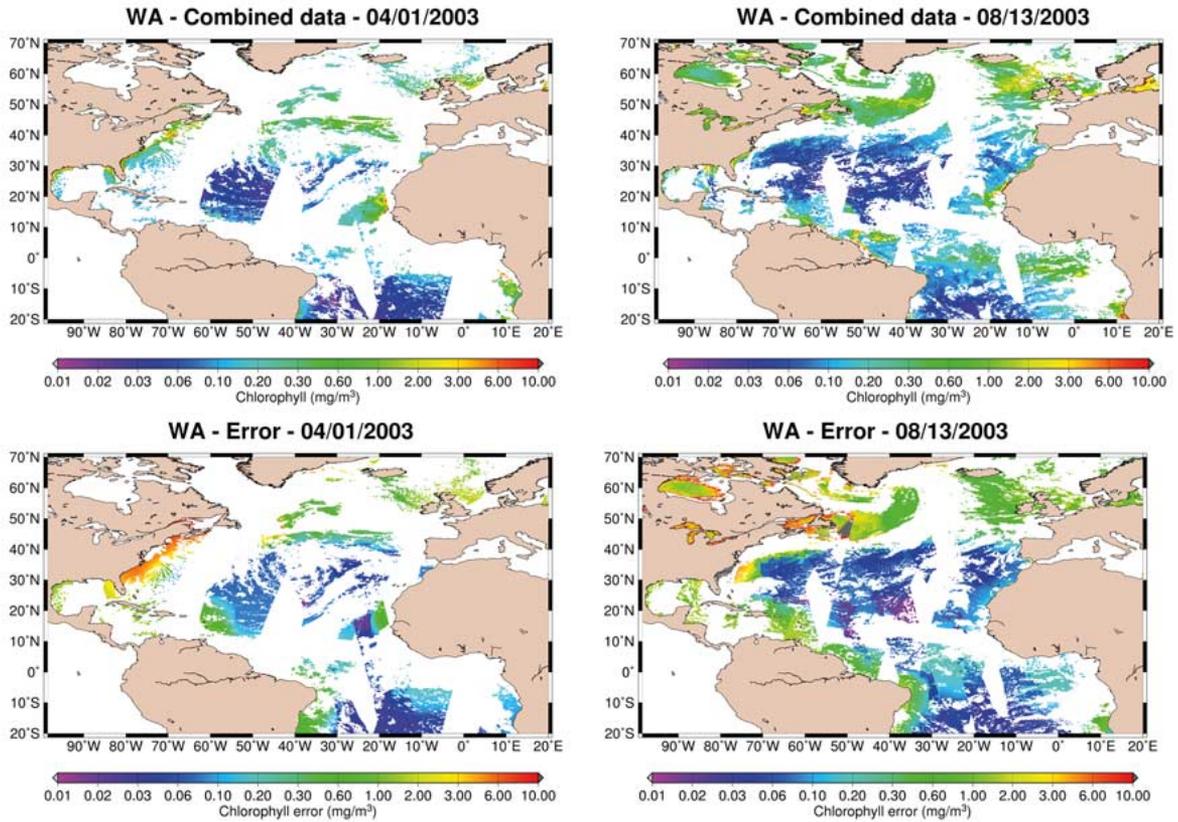


Fig. 8. Weighted averaging. Results of the application of Algorithm WA2 to untransformed values for (left) April 1, 2003 and (right) August 13, 2003. (Top) Map of the combined data. (Bottom) Map of the associated error. (Clear gray) Values lower than 0.01 mg/m³. (Dark gray) Values greater than 10 mg/m³.

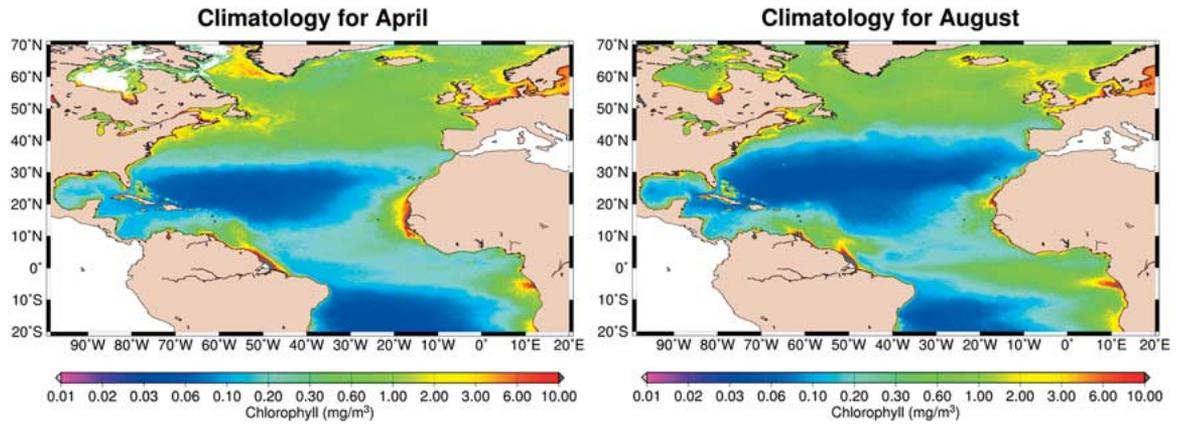


Fig. 9. Climatology for (left) April and (right) August.

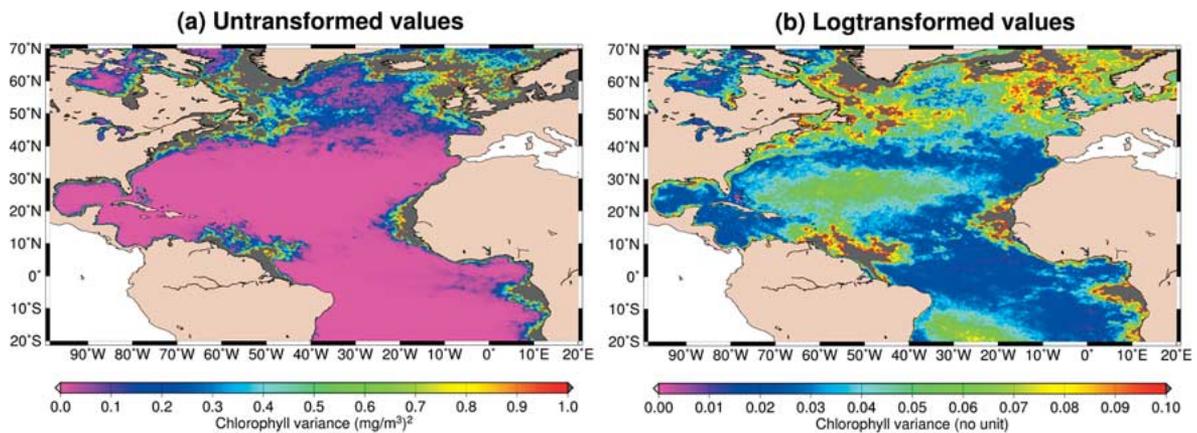


Fig. 10. Temporal variance of SeaWiFS chlorophyll for the year 2003, which is computed on (left) untransformed and (right) log-transformed values. Dark gray is for values greater than 1 (milligrams per cubic meter)² and 0.1, respectively.

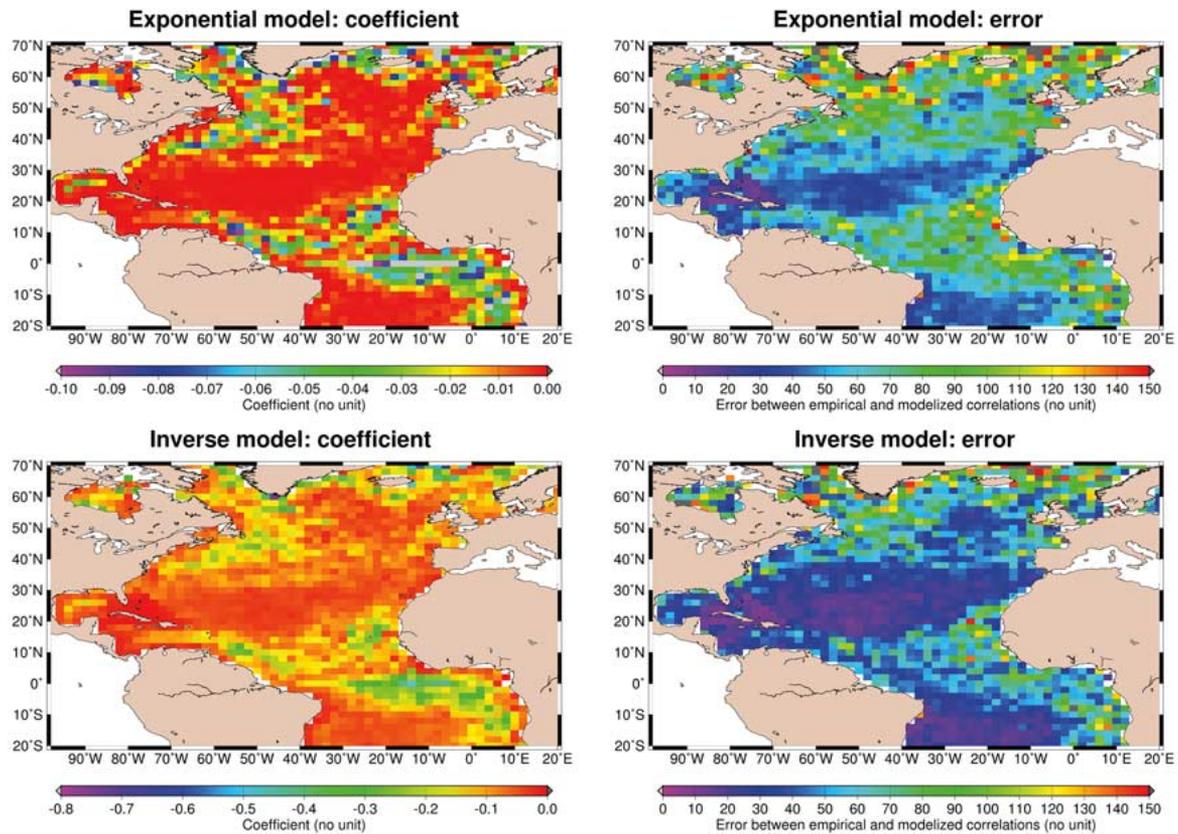


Fig. 11. (Top) Spatial distribution of (left) the coefficients α of the exponential model in the North Atlantic basin and (right) the resulting absolute error from the empirical model. (Bottom) Same for the inverse model.

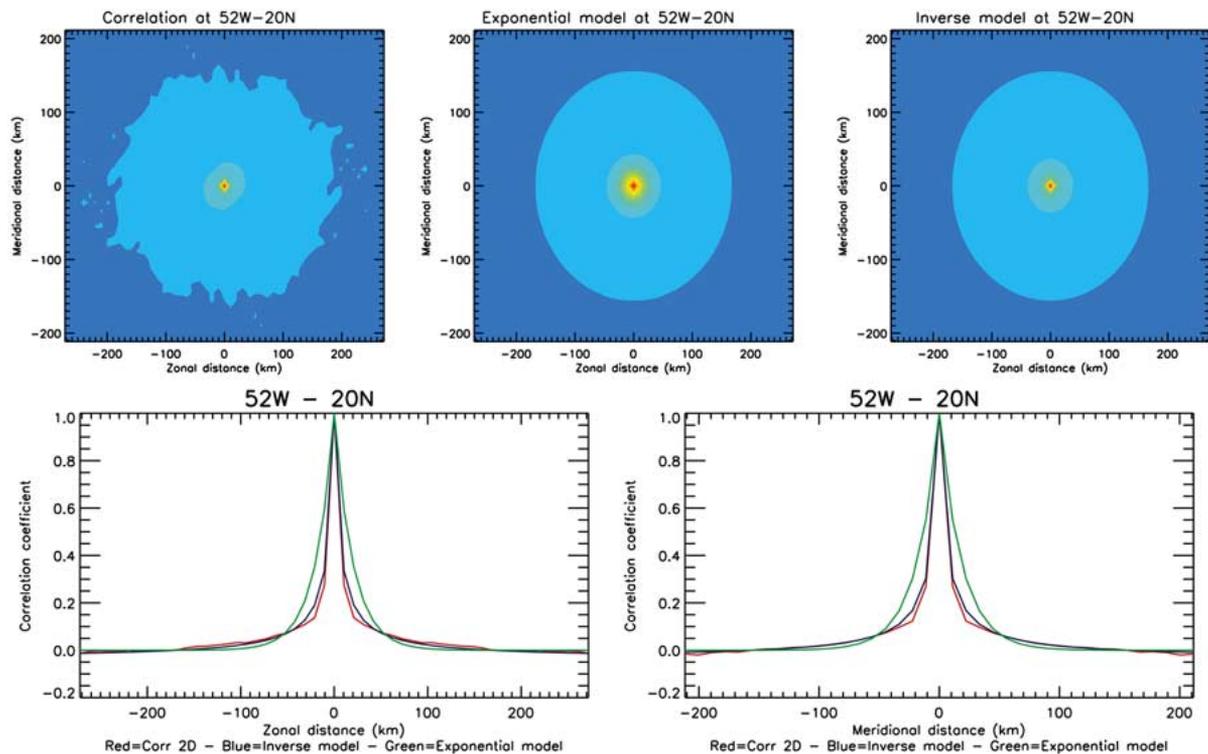


Fig. 12. Spatial correlation coefficients of SeaWiFS anomalies for the year 2003 at the grid point 52°W–20°N (Province 6). (Top) Value of the correlation coefficients versus zonal and meridional distances (in kilometers) for (left) the empirical function, (middle) the exponential model, and (right) the inverse model. Color scale: Red is for the value 1.0; dark blue is the zero field. (Bottom) Section for (left) a null meridional distance and (right) a null zonal distance; the empirical function is in red, the exponential model in green, and the inverse model in blue.

which gives the following, considering the above constraints:

$$\text{CORR}_{\text{exp}}(r) = \beta + \frac{\beta(1-\beta)}{\beta-r}. \quad (17)$$

Here, again, the smaller is the β , the more sharp pointed is the function.

The associated spectrum of both correlation models is a positive number for each wavenumber. Therefore, they are both nonnegative definite forms, which are a requisite [15].

At each point of the $2^\circ \times 2^\circ$ grid, α and β were computed so that the model correlation function gives the best fit to the empirical one. Fig. 11 shows the spatial repartition of these coefficients in the North Atlantic basin. For both models, they depend obviously on the biogeochemical provinces: In the subtropical gyres, they are close to zero, and the correlation model is very sharp pointed (i.e., the points are not well correlated between each other).

In the equatorial zone, the coefficients are higher in absolute value, and the correlation model is wider (i.e., the points are better correlated with each other). A monthly analysis reveals an interesting seasonal variation: The correlation scale is larger in summer (i.e., wide correlation model) than in winter (i.e., sharp-pointed correlation model).

Fig. 12 shows sections of the two-dimensional (2-D) correlation function at the grid point $52^\circ \text{W}-20^\circ \text{N}$. Obviously, the inverse model fits better the empirical function than the exponential one. This is confirmed by the maps of the absolute error between the models and the empirical function in Fig. 11.

The same computation has been carried out on log-transformed anomalies, giving the same results.

4) *Correlation Scales*: Zonal (respectively meridional) correlation scales are given by the first zero crossing of the correlation function computed above with the zonal (respectively meridional) axis. They define the size of the influence bubble.

Fig. 13(a) displays the zonal correlation scales for the whole year 2003. They clearly depend on latitude and not on the biogeochemical-province partitioning. A study per month points out that they do not show any seasonal variation. Zonal correlation scales R_x were thus modeled with the following function: $R_x = -0.03 \text{ lat}^2 + 220 \text{ km}$. Thus, the coefficient of determination r^2 is 0.914, showing the excellent fit of the modeled function.

Fig. 13(b) shows the meridional correlation scales for the year 2003. They depend neither on the latitude nor on the biogeochemical-province repartition. No seasonal variation emerges. The scales are close to 150 km all over the basin, and a constant value (150 km) has been used.

The time correlation scale computed (not shown) is about one day everywhere and until five days at few locations in the subtropical gyres, confirming the results in [26]. Therefore, the value of the time correlation scale has been set here to one day.

C. Results

The objective analysis was applied on the North Atlantic basin for the whole year 2003. It was used on untransformed and log-transformed anomalies. The computation was made in order to have a combined data with a resolution of $0.1^\circ \times 0.1^\circ$.

Combined data obtained by the application of the objective analysis to untransformed anomalies (not shown here) are worse than those obtained by the application to log-transformed anomalies. Indeed, unrealistic values can be obtained at locations where the variance between points within the influence bubble is highly different. By this way, the covariance matrix is numerically ill conditioned, and its inversion gives a wrong result. This does not occur with lognormal anomalies because such differences do not exist in this case. Moreover, this method is based on the Gauss–Markov theorem, which is optimal only if applied to Gaussian variables. Therefore, objective analysis must be applied to log-transformed anomalies.

As for the weighted averaging, matchups between the combined and *in situ* data were examined for both cases, and rms error and bias were computed on untransformed and log-transformed chlorophyll. Coefficients of determination, rms errors, and biases computed on anomalies and log anomalies (Table III) confirm that better results are obtained by applying the objective analysis to log-transformed than to untransformed anomalies.

Naturally, for operational purposes, it is impossible to use all the valid points (up to 2000) in the influence bubble. Therefore, to decrease the CPU time, another parameter was added: the number of observations in the influence bubble considered for the computation at one given location. The CPU time increases of about 30 min with a 50-point increment. Different values were tested, and a value of 150 points was chosen. Indeed, selecting more points in the influence bubble does not change any more the chlorophyll value. The CPU time is about 1 h and 30 min for a daily map on a 700-MHz processor, depending clearly on the season (because of the number of points on the original grids).

As an example, Fig. 14 shows the result of the application of the objective analysis on log-transformed anomalies for April 1, 2003 (day for which the spatial coverage is the worst, left panel) and August 13, 2003 (day for which the spatial coverage is the best, right panel). The spatial coverage of the combined data is 50% and 71%, which is an improvement of 38% and 48%, respectively, compared to the initial SeaWiFS coverage.

V. DISCUSSION AND CONCLUSION

The objective here was to present and test two approaches of ocean color data merging for operational purposes. Both were applied on the North and Equatorial Atlantic basins for the combination of SeaWiFS and MODIS/Aqua for the year 2003. Input values were untransformed and log-transformed chlorophyll-concentration values.

The first approach is an error-weighted averaging. Even considered as basic, it improves the spatial coverage by taking into account only the existing satellite values. It consumes low CPU time. Its main limitation is the discontinuity in the accuracy of the combined data, as shown in Fig. 8. If there is only one observed value, the accuracy of the combined value is the accuracy of this sensor's value. If there are two estimates (i.e., an observed value for both sensors), the accuracy of the combined data is improved. Its application on the North

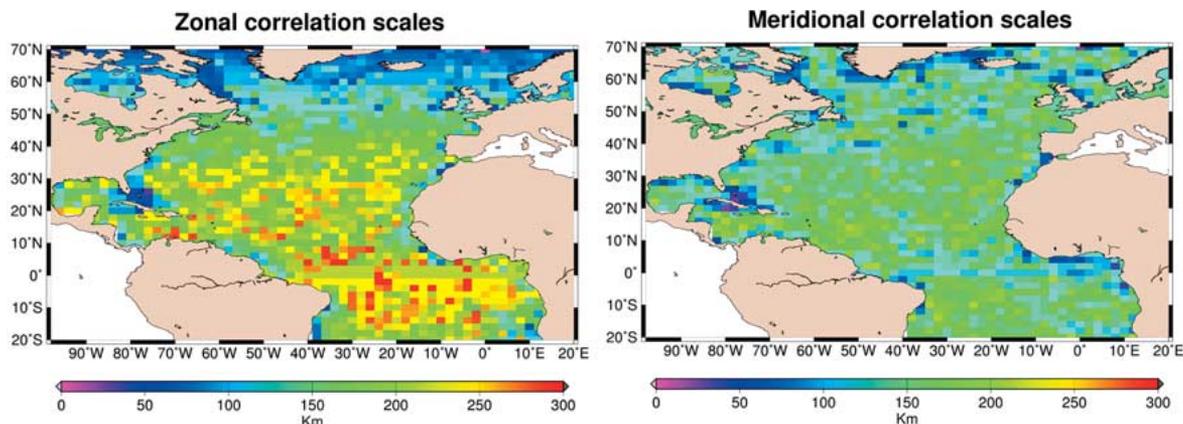


Fig. 13. Spatial repartition in the North Atlantic basin of (left) the zonal and (right) the meridional correlation scales.

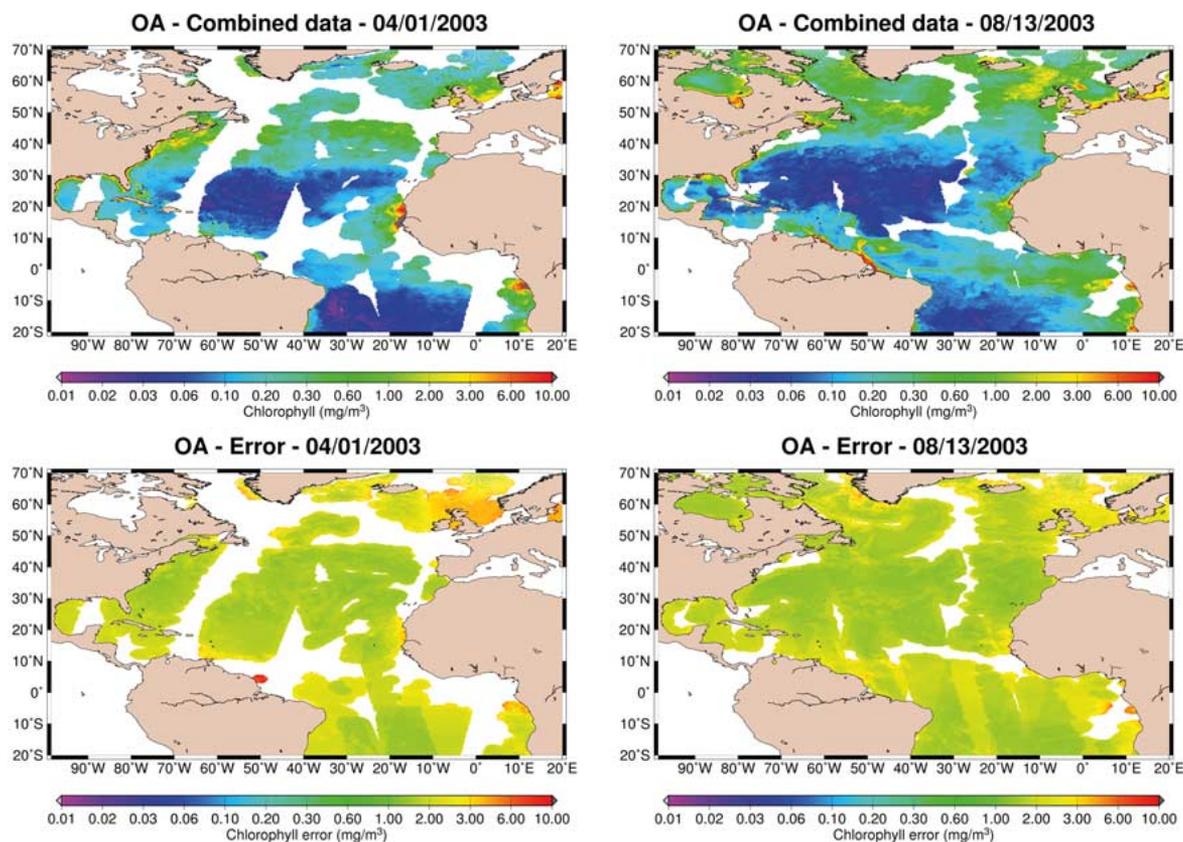


Fig. 14. Objective analysis. Results of the application to logtransformed values for (left) April 1, 2003 and (right) August 13, 2003. (Top) Map of the combined data. (Bottom) Map of the associated error. Clear gray is for values lower than 0.01 mg/m^3 , and dark gray is for values greater than 10 mg/m^3 .

Atlantic basin shows that the quality of the combined data does not differ much between the applications on untransformed and log-transformed values.

The second approach is an objective analysis. Given a first guess and statistical parameters computed directly on satellite data, this approach interpolates values where there is none. As a result, the spatial coverage of the combined data is improved compared to the error-weighted averaging. Its application on untransformed anomalies sometimes gives poor results, especially because of the lognormal distribution and the time dynamics of the chlorophyll. Therefore, the objective analysis must be applied to log-transformed data. The results shown

here constitute a preliminary application of this method on chlorophyll-*a* data. It thus needs to be improved: tuning, better characterization of signal and noise, correlation function, selection of the observations considered for the computation, etc.

Assessing the quality of the combined data is quite difficult. The best way to do it is to compare the combined data to *in situ* measurements, as we made here. Table III shows that the rms errors and biases of the error-weighted averaging are slightly smaller than for the objective analysis. But, since the matchups are very sparse and rare in space and time, an estimation of the global quality (extreme values, mean, and standard deviation) of the combined data can also be computed.

The CPU time is much quicker for the weighted averaging than for the objective analysis because of the difference of the space coverage between both approaches. Nevertheless, the CPU time is quite low, and both techniques can be applied for operational purposes.

In conclusion, we have tested and compared two approaches for ocean color data merging of SeaWiFS and MODIS/Aqua in the North Atlantic basin for the year 2003: weighted averaging and objective analysis. The main advantage of the objective analysis is its ability to interpolate in space (and time) by taking into account the statistical properties of chlorophyll-*a*. Moreover, matchups between *in situ* and combined data show that both approaches yield about the same rms errors and biases, although the spatial coverage is at least twice as large in the case of the objective analysis as the weighted averaging. It also seems that, due to the nature of the chlorophyll distribution, it is better to use mergers on log-transformed values.

ACKNOWLEDGMENT

The authors would like to thank Orbimage for the SeaWiFS mission, the OceanColorWeb support, and the SeaBASS and the AMT archivers (especially G. Moncoiffé for her efficient help) of chlorophyll, and the contributors to these databases. The authors would also like to thank W. Gregg for giving access to the recent SeaBASS datasets and for the helpful comments. They would also like to thank P. Mazzega (LEGOS), S. Ruiz (CLS), and F. Mélin (JRC, Ispra, Italy) for fruitful discussions, as well as the three reviewers for the very constructive comments that improved this paper.

REFERENCES

- [1] H. R. Gordon and M. Wang, "Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with SeaWiFS: A preliminary algorithm," *Appl. Opt.*, vol. 33, no. 3, pp. 443–452, Jan. 1994.
- [2] A. Morel, D. Antoine, and B. Gentili, "Bidirectional reflectance of oceanic waters: Accounting for Raman emission and varying particle phase function," *Appl. Opt.*, vol. 41, pp. 6289–6306, Oct. 30, 2002.
- [3] J. E. O'Reilly, S. Maritorena, B. G. Mitchell, D. A. Siegel, K. L. Carder, S. A. Garver, M. Kahru, and C. McClain, "Ocean color chlorophyll algorithms for SeaWiFS," *J. Geophys. Res.*, vol. 103C, no. C11, pp. 24 937–24 953, Oct. 1998.
- [4] W. W. Gregg, W. E. Esais, G. C. Feldman, R. Frouin, S. B. Hooker, C. R. McClain, and R. H. Woodward, "Coverage opportunities for global ocean color in a multimission era," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 5, pp. 1620–1627, Sep. 1998.
- [5] J. Xiong, G. Toller, V. Chiang, J. Sun, J. Esposito, and W. Barnes, *MODIS Level 1B Algorithm Theoretical Basis Document*, (2005). [Online]. Available: http://www.mcst.ssaai.biz/mcstweb/info/online_doc.html
- [6] G. Meister, E. J. Kwiatkowska, B. A. Franz, F. S. Patt, G. C. Feldman, and C. R. McClain, "Moderate-Resolution Imaging Spectroradiometer ocean color polarization correction," *Appl. Opt.*, vol. 44, no. 26, pp. 5524–5535, Sep. 2005.
- [7] C. R. McClain, W. Esaias, G. Feldman, R. Frouin, W. Gregg, and S. Hooker, "The proposal for the NASA sensor intercalibration and merger for biological and interdisciplinary oceanic studies (SIMBIOS) program," NASA Goddard Space Flight Center, Greenbelt, MD, NASA TM-2002-210008, 2002.
- [8] E. J. Kwiatkowska, "Comparisons of daily global ocean color data sets: MODIS-Terra/Aqua and SeaWiFS," in *MODIS Validation, Data Merger, and Other Activities Accomplished by SIMBIOS Project: 2002–2003*, NASA Goddard Space Flight Center, Greenbelt, MD, NASA-TM-2003.
- [9] B. A. Franz, "A long-term intercomparison of oceanic optical property retrievals from MODIS-Terra and SeaWiFS," in *MODIS Validation, Data Merger, and Other Activities Accomplished by SIMBIOS Project: 2002–2003*, NASA Goddard Space Flight Center, Greenbelt, MD, NASA-TM-2003.
- [10] S. Maritorena and D. A. Siegel, "Consistent merging of satellite ocean color data sets using a bio-optical model," *Remote Sens. Environ.*, vol. 94, no. 4, pp. 429–440, Feb. 2005.
- [11] W. W. Gregg and M. E. Conkright, "Global seasonal climatologies of ocean chlorophyll: Blending *in situ* and satellite data for the coastal zone color scanner era," *J. Geophys. Res.*, vol. 106, no. C2, pp. 2499–2516, Feb. 2001.
- [12] E. J. Kwiatkowska and G. S. Fargion, "Merger of ocean color data from multiple satellite missions within the SIMBIOS project," in *Proc. SPIE Symp.—Remote Sensing Atmosphere, Ocean, Environment and Space*, 2002, pp. 168–182. Invited Paper.
- [13] —, "Merger of ocean color data from multiple satellite missions within the NASA SIMBIOS project office," in *Proc. 5th IEEE Int. Conf. Inf. Fusion*, 2002, vol. 1, pp. 291–298.
- [14] —, "Application of machine-learning techniques toward the creation of a consistent and calibrated global chlorophyll concentration baseline dataset using remotely sensed ocean color data," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 12, pp. 2844–2860, Dec. 2003.
- [15] F. P. Bretherton, R. E. Davis, and C. B. Fandry, "A technique for objective analysis and design of oceanographic experiments applied to MODE-73," *Deep-Sea Res.*, vol. 23, no. 7, pp. 559–582, Jul. 1976.
- [16] P.-Y. Le Traon, F. Nadal, and N. Ducet, "An improved mapping method of multisatellite altimeter data," *J. Atmos. Ocean. Technol.*, vol. 15, no. 2, pp. 522–534, Apr. 1998.
- [17] J. W. Campbell, "The lognormal distribution as a model for bio-optical variability in the sea," *J. Geophys. Res.*, vol. 100, no. C7, pp. 13 237–13 254, Jul. 1995.
- [18] P. J. Werdell and S. W. Bailey, "The SeaWiFS bio-optical archive and storage system (SeaBASS): Current architecture and implementation—Introduction," NASA Goddard Space Flight Center, Greenbelt, MD, NASA Tech. Memo 2002-211617, 2002.
- [19] D. B. Robins and J. Aiken, "The Atlantic Meridional Transect: An oceanographic research programme to investigate physical, chemical, biological and optical variables of the Atlantic Ocean," *Underwater Technol.*, vol. 21, no. 4, pp. 8–14, 1996.
- [20] W. W. Gregg and N. W. Casey, "Global and regional evaluation of the SeaWiFS chlorophyll data set," *Remote Sens. Environ.*, vol. 93, no. 4, pp. 463–479, Dec. 2004.
- [21] A. Longhurst, *Ecological Geography of the Sea*. San Diego, CA: Academic, 1998, p. 398.
- [22] B. A. Franz, P. J. Werdell, G. Meister, S. W. Bailey, R. E. Eplee, G. C. Feldman, E. Kwiatkowska, C. R. McClain, F. S. Patt, and D. Thomas, "The continuity of ocean color measurements from SeaWiFS to MODIS," *Proc. SPIE*, vol. 5882, pp. 1–13, 2005.
- [23] G. Zibordi, F. Mélin, and J. F. Berthon, "Comparison of SeaWiFS, MODIS and MERIS radiometric products at a coastal site," *Geophys. Res. Lett.*, vol. 33, no. 6, L06617, DOI: 10.1029/2006GL025778, 2006.
- [24] R. W. Reynolds and T. M. Smith, "Improved global sea surface temperature analyses," *J. Clim.*, vol. 7, no. 6, pp. 929–948, Jun. 1994.
- [25] S. C. Doney, D. M. Glover, S. J. McCue, and M. Fuentes, "Mesoscale variability of Sea-viewing Wide Field-of-view Sensor (SeaWiFS) satellite ocean color: Global patterns and spatial scales," *J. Geophys. Res.*, vol. 108, no. C2, 3024, DOI: 1029/2001JC000843, 2003.
- [26] B. M. Uz and J. A. Yoder, "High frequency and mesoscale variability in SeaWiFS chlorophyll imagery and its relation to other remotely sensed oceanographic variables," *Deep-Sea Res. II*, vol. 51, no. 10/11, pp. 1001–1017, May/June 2004.



Claire Pottier received the engineering degree from the Ecole Supérieure d'Electronique de l'Ouest, Angers, France, in 2003, and the diploma in signal and image processing from the University of Rennes, Rennes, France. She is currently working toward the Ph.D. degree at the University of Toulouse, Toulouse, France.

Since 2003, she has been working with Collecte Localisation Satellites (CLS) and LEGOS, Toulouse. Her research interests are in the areas of signal and image processing adapted to oceanographic data.



Véronique Garçon received the Ph.D. thesis degree in energy and environmental pollution from the University of Paris VII, Paris, France, in 1981, and the Habilitation à Diriger les Recherches in oceanography from Toulouse University, Toulouse, France, in 1995.

She is currently a Centre National de la Recherche Scientifique (CNRS) Senior Scientist and Head of the Group Physical Dynamics/Marine Biogeochemistry with the LEGOS Laboratory, Toulouse, France. Her research interests are in the areas of marine biogeochemical cycles and dynamics of ecosystems in response to climate changes. This implies activities in remote sensing, coupled physical/biogeochemical modeling, analysis of *in situ* data, data assimilation, and experimental work at sea.



Joël Sudre received the Master's degree from the University of Méditerranée, Aix-Marseille II, France.

He is currently a CNRS Engineer with the LEGOS Laboratory, Toulouse, France, working on satellite data analysis and image processing for oceanographic studies.



Philippe Schaeffer received the engineering degree from the Institut National des Sciences Appliquées (ENSAIS), Strasbourg, France, which is a French superior engineering school, in 1991.

He then worked six years with the French National Research Council (CNRS) as a Research Engineer in the domains of orbitography, geodynamics, and satellite altimetry. Since 1997, he has been a Research Engineer with the CLS Space Oceanography Unit, Ramonville-Saint-Agne, France.



Gilles Larnicol received the engineering degree from the Ecole Nationale Supérieure de Techniques Avancées, Paris, France, in 1993, and the Ph.D. degree from the University of Bretagne Occidentale, Brest, France, in the field of ocean data analyses, in 1998.

Since 1998, he has been working with Collecte Localisation Satellites (CLS) Space Oceanography Division, France. He has experiences both in operational and research oceanography. He is currently the Head of the Oceanography Department of the Space Oceanography Division.



Pierre-Yves Le Traon graduated the French Civil Engineering School of Saint Etienne in 1984. He received the Master thesis degree in dynamical and coastal oceanography of Brest, in 1985, and the Ph.D. degree in physical oceanography from Toulouse University, Toulouse, France, in 1990.

He is formerly the Vice-Director of the CLS Space Oceanography Division. He is now Program Director for operational oceanography systems with the Institut Français de Recherche pour l'Exploitation de la Mer (IFREMER), Plouzané, France. His fields of

interest include operational oceanography, data assimilation, and *in situ* and satellite ocean observing systems.

Claire POTTIER

**MERGING OF OCEAN COLOR DATA FROM MULTIPLE SENSORS :
APPLICATION TO OPERATIONAL OCEANOGRAPHY**

Ph.D. Advisor: Véronique GARÇON, Directeur de Recherches CNRS, LEGOS

Ph.D. defended on December 22, 2006, at the Observatoire Midi-Pyrénées, Toulouse, France

~ Abstract ~

Phytoplankton play an important role in the carbon cycle on Earth, via the carbon dioxide absorption during photosynthesis. If scientific cruises provide high frequency data at high spatio-temporal resolution, the observation from space allows a synoptic description and on long periods of chlorophyll-a, the ocean phytoplankton's main pigment. Each space mission measuring ocean color is limited in ocean spatial coverage (satellite tracks, clouds, etc.). The daily spatial coverage may increase significantly by merging datasets from several sensors. The objective of this Ph.D. was to design, develop and test different methods which combine ocean color data, from the American sensors SeaWiFS and MODIS/Aqua, for near real time applications for operational oceanography. Three concepts have been studied : error-weighted averaging (keeps the structures accuracy but uses only existing data), objective analysis (increases spatial coverage but, in return, smoothes the field), and a last new approach based on the wavelets transform (keeps the structures accuracy and increases the spatial coverage). The operability of these three methods has been demonstrated.

The relevance of using combined data was studied by highlighting the dominant modes of oceanic variability in physical and biological dynamics in the Southern Ocean, by using SeaWiFS + MODIS/Aqua combined data from the Antarctic circumpolar belt during the 2002-2006 period.

Keywords : ocean color, space data / *in situ* data, weighted averaging, objective analysis, wavelets, Southern Ocean

Research field : Image Processing in Space Oceanography

Laboratoire d'Etudes en Géophysique et Océanographie Spatiales (LEGOS)
14 avenue Edouard Belin – 31400 Toulouse, France

CLS / ARGOS

8-10 rue Hermès – Parc Technologique du Canal – 31520 Ramonville Saint-Agne, France

Claire POTTIER

**COMBINAISON MULTI-CAPTEURS DE DONNEES DE COULEUR DE L'EAU :
APPLICATION EN OCEANOGRAPHIE OPERATIONNELLE**

Directrice de thèse: Véronique GARÇON, Directeur de Recherches CNRS, LEGOS

Thèse soutenue le 22 décembre 2006 à l'Observatoire Midi-Pyrénées, Toulouse

~ Résumé ~

Le phytoplancton joue un rôle important dans le cycle du carbone sur Terre, de par l'absorption du dioxyde de carbone au cours de la photosynthèse. Si les campagnes en mer offrent la possibilité d'acquérir des données à haute fréquence et à fine échelle spatio-temporelle, l'observation spatiale procure une description synoptique et sur de longues périodes de la chlorophylle-a, pigment principal du phytoplancton océanique. Chaque mission satellitaire qui mesure la couleur de l'eau est limitée en couverture océanique (traces du satellite, nuages, etc.). La couverture spatiale journalière peut augmenter considérablement en combinant les données issues de plusieurs satellites. L'objectif de cette thèse a été de concevoir, développer et tester des méthodes de combinaison de données couleur de l'eau, provenant des capteurs américains SeaWiFS et MODIS/Aqua, pour des applications en temps réel relevant de l'océanographie opérationnelle. Trois concepts ont été retenus : la moyenne pondérée par l'erreur capteur (conserve la netteté des structures mais n'utilise que les données existantes), l'analyse objective (améliore la couverture spatiale, mais lisse le champ en contrepartie), et une dernière approche innovante basée sur la transformée en ondelettes (conserve la netteté des structures et améliore la couverture du champ). L'opérationnalité de ces trois méthodes a été démontrée.

L'intérêt d'utiliser des données combinées a été montré à travers la mise en évidence des modes de variabilité dominants de la dynamique océanographique et biologique dans l'Océan Austral, en utilisant les données combinées SeaWiFS + MODIS/Aqua de la ceinture circumpolaire pour la période 2002-2006.

Mots-clés : couleur de l'eau, données satellitaires / données *in situ*, moyenne pondérée, analyse objective, ondelettes, Océan Austral

Discipline : Traitement d'Image en Océanographie Spatiale

Laboratoire d'Etudes en Géophysique et Océanographie Spatiales (LEGOS)
14 avenue Edouard Belin – 31400 Toulouse

CLS / ARGOS
8-10 rue Hermès – Parc Technologique du Canal – 31520 Ramonville Saint-Agne